THE PEDAGOGICAL ORIGIN OF HUMAN COMMUNICATION

NIMA MUSSAVIFARD

Submitted to:
Central European University
Department of Cognitive Science

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Primary supervisor: Gergely Csibra
Secondary supervisor: Dan Sperber

Vienna, Austria, 2023

Abstract

Human ostensive communication is believed to be a distinctive cognitive and behavioral trait. This thesis investigates the origin of this communicative system from an evolutionary and developmental perspective. Since addressing this question requires a specified explanandum, chapter 2 explores various accounts of ostension. I suggest that the difficulty of characterizing ostension may originate in evoking mostly mechanistic notions. For instance, *higher-order intentionality* focuses on the internal processes underlying communication and obscures their function. Likewise, *attention manipulation* draws on attentional mechanisms and fails to account for cases that go beyond highlighting perceptual phenomena. I propose, instead, to define ostension as involving *nonnatural marking*, i.e., the act of marking actions as communicative. This function can be implemented in various mechanisms and enables generating communicative means open-endedly.

Chapter 3 discusses the diverse selective scenarios proposed for the evolution of human communication. These scenarios are assessed against five criteria suggested by scholars for sound evolutionary accounts: uniqueness, immediate utility, generality, honesty, and cooperativeness. I show that scenarios stressing non-verbal modalities and the transfer of semantic contents are more successful in satisfying immediate utility and generality. Overall, most theories do not offer convincing explanations for the uniqueness of ostensive communication.

In chapter 4, I present my proposal for the origin of ostensive communication: it evolved to enable teaching technological knowledge. Hominin technology involved opaque skills that demanded demonstration and ostensive marking. Unlike other scenarios, demonstrating technological knowledge uniquely requires open-ended communication. Demonstrations are of immediate utility due to using objects and actions, rather than conventions. Demonstrations also satisfy the generality criterion, because they flexibly exploit stimuli to represent displaced generic contents. Finally, inclusive fitness explains why honest, cooperative communication was possible in early interactions.

If ostensive communication evolved in demonstrations, these may already contain some of the properties of language. In chapter 5, I argue that demonstrations possess a predicate-argument structure. Previous work has suggested that objects in

demonstrations may act as exemplars symbolizing their kind. I propose that actions on these object-symbols work like predicates in revealing hidden properties. The effect of actions on objects and their relations can be interpreted by infants as predicates ascribable to the kinds.

Chapter 6 investigates what a concept of communication involves. The standard answer is that communication requires complex metarepresentations of mental states. This complexity is at odds with the limited abilities of infants. But mentalistic metarepresentations are neither necessary nor sufficient in explaining communication. Ostensive signals rest on decoding rather than metarepresentational inferences—thus, some metarepresentations may be unnecessary. However, mentalizing is also insufficient for explaining communication: the logic of instrumental actions permits interpreting their effect as following from intentions. However, communicative effects are often unavailable for inferring meaning. My proposal is that the primitive concept of communication targets, instead, representational action. When we communicate, we typically convey a content that is detached from our acts. This representational property is absent in ordinary goal-directed actions. This account additionally raises the possibility that metarepresentation emerged for representing external, communicative representations and was later exapted for other purposes.

Declaration of Authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

_____

Nima Mussavifard

Acknowledgements

Many elements worked together to stop this thesis from happening—from the sandfly that bit my foot in the desert and hospitalized me for days to those trying to destroy our university. But the people mentioned here helped make it possible.

I should thank Gergő for bearing with me throughout this turbulent PhD. He taught me how to think critically, how not to take anything for granted, and that whatever follows “It is difficult to imagine” is likely an unfounded argument. He was always open to new ideas and was willing to reconsider common assumptions. He encouraged me to be creative and showed me the way to develop my own ideas. I am grateful for his support and for his guidance.

I should also thank Dan who introduced CEU to me. Before he replied to my first email, I did not know that it was possible for me to study next to such big names! He was always enthusiastic about discussing ideas. He encouraged me to be bold in building theories even when he did not agree with my conclusions.

I thank Vida for giving life meaning, for believing in me even when I didn’t, for always having my back. Things started working out when she joined my life. I would certainly not be here without her.

I am grateful to my family for always being there. I would not be studying science if my dad didn’t explain to me how a honeybee colony works when I was five. The fossils he found me were the best gifts I ever received. I am also grateful to my brother for supporting me from afar, for listening to my complaints, and for checking on me in desperate times.

My old friends have been an important part of my family. Hamed, friend and comrade, would listen to my random ideas for hours on end without showing signs of boredom. His honesty and friendship never stop surprising me. Siavash, friend and comrade, was there in hikes, in fights, in celebrations, in mournings. I always knew there was someone I could count on. Ramin, friend and comrade, would engage with me in long literary, political, philosophical, and scientific discussions—he never let me get away with a bad argument. Talking to him was my university before CEU.

I am also grateful to my colleagues and friends in CEU and elsewhere. There are so many important people, but I especially thank (in alphabetical order): Adrien, Antonio, Behnam, Boris, Christophe, Cristina, Dóri, Edit, Fati, Fra, Giorgia, Ieva,

Laura, Martin, Martín, Mikołaj, Milad, Morteza, Nazlı, Otávio, Peyman, Rachel, Szabad Egyetem, Simily, Zsuzsanna. They made life easier and more memorable in the last years. Last but not least, I thank Réka and Ági for their immense kindness and willingness to help.

To the brave women of my country

Table of Contents

Chapter 1. General Introduction

1.1. What is the Question?

Humans stand out from other animals in many cognitive and behavioral domains. These include culture, technology, cooperation, metarepresentation, and communication. The present dissertation is concerned with the latter. Animals of course communicate in various interesting ways. They use displays, vocalizations, and gestures to influence one another's behavior, and they use others' communication to gain fitness-relevant information. Thus, communication, like most of the other traits, is present in non-human animals too. Yet, we communicate in such distinctive ways that may suggest ours is qualitatively unique. This uniqueness is often sought in the faculty of language (e.g., see entries in Tallerman & Gibson, 2011). Phonology combines auditory elements in systematic ways to make meaningful words. Words are used symbolically to denote concepts and entities. And syntax combines words to convey propositional contents. Neither of these interrelated components is found elsewhere in nature. But even the non-verbal communication that is used when language is not available or not required reveals properties that may be unique to humans.

When we are in a foreign country, we may mime the act of eating to ask where we can find a restaurant. When we cannot talk, say, because we are in a library, we may use pointing and other gestures to request things. We may even use in communication well-formed actions that are otherwise performed for attaining instrumental goals. Moving aside, opening the door, placing an object on the table, and using a tool to carry out its function are all meaningful goal-directed actions independently of communication. However, by embedding them in a communicative context, one can also use these same behaviors to transmit information. Then, moving aside could communicate that someone we know is approaching (Sperber & Wilson, 1995); opening the door conveys that one can now enter the room; placing the object could remind a friend of something they may forget; and tool use may be demonstrated to inform about the function of the tool. Neither of these instances of communication is dependent on language. But the ease with which we turn these goal-directed acts into communicative ones on the fly and the range of information we can thereby transmit suggest that even our non-verbal communication is an evolutionary anomaly. Although other animals and especially non-human primates may also engage in diverse, more

or less flexible, interactions, these arguably do not come anywhere near the flexibility of communication in humans.

That we communicate so flexibly in non-conventional modalities as well may imply that the same communicative system underlies both linguistic and non-linguistic communication. This system is sometimes called *ostensive communication* (Sperber & Wilson, 1995). Explaining the evolution of ostensive communication is important not only due to its uniqueness but also because it is a chief contributor to other seemingly unique domains. Human cultural and technological achievements are dependent to a large extent on communication for transmission of knowledge. Advanced cooperation, too, depends on flexible communication for coordinating actions, especially those directed at future goals. Likewise, the development of mindreading and its metarepresentational format may be argued to rely on some form of communication. However, the converse can also be argued to hold: languages are largely the results of cumulative cultural evolution; the stability and functioning of ostensive communication may depend on cooperative social structures and cooperative cognitive inferences (Bohn et al., 2022; Heintz & Scott-Phillips, 2022); and communication is to a large extent founded upon metarepresentational capacities for computing both mental and public representations (Sperber, 2000). That these human-unique abilities are so inextricably intertwined makes the task of explaining the evolutionary origin of ostensive communication all the more difficult—the causal direction can potentially go either way. Hence, evolutionary theories may radically diverge as to which factor has explanatory priority. The apparent uniqueness of ostensive communication could also render comparative research less informative (see also Bolhuis et al., 2014). All of these have made the evolution of human communication one of the hardest problems in science (Christiansen et al., 2003).

A similar problem may be encountered in studying the development of ostensive communication. Humans acquire knowledge not only through direct experience with the world but also indirectly through communicated information. The representational devices that language provides may additionally have a pivotal role in the conceptual development of children (Carey, 2009; Spelke, 2003). If communication is so central to the acquisition of knowledge in various domains, one must be able to explain it relatively independently. And like evolution, opting for either metarepresentation or communication to account for the other removes a major source of explanation.

Moreover, the flexibility of ostensive communication is difficult to spell out. As I will argue, human communication is an open-ended system. Accordingly, one must explain how a limited cognitive system can meet the productive and inferential demands of such communication. Open-endedness also complicates functional approaches because it could permit usage across adaptive domains. As such, deciding which domain has functional priority over the other remains obscure. Evolutionarily, no selective scenario can be rejected a priori, for our general-purpose communication could confer selective advantage in possibly any adaptive domain, from mate selection to collaborative foraging. One must therefore demonstrate why one functional or phylogenetic account has primacy over alternatives.

Aside from substantive problems, the generality of our communicative system also gives rise to definitional issues. Since we use diverse means for transmitting the intended content (e.g., pantomime, depiction, pointing, spoken language, demonstration, emotional expressions, etc.) and each means reveals distinct problems (e.g., iconicity, attention manipulation, syntactic processes, symbolic representation, turn-taking, etc.), choosing which feature to highlight will not only affect one's specification of the explanatory problem, it will also change one's view on the very nature of the capacities that enable our communication. For example, focusing on pointing or other gestures may lead one to define human communication based on its attention-manipulating or intentional qualities, which may in turn give rise to the conclusion that homologous traits can be found in other great apes. Focusing on spoken language and its symbolic and syntactic qualities, on the other hand, may lead to the conclusion that no homologous or analogous system exists in nature.

My aim in this dissertation has been to navigate through these substantive and definitional problems, by specifying the explananda as clearly as possible and offering explanations that address the problems parsimoniously without taking too much for granted. Thus, the central question of the dissertation is *why humans communicate ostensively*. This question can be approached from various angles. It can be approached from an evolutionary perspective to specify the unique selection pressures that drove the evolution of ostensive communication, the processes that it underwent to arrive at its current state, and the functions of its components. The question can also be approached to spell out the proximate working of the cognitive mechanism responsible for its development. I discuss both general questions throughout the dissertation, albeit with an emphasis on evolutionary issues.

As I state elsewhere, this account, like any other account, cannot be exhaustive in its treatment of human communication. As mentioned, our general-purpose communication is utilized in a diversity of domains drawing on various means. Each of these factors raises distinct questions and demands distinct explanations. For instance, I say little about the spoken modality or syntax, both crucial questions. This is partly because there is much more work done on these questions compared with the more pragmatic questions raised here. But this biased selection of components is also partly due to reverse engineering the processes (Jackendoff & Wittenberg, 2017). Although we may not have access to the exact developmental or evolutionary processes leading to ostensive communication, we can infer which of the building blocks had to be in place to permit further processes. As discussed above, we can communicate even in the absence of a shared code. Moreover, arguably there must be a more or less specified communicative concept that can guide the acquisition of the linguistic devices and put them to communicative use. Thus, even if we attribute the origin of some linguistic elements to non-communicative capacities, explaining linguistic *communication* prompts us to put pragmatics front and center: recognizing and interpreting instances of communication are conceptually prior to the communicative use of syntax, phonology, words, and so on. Thus, my aim to explain human communication leads me to target the pragmatic building blocks that allow for the ostensive open-ended communication that is characteristic of humans. As suggested by the title of the dissertation, I draw inspiration mainly from the theory of *natural pedagogy* (Csibra & Gergely, 2009, 2011) in answering the questions.

1.2. Overview

As discussed, our evolutionary account of human communication is dependent on how we define and delineate the trait. This is sometimes done implicitly. But in chapter 2, I try to offer an explicit characterization of ostensive communication. My main goal is to provide a clear explanandum for studying the origin of ostensive communication. Thus, I find it useful to avoid terminology that is loaded with commitment to one or another theory about the cognitive mechanism underlying communication. To put forward a cognitively pretheoretical definition, I suggest that instead of using proximate mechanistic notions, we define ostensive communication in terms of its ultimate function(s). This would yield a definition that can be utilized in evolutionary and comparative accounts, whatever our theoretical convictions about the mechanisms

involved. Because ostensive communication is (trivially) a type of communication involving ostension, we may first define ostension. I suggest that, if we look at the role of Gricean higher-order intentions in *distinguishing* ostensive communication from other types of behavior that may also influence others (physical coercion, hidden authorship, etc.) or, conversely, the role of ostension in expressing those distinctive intentions, we can take this distinguishing role as the defining feature. Thus, a crucial feature of ostension is that it distinguishes or marks communicative actions. One may, of course, object that they find other features more important or interesting. Notwithstanding, our evolutionary account of human communication must eventually say something about this characteristic feature. Having this in mind, I then discuss other suggested notions to see whether they can offer a fruitful account of ostensive communication. Regarding ostension as involving *higher-order intentionality* is a common approach. However, this notion is used mostly mechanistically to refer to the (meta)representational structure underlying ostensive behavior. As such, it may obscure what these representations are supposed to enable, it does not offer clear criteria to find similar behaviors in other species, and it may rule out simpler implementations of the same function in infants or other primates. *Attention manipulation* is another common notion. This may similarly commit us to (attentional) mechanisms. Although it may be argued that manipulating attention is indeed the function of ostension, there are limits to the explanatory power of this approach. We do regularly draw attention to entities in communication. But when communication is about perceptually absent entities, we are forced to talk about drawing attention to non-perceptual phenomena such as informative intentions. This, I believe, changes the standard meaning of attention and makes the evolutionary processes less straightforward. Importantly, neither notion directly targets the feature of communicative marking. Thus, I suggest that we define ostensive communication as involving what I call *nonnatural marking* or marking$_{NN}$. This is used to refer to the act of flexibly marking actions as communicative. (I analyze Grice's intentionalist meaning$_{NN}$ into non-intentional terms: marking$_{NN}$ and informing$_{NN}$. I discuss the latter in chapter 6.) This functional definition, I argue, can be operationalized to study communication in both non-human animals and human infants without committing to the more controversial and ambiguous mechanisms. Nonetheless, as function and mechanism are complementary, the definition can potentially shed light on the mechanism too.

Having now an explanandum, I explore the evolution of ostensive communication in chapters 3 and 4. In chapter 3, I discuss other theoretical work on the evolutionary origin of human communication. I do this against five criteria that have been proposed in the literature as standards for appraising the soundness of evolutionary theories. These include: uniqueness, immediate utility, generality, honesty, and cooperativeness. There are theories that fare better against these criteria. Specifically, theories that focus on the evolution of the basic pragmatic competence in non-verbal channels can meet the immediate utility and generality criteria more easily. Likewise, theories stressing the transmission of propositional contents, rather than less direct sociopolitical gains, account for generality (including open-endedness) more successfully—for limited animal-typical signals could arguably serve those functions. Overall, uniqueness creates problems for most accounts, as the proposed pressures are either present for other animals or they do not necessitate an open-ended system. It seems that the strongest alternative to compare with my own proposal is that ostensive communication evolved as a non-verbal, productive system for enabling diverse collaborative and cooperative activities.

In chapter 4, I present my arguments in favor of the proposal that our communication evolved initially to enable teaching various types of technological knowledge. Teaching is used here as a behavior with the function of facilitating long-term learning. Thus, it is meant in a broader sense than an institutional or cultural practice. Specifically, I draw on the theory of natural pedagogy which proposes that humans are endowed with cognitive adaptations for facilitating teaching opaque knowledge. Ostensive signals (i.e., eye contact, motherese, contingency) are among these adaptations. These serve to mark as communicative object-directed demonstrations, which are in turn interpreted by children as containing generic knowledge. Thus, teaching variable technology-related knowledge called for an open-ended system that used ostensive signals as markers$_{NN}$ and physical entities, namely actions and objects, as external representational media. This proposal can satisfy the five criteria. Most importantly, the uniqueness of ostensive communication suggests that it has likely evolved for a unique domain. Our dependence on complex technology and other types of socially transmitted knowledge provides this unique domain. Demonstrations are, moreover, of immediate utility in absence of a developed code. They are also general, open-ended systems that use limited components (actions and objects) combinatorially to convey unlimited messages. The teaching scenario relies

on the transfer of knowledge between kin. So, honesty and cooperativeness can be explained by kin selection and inclusive fitness.

Chapter 5 and 6 deal with some of the more proximate consequences of taking non-linguistic teaching as the primary function of communication. In chapter 5, I attempt to show that the ostensive demonstrations that are used by modern humans to teach are capable of transmitting relatively complex messages and that they already possess some of the features of more developed linguistic communication. One such feature that is at the core of our compositional semantics is predicate-argument structure. I suggest that this distinction might already be required for interpreting action demonstrations. The objects in demonstration could be viewed as standing for their kind like verbal symbols, so that any knowledge ascribed to them is ascribed to the whole kind. Thus, objects function as arguments in demonstration. Actions, on the other hand, reveal the hidden properties of object-symbols (e.g., in tool use) or position the objects in spatiotemporal relations to one another. The resulting states and relations may be interpreted by infants as predicates of those objects. Accordingly, actions seem to have a predicative function in demonstration. This creates a compositional system that already departs from holistic animal signals and perhaps provides the foundation for linguistic communication. If our ancestors possessed this depictive system, they could employ it to convey various types of generic knowledge. That demonstration, like language, uses representational devices suggests that their underlying mechanism may be identical. I return to this in chapter 6.

Chapter 6 pertains to the more traditional debates in pragmatics on Grice's meaning$_{NN}$ and the contribution of metarepresentation. Since Grice's (1957) seminal paper, the mechanisms behind our concept of communication have become the subject of heated debate. Taking his formulation of communication as cognitive processes may imply that communication demands several orders of metarepresentation. This poses problems for a cognitively and developmentally realistic account. For instance, it is not obvious that infants who participate in communication entertain any more than one level of metarepresentation. I first argue that many of these metarepresentations are not necessary. Appealing to my notion of marking$_{NN}$, for instance, allows that the marking and interpretation of actions as communicative be performed by specialized, coded ostensive signals without the need to make metarepresentational inferences about the intention of the communicator. Some cases of communication (e.g., conventional speech acts and infant-directed

demonstration) could, furthermore, be parsimoniously explained without relying on intention attribution. More importantly, however, metarepresentation of mental states is not sufficient for explaining ostensive communication. There are famous reasons for it, such as the requirement to add ever more representational levels or conditions to rule out counterexamples to Grice's formulation. There are also developmental reasons for insufficiency. For instance, even if we do not find multiple metarepresentations difficult, the analysis of communication into intentions and beliefs suggests that a link between the two is made in conceptual development. There is however no evidence that first we develop these and then we link them: communication develops simultaneously with, if not earlier than, mental state attribution. A deeper problem concerns whether intention attribution in itself is capable of interpreting communication. In understanding instrumental acts, we may attribute intentions. The content of these attributed intentions could then be leveraged to predict what the agent will do. But unless very standard means are used, we cannot do the same thing with open-ended communicative acts. This is especially so for infants learning novel means of communication who would not know which means have which effects to predict communicative behaviors and their intended effect. Action explanation, on the other hand, typically starts from a perceivable effect that is attributed as the goal of the agent. But the effect on the belief of the receiver is not available in interpreting third-party interactions. Thus, it cannot be used for attributing a communicative intention to the communicator. When the interpreter is the addressee, this process would require metacognitive access to the belief that has been formed as a result of communication and then attributing that to the communicator. Therefore, the working of communicative intentions is less straightforward than it might appear. I then propose that seeing the concept of communication as a primitive concept targeted at external representations is a more viable alternative. When we communicate, we use external stimuli (e.g., actions and objects) to communicate. This fits well with Grice's exclusion of showing as an instance of meaning$_{NN}$. I argue that even when it appears that communication does not involve external representations, for instance because pointing draws attention to non-representational referent objects, the predicate (e.g., BRING the object) is nonetheless detached from the medium and must be inferred. Thus, typically communication *represents* its content. This prototypically happens in linguistic utterances using symbolic lexical items. But it happens also in demonstration where

the action represents, say, a kind of action. Communication allows us to represent concepts in our actions and thereby stipulate informational links between the represented concept and the referent, where these links are not already there. As an example, we can make a link between an object and the predicate PHONE (e.g., by holding it like a phone or by using the word *phone*) to establish that it is indeed a phone or to pretend that it is a phone within the play context. Because of this capacity to make novel informational links on the spot, this representational component of ostensive communication can also be called *informing*$_{NN}$. Hence, I suggest, communication in humans involves marking$_{NN}$ (which enables rendering communicative an open-ended range of actions and other entities targeted by action) and informing$_{NN}$ (which enables using those actions and entities to convey an open-ended range of information). If my claim that basic ostensive communication does not necessitate the attribution of mental states and it requires instead understanding representational actions is correct, then it offers the possibility that metarepresentation evolved not to enable metapsychology but for representing external, communicative representations. The evolution of metapsychology demands a transition from primary representations to a novel higher-order representational format—which is difficult to spell out. The proposal that linguistic representations permitted metarepresentation seemingly avoids this but leaves unanswered the question of how representing linguistic representations was possible without metarepresentation. External representations, on the contrary, allow incremental evolution from beings that can only use and learn from external representations (typically actions and objects) to beings that represent those *as* representations in a full-blown metarepresentational format. Such a format would then be used to produce and represent linguistic devices and exapted to enable metapsychology. Chapter 7 concludes the dissertation.

Chapter 2. Marking$_{NN}$ as a Distinctive Feature of Human Communication

2.1. Introduction

The expressive power of human communication is unrivaled in nature. Whether in the verbal or gestural modality, conventional or ad hoc, our communication permits us to interact in much more flexible ways than any other animal, including phylogenetically close non-human primates (henceforth, primates). Thus, an evolutionary account of this seemingly unique communicative system is of utmost interest. Naturally, to have a well-grounded account, one must first and foremost specify what it is for which an explanation is sought. However, there does not seem to be any consensus in the literature on the explanandum (reminiscent of Ferdinand de Saussure's verdict that in the study of language our theory precedes the object of study). The communicative system that enables our flexible interactions is variously called ostensive or Gricean communication, *inter alia*. But what is ostensive communication? A correct but trivial answer is that it is a communicative system that involves ostension. The question is then: what is ostension? Although this notion, first introduced by Sperber and Wilson (1995), appears to touch on important intuitions, it has proved notoriously difficult to define (Bohn & Frank, 2019; Scarafone & Michael, 2022; Sperber, 2018; Zuberbühler & Gomez, 2018). The definitional disagreements have led researchers to make diverging conclusions about the evolutionary origin of ostensive communication, with some emphasizing its human specificity and others arguing for the existence of precursors in non-human animals.

I believe that the source of these disagreements about the proper characterization of ostensive communication and its uniqueness are largely due to the fact that the existing accounts define ostension and related notions in proximate, mechanistic terms. It is a common practice in behavioral biology to define behaviors in ultimate, functional terms. This allows researchers to find similar behaviors across a wide range of taxa. For instance, signals are sometimes defined as traits that alter the behavior of other organisms, which evolved because of that effect, and are effective because the receiver's response has also evolved (Maynard Smith & Harper, 2003). Or, as another example, cooperation is defined as behavior which provides a benefit to another individual, and which is selected for because of its beneficial effect on the recipient

(West et al., 2007). These characterizations make no strong claims (although they can inform) about the mechanistic implementations of the relevant traits or behaviors, allowing one, for example, to compare communication in taxa as distant as humans and bacteria (Scott-Phillips et al., 2014), and to find analogous cooperation in humans (Palmer, 1991) and vampire bats (Wilkinson, 1984). Likewise, if we define ostension and ostensive communication functionally, we might be able not only to make better judgments as to whether this behavior is human-specific or analogously present in other species, but also to examine whether the postulated function can be served in relatively simpler ways in humans as well. Moreover, such a functional definition can also help us to make sound hypotheses about the ecology in which the ostensive function could be of adaptive value, shedding light on the selection pressures that resulted in the evolution of human communication.

The notions that have been appealed to in the literature for characterizing ostension include, among others, (higher-order) intentionality, overtness, attention manipulation, and audience specification. Whether or not we accept the importance of these phenomena for ostensive communication, they have been used in mostly proximate senses and so may not be ideal for characterizing ostensive communication. In this chapter, I will first discuss these notions and their shortcoming in accounting for what is distinctive of human communication and then suggest that *nonnatural marking* (or marking$_{NN}$ for short) is a more useful notion. This is used here to refer to the ability in humans to mark their actions as communicative (Bohn & Frank, 2019; Fitch, 2015, 2017; Scott-Phillips et al., 2009). This ability enables humans both to use ad hoc means of communication and also to establish channels of communication in development. A genuinely functional notion, marking$_{NN}$ can work as a proximately and cognitively pretheoretical explanandum that may potentially facilitate debates on the uniqueness and evolutionary origin of ostensive communication. Focus on this phenomenon may nonetheless lead us to rethink the cognitive demands of ostensive communication as well. But even if it is not ultimately accepted as corresponding to the research goals of others using the term “ostensive communication”, this phenomenon is evolutionarily interesting in its own right. As I will argue, the flexibility of our communicative system can be understood as largely due to marking$_{NN}$.

2.2. Ostension as Higher-Order Intentionality

Paul Grice's 1957 paper, entitled "Meaning", is the primary source of what we now know as ostensive (or eponymously, "Gricean") communication. There he points out that the word "meaning" is used in at least two distinct senses. Sometimes it is used to refer to how environmental cues convey information to us. In such cases, we are concerned with a factive sense of the word. For example, if smoke means fire, then the presence of smoke entails that there is fire. This he terms *natural meaning*. Human communication, however, informs us in a non-factive fashion. The utterance "there is fire" does not necessarily entail that there is fire. This is *nonnatural meaning* (meaning$_{NN}$ for short). Grice and other advocates of *intentionalism* maintain that the distinction originates in the intentionality of communication, that is, utterances (used broadly to include instances of non-verbal communication as well) are produced with the intention to induce a response or belief in an audience. More relevant to the present chapter, he notes that this first-order intention is not sufficient, as cases of *hidden authorship* (Grosse et al., 2013; Tomasello, 2008), too, are produced with such an intention. This happens, for example, when A leaves B's handkerchief by the scene of a murder to induce in a detective the belief that B is the perpetrator. When one communicates, it is argued, one also has a second-order intention. Thus, "*x* means$_{NN}$ something" is true if one intends the utterance of *x* to produce a certain effect in an audience by means of the recognition of this intention. Consequently, communicative behavior is to be distinguished from other actions by the second- or higher-order intentions that are at work in its production and comprehension. These intentions are aimed at making the otherwise covert first order-intention *overt* (Strawson, 1964).

In *relevance theory*, Sperber and Wilson (1995) develop Grice's insights further by grounding them in a cognitive framework. They argue that when we communicate, we simultaneously express two types of intention, an *informative intention* which aims at informing an audience of something, and a *communicative intention* which aims at informing the audience of the embedded informative intention (roughly corresponding to Grice's first- and second-order intentions, respectively). Because these intentions involve inducing beliefs in the audience, entertaining them fully requires four orders of metarepresentation (e.g., A intends B to believe A intends B to believe *p*). Accordingly, they introduce the notion of *ostension* (or ostensive behavior) which refers to behavior that makes manifest an intention to make something manifest. In other words, ostensive behaviors are those that express communicative (including informative)

intentions (Scott-Phillips, 2015a). Thus, ostension provides evidence of one's thoughts, which the audience is then to infer. As such, *ostensive-inferential communication* could be contrasted with the largely coded signaling characteristic of other animals, which involves encoding a message in a signal by the signaler and decoding it by the receiver (but see Proust, 2016; Sperber, 2018; Warren & Call, 2022). It is argued that what enables the recipient to comprehend ostensive behavior is that it is interpreted as guaranteeing the relevance of its content.

Intentionality is also frequently used in distinguishing communicative systems of varying flexibility among non-human animals. Following Dennett (1983), distinctions are made between different orders of intentionality in production (Zuberbühler & Gómez, 2018). A communicative system of zero-order intentionality is thought to be one in which signalers emit signals without any intention or goal in mind, that is, they only react automatically to environmental stimuli. Further up the scale, first-order intentional systems are those that are produced with the goal of influencing others' behavior. Such communication requires the signaler not only to act according to a functional goal, but also to mentally represent that goal and try to achieve it through various means. Inspired by Bretherton and Bates' (1979) work in developmental psychology, several criteria are utilized by zoologists studying animal communication to test the goal-directedness of signals (e.g., Townsend et al., 2017). These include, among other things, persistence and elaboration in signaling until the goal is reached, and selectively producing or withholding a signal in response to social factors such as audience composition (e.g., whether allies or kin are present) and behavior. First-order intentional communication is ubiquitous in primates, both in the gestural and vocal modality (Fischer & Price, 2017). For instance, bonobos produce a beckoning gesture, resembling that of a human, in which they stretch the arm toward a recipient, sweep it toward themselves, and then twirl the wrist downward. This is usually preceded by a sexual initiation posture. Bonobos show persistence in this gesture until they achieve their goal of prompting the recipient to approach and follow them to a different location for copulation, suggesting that the gesture is produced intentionally (Genty & Zuberbühler, 2014). Although vocal communication used to be viewed as less flexible than gestures (Tomasello, 2008), there is now a growing literature demonstrating intentional calling in animals (Seyfarth & Cheney, 2018). For example, Thomas langurs emit alarm calls when they detect tigers and continue to do so until all group

members have responded with an alarm call, likely to ensure that the predator has been perceived (Wich & Vries, 2005).

Second-order intentionality is true of a communicative system in which signalers and receivers take one another's mental states into account. Thus, a second-order communicator would not only aim at changing the behavior of a recipient but also pursue changing their mental states. This corresponds to Grice's first-order intention and informative intention in relevance theory. As an example, toddlers have been shown to attempt at correcting the misunderstanding of an experimenter, even when the experimenter accidentally hands them a requested object (Grosse et al., 2010)—suggesting that they are sensitive to the mental state of their audience. Since the presence of mental state attribution in non-human animals is debated (Call & Tomasello, 2008), its use in their signaling has not been easy to demonstrate either. Nonetheless, chimpanzees appear to be sensitive to the knowledge state of group members in their alarm calls. When they detect snakes, they are more likely to signal in the presence of unaware group members than in the presence of aware group members. That is, alarm calls are significantly more common if the caller is with group members who have either not seen the snake or have not been present when the calls were given, argued to show that they recognize knowledge and ignorance in others (Crockford et al., 2012).

Unsurprisingly, cases of higher-order intentionality in non-human animals are rarely reported. Consequently, overt intentionality, which demands fifth-order intentionality (or, in other words, fourth-order metarepresentation) and is deemed necessary for enabling ostension, would be absent in other species (Scott-Phillips, 2015b, 2016). However, the complexity of the apparent representational demands creates problems also for ascribing them to prelinguistic children, who nevertheless seem capable of basic forms of ostensive communication (Breheny, 2006; R. Moore, 2014). Despite evidence supporting the presence of first-order mental state attribution in infancy (Kovács et al., 2010; Onishi & Baillargeon, 2005; Surian et al., 2007), it seems unlikely that fourth-order metarepresentation is available early in ontogeny for supporting the development of communication, itself a very early-emerging capacity. However, Scott-Phillips (2015a) argues that children's understanding of "covert intentionality" might suggest that they can also appreciate overt intentionality, as their underlying representational structure is similar. For instance, in one study, 3- and 5-year-old children exhibited the ability to hide communicative cues while trying to prosocially

inform an adult who did not wish to be helped (Grosse et al., 2013). Further studies are required to see whether such sensitivity is present in young infants as well. Regardless, the inherent difficulty of demonstrating mental states in observable behavioral evidence has hampered comparative and developmental studies about the question—leading, for instance, Moore (2016) to say: “the claim that preverbal infants grasp the fourth-order meta-representations considered necessary for Gricean communication is not supported by any empirical data” (p. 226).

From its inception, the Gricean framework has been met with numerous counterexamples questioning the necessity or sufficiency of the reduction of the concept of communication to intentions and beliefs (see also chapter 6). One common counterexample regards a type of situation in which one person A arranges things in a certain way to induce a belief in another person B who, although aware of A’s informative intention, does not realize that A is indeed intending B to realize that A has such an informative intention. That is, the informer has the same higher-order representations as communication proper, but these are nonetheless covert. One popular solution has been to add a mutual-knowledge condition to the formulation, so that for communication to occur, the communicator and the audience should be mutually aware of the intentions involved in the act (Schiffer, 1972). This philosophical solution has, however, been taken as corresponding to the actual cognitive processes in the minds of the interlocutors. But, since mutual knowledge or mutual awareness appear to involve infinite mutual mental states (knowledge of knowledge of knowledge, etc.), several attempts have been made to minimize the cognitive load for communication. For instance, relevance theorists have replaced (mutual) knowledge with the less demanding notion of (mutual) manifestness. And Gómez (1994) suggests that eye contact creates an attentional loop (A attends to B attending to A attending to B, etc.) for the interlocutors resulting in mutual awareness without their engaging in sophisticated metarepresentations—thus permitting overtness in prelinguistic children and primates.

There are good reasons to believe that intentions are indeed at play in human communication. We form intentions in our minds to influence others. And we use various means to act on these intentions. When we are on the receiving side, we might also use our folk intuitions about the intention formation process to yield the best interpretation of the communicative act. However, we would still need reliable behavioral cues through which to recognize communicative from non-communicative

episodes, since, for obvious reasons, we do not have direct access to mental states. It is based on such a recognition that we might be able apply the metarepresentations deemed necessary for a full grasp of communication. Whether the expression and attribution of intentions is indeed the primary mechanism for ostensive communication is a separate question. The relevant point here is that the intentionality of communication is a proximate mechanism for communication (Scott-Phillips, 2015b, p. 803). As such, we should not lose sight of the ultimate function or the adaptive problem that intentionality is supposed to solve. Otherwise, our account becomes circular (to simplify: "What is ostension? The expression of intentions. How is it done? Through the expression of intentions."). As I will argue in more detail, I believe that the function of ostension is the flexible marking of actions as communicative. How this marking$_{NN}$ is implemented may be through forming complex communicative intentions, on the production side, and attributing intentions, on the recognition side. However, since marking$_{NN}$ is the function, it could, in theory, be implemented without the intentional mechanism too. Firstly, even if intentions play a central role in production, it does not follow that comprehension should mirror this productive process. It could, instead, use simpler, less mentalistic heuristics. Secondly, the marking$_{NN}$ function can potentially be present in non-intentional communicative systems (but see Scott-Phillips, 2015a). For example, we can imagine a species that marks$_{NN}$ its action by non-intentionally emitting a sound whenever it is beneficial for both the actor and receiver that the action be interpreted as communicative (e.g., in a teaching context). Thirdly, the mechanism for detecting marking$_{NN}$ in humans might be separate from the mechanism responsible for interpreting the content of the communicative act (see also Csibra, 2010); and, related to the previous point, the production or recognition of marking$_{NN}$ could be through non-mentalistic processes, while recovering the meaning is, say, fully metapsychological.

Without a functional account, and only relying on intentionality, it becomes a cumbersome, if not impossible, task to decide convincingly whether a communicative system is ostensive or not. The same is true of the related notion of overtness (or overt intentionality) which is dependent upon the intentions to be made explicit. Are the well-known alarm calls of vervet monkeys overt? Adult vervet monkeys seem to punish younger individuals that produce alarm calls in the absence of the relevant predator (Caro & Hauser, 1992). There is thus some tacit expectation about the proper use of communicative means. If informing about the presence of predators is accepted as

fulfilling Grice's first-order intention to elicit a response, one might be tempted to say that the punishment shows that vervet monkeys take alarm calls as transmitting information overtly and so retaliate against those that do not follow the "norms" of communication—for why would they punish covert information transfer? Hence, while overtness might be a useful notion for distinguishing between the different ways that humans convey information and their social implications, it is less so for distinguishing ostensive from non-ostensive communicative systems. An account that explicitly states the function which ostensive behavior serves, rather than relying on implicit assumptions, would arguably not face the same issues, since it would urge us to assess whether the observed behavior performs the said function.

2.3. Ostension as Attention Manipulation

As suggested by its etymology, ostension is also accounted for by drawing on the notion of attention and other related notions. Sperber and Wilson (1995) introduce *ostensive stimuli*, as stimuli, such as hand waves and spoken utterances, which are used to make an informative intention mutually manifest. They note that these stimuli must satisfy two conditions. The first condition is that they must attract the audience's attention. For this reason, most ostensive stimuli, they believe, involve attention-grabbing features like loud sounds and striking visual or tactile stimulation. Similarly, spoken utterances are difficult to ignore by native speakers of the language used. Only when the audience would pay attention to an ostensive stimulus voluntarily can perceptually weaker stimuli be used. Therefore, they conclude, ostensive communication cannot achieve its effect sub-attentively.

In line with this first condition, Gómez (1994, 1996) and Moore (2016, 2018) emphasize the attention-manipulating nature of ostension and conclude that ostensive communication is present in primates without the complex metarepresentational demands envisaged by accounts focused on adult humans. Highlighting the communicative function of eye contact in humans, Gómez maintains that other great apes, too, often accompany their gestures with eye contact. In his view, although eye contact in monkeys is usually interpreted as a sign of aggression, in great apes it serves also to deliberately and openly address referential behavior to the attention of the audience. Thus, eye contact endows non-human great apes with an overtness believed to be characteristic of human communication. As such, it must be seen as a signal of ostension or communicative intent. However, as Scott-Phillips (2015a)

remarks, that eye contact is used ostensively in humans does not necessarily mean that its use in great apes is also ostensive (but see Gómez, 2010).

Another behavior that Gómez and Moore see as contributing to the overtness of primate communication is various types of attention-getters. A distinction is sometimes made between two classes of gestures, *intention-movements* and *attention-getters* (Tomasello, 2008; Tomasello & Call, 2019). Intention-movements, such as begging gestures, emanate from the communicator's social intention that the recipient do something; attention-getters, on the other hand, draw the attention of the audience to the actor. Apparently unique to great apes, the latter class includes slapping the ground, poking, leaf-clipping, and throwing objects, *inter alia*. These could dyadically bring attention to the action itself or to an accompanying display, such as a play face or an erect penis in order to initiate play or sexual intercourse, respectively. Thus, the actor wants the performance of some action from the recipient (the "social intention"), and for this she draws the recipient's attention to something (the "referential intention") so that, noticing it, the recipient will do as she wishes. Tomasello (2008) believes that this two-tiered intentional structure is an evolutionary novelty in great apes and could be regarded as a "missing link" between non-human primate communication and the human attention-directing referential communication.

However, as Moore (2016) and Tomasello (2008) note, great apes do not systematically produce gesture sequences in which attention-getters are used to obtain the recipient's attention followed by other intentional gestures that draw the attention to something else. Yet they use another strategy to address their gestures to the attention of the audience: when the audience is not facing them, they walk around to gesture in front of the audience (Liebal et al., 2004). Moreover, chimpanzees produce more attention-getting signals when a recipient is oriented away from them, compared to when the recipient is oriented towards them (Leavens et al., 2010). But when the human audience is oriented towards them, they prefer to produce visually-based gestures. These examples show that great apes have some sensitivity to the attention of their audience and can address their communication appropriately (but see Povinelli et al., 1996). Based on such findings, Moore (2016, 2018) suggests that, in a functional reading of Grice's formulation of communication, non-human primates could be credited with Gricean or ostensive communication. In Moore's minimal account, Grice's first-order intention is fulfilled when a great ape engages in *sign production*. Sign production refers to behavior (e.g., any gesture or facial expression)

targeted at inducing a response in an audience. The second-order intention is, then, enacted by *acts of address* (e.g., eye contact and attention-getters) which bring the sign production to the attention of the audience and make it overt. Because these two acts are functionally separate (rather than embedded as in many other accounts), they can significantly minimize the representational demands of ostensive communication. In this picture, attributing communicative intentions to others requires entertaining only a pair of first-order metarepresentations of the goals that the communicator has in relation to the audience. One can thus uncontroversially ascribe ostensive communication to preverbal infants and non-human primates (see also chapter 6).

Despite offering a parsimonious picture, this functional separation comes at the expense of missing the unique quality of human communication to open-endedly create novel communicative means. What is gained by these minimal components is only more success in achieving the communicative goal due to securing the receiver's attention. As acts of address merely draw attention to behavior rather than being communicatively and conceptually tied to them, sign production cannot go beyond what is already in the communicative repertoire of the animal. Sperber and Wilson's second condition for ostensive stimuli highlights this point: ostensive stimuli must also focus the attention on the communicator's intention or meaning. This feature of ostensive stimuli and, relatedly, the embeddedness of informative intention within communicative intention permits one to recursively communicate that one intends to communicate a content. When the communicative nature of the behavior is established, that is, a communicative channel is secured between the interlocutors, the communicator can use almost any behavior for conveying information. Csibra (2010), like Moore and Gómez, suggests that ostension may sometimes be behaviorally separated from information transfer, due to the existence of specialized *ostensive signals* (e.g., eye contact, child-directed speech, and contingent responsivity) that convey the occurrence of communicative episodes. However, like Sperber and Wilson, he conceptually links these to the informative intention of the communicator. This link guarantees that the accompanying action is interpreted as part and parcel of the same communicative event, permitting flexibility in production, which would not be achieved if ostensive signals were simply attention-getters like those of non-human great apes.

Nonetheless, recently some relevance theorists have moved closer to the ideas of Moore and Gómez in emphasizing the attention-manipulating feature of ostensive

behavior and considering the possibility that primates possess simpler forms of ostensive communication. They propose that although precursors of ostension, or "proto-ostension" (Sperber, 2019), may not be found in other signaling systems in nature, homologous behavior may be found in the ancient forms of attention manipulation (Scott-Phillips & Heintz, 2023). Just as humans sometimes ostensively "show" (as opposed to "mean") things, where ostension does not contribute much to how the act is interpreted (Sperber & Wilson, 2015), so great apes may draw attention to things which are in themselves direct evidence of the intended content (Sperber, 2019). For instance, when attention is drawn to a state of sexual arousal by a male primate (Genty & Zuberbühler, 2014), this is enough for the female to figure out what the male intended to convey. With only rudimentary mindreading abilities, the recipient may recognize that this attention manipulation is intentional—a recognition which may in turn enhance the expectation of relevance. Consequently, attracting the female's attention may increase her willingness to mate with him. Thus, proto-ostension is "a form of interaction where A draws B's attention to some state of affairs in a manifestly intentional way and this elicits in B the expectation that this state of affairs is relevant to him or her" (Sperber, 2019). This requires neither a communicative intention in A, nor the attribution of an informative intention by B.

At the other extreme are cases of "meaning", in which, it is argued, evidence for a communicated content is found only in the communicator's provision of overt evidence of her *intention* that the addressee should believe that content. Here, attention is intentionally manipulated towards one's own intentions, giving it an "auto-deictic" nature (Scott-Phillips & Heintz, 2023). In other words, it brings attention to the fact that the actor intends to communicate (Scott-Phillips, 2015a; Vesper et al., 2021). As can be seen, attention manipulation is sometimes used in senses beyond the standard usage of drawing attention to perceptual stimuli. In this broader sense, attention can be drawn not only to objects and actions, but also to non-perceptual phenomena, such as mental states and facts, which are much more abstract in nature (for a similar usage see Leslie et al., 2004). When we draw attention to objects, we often intend to make others notice things that, although perceivable, might otherwise be overlooked by them. Since mental states can only be inferred, rather than directly perceived, drawing attention to the already inferred intentions seems mostly redundant. Thus, attention manipulation should be understood differently, perhaps in the sense of prioritizing or, alternatively, signaling mental states. Even if such variation in usage is ultimately

accepted, it makes the evolutionary continuity between drawing attention to perceivable stimuli, on the one hand, and drawing attention to intentions, on the other, much less obvious. As a result, both relevance theorists and other scholars emphasizing attention manipulation would still need to explain how this transition occurred.

Like intentional accounts, attention-based accounts of ostension characterize ostension in proximate, mechanistic terms. Even if they are correct in the proximate explanation they offer, they come short of proposing a clearly specified function for the suggested mechanisms. This neglect might result in overlooking the crucial feature of marking the occurrence of communication. Considering marking$_{NN}$ as the function, one can study how attentional, as well as intentional, processes serve to achieve it. Of course, one may argue that manipulating the attention *is* the function of ostension. However, this would leave unexplained a significant design feature of human communication that contributes to its distinctive flexibility.

Function and mechanism are sometimes conflated in developmental psychology, too. For instance, it has been suggested that eye contact leads to gaze following, not due to the communicative nature of eye contact, but rather because eye contact is particularly attention-grabbing (R. Moore, 2014; Szufnarowska et al., 2014). However, there is nothing intrinsically attention-grabbing about eye contact. Indeed, in humans it is attention-grabbing likely *because* it is communicative (see also Csibra, 2010). Besides, it is plausible that ostensive stimuli evolved from behavior that initially had a primarily attention-modulating function. The evolution of marking$_{NN}$ would benefit from allocating attentional resources both to the marking act and (if it is enacted separately) the communicatively marked stimuli. But defining ostension independently of attentional terms allows the unexplored possibility that behavior be marked$_{NN}$ sub-attentionally. One can imagine a species that, say, emits a pheromone while performing a wide range of fitness-relevant actions. If, as a result, these actions acquire a communicative status, the sub-attentional olfactory signal can, according to the present account, be viewed as an instance of ostension. Thus, minimal accounts should not focus solely on whether a behavior matches a certain proximate formulation of human communication. They should additionally consider whether it achieves the same outcome or not.

2.4. Ostension as Audience Specification

A final notion to consider pertains to the fact that ostensive communicators often specify who is targeted by their communicative act. *Audience specification* is thus seen as an integral component of ostension. For example, Moore (2018) notes that for an act of attention soliciting to work, it should be apparent to the recipient that he is being addressed. Once the recipient recognizes himself as the addressee of an utterance, he can then figure out why the communicator wanted him to attend to her communicative act. Similarly, Gómez (1996) asserts that ostension reveals for whom a message is intended.

A somewhat similar phenomenon, *audience effect*, which refers to changes in the signaling behavior of individuals caused by the presence of other individuals (e.g., sexual partners, relatives, and friends), has been observed in primates (Zuberbühler & Gómez, 2018). And, on the receiving side, members of at least one species, chacma baboon, have been shown to infer whether they are a threat-grunt's target based on contextual cues such as the signaler's gaze direction and the nature of recent interactions with the signaler (Engh et al., 2006; Seyfarth & Cheney, 2018).

Audience specification is particularly crucial for human child-directed communication in which children must recognize when they are being addressed to be able to learn from communicative episodes produced for them (Csibra, 2010). Thus, it is argued, a limited set of ostensive signals must unambiguously specify that the infant is the addressee. Eye contact, for instance, shows to the gazed-at infant that he is being addressed. Likewise, the special prosody of child-directed speech ("motherese") and calling the infant's name indicate to him that he is the audience. This recognition can, then, enhance learning by motivating infants to interpret the communicative act as conveying generalizable content.

Audience specification, like the notions discussed above, is no doubt an interesting and important phenomenon. Indeed, signals such as eye contact are frequently used in adult linguistic interactions mainly for specifying the addressee, as the communicative nature of the behavior is evident. However, if one grants that distinguishing communicative from non-communicative action is the main function of ostension, audience specification can be viewed as a proximate means for ensuring that ostensive communication achieves its goal. That is, to communicate a message, you would benefit from addressing it to your target audience so that they attend to it and make the required inferences. Nonetheless, one can communicate the occurrence

of communication, in a broadcast fashion, without a specific audience in mind. For instance, one can pantomime in a public square so that whoever passes by interprets the act as communicative. Moreover, we can interpret communicative acts appropriately when they are addressed to others and even when (as in soliloquy) they are addressed to the communicator herself.

2.5. Ostension as Marking[NN]

As shown above, ostension and ostensive communication are usually characterized in proximate, mechanistic terms. Scott-Phillips (2015a) puts this explicitly: the difference between ostensive communication and the coded communication of other animals is “a difference in the internal mechanisms that makes communication possible” (p. 59). And similarly, “ostension is ultimately a psychological construct, i.e. ostension is not any particular behaviour, but rather any behaviour motivated by a particular cognitive phenomenon, namely informative intentions” (Heintz & Scott-Phillips, 2022, p. 23). As the authors state, this makes it impossible to fully isolate behavioral characteristics, because any behavior could potentially be reinterpreted in a non-mentalistic way.

If one complements mechanistic accounts of ostension with functional accounts, such methodological problems may be avoided—for one would then be able to find various behavioral solutions to the same adaptive problem. Moreover, mechanisms are mostly designed to perform functions. If the function is not clearly specified, the mechanism may also remain obscure. Besides, tying our account to specific mechanisms would make comparative research a cumbersome task. It would be tantamount to defining cooperation only with respect to human social cognition. This would, of course, inhibit the search for analogous behavior in other taxa. Functional accounts would not only enable comparative study, but also permit simpler formulations of the problem—which may ultimately reveal simpler mechanisms in humans too.

Although the ultimate-proximate distinction could be traced to Aristotle (Hogan, 1994), it found its place in biology thanks primarily to the work of Mayr (1961) and Tinbergen (1963). Ultimate and proximate explanations are logically distinct explanations that are necessary for a full account of a biological phenomenon (Bateson & Laland, 2013; Nesse, 2013; Scott-Phillips et al., 2011). Proximate explanations, further divisible into ontogenetic and mechanistic explanations, mainly

have to do with *how* different traits develop and work. Ultimate explanations, on the other hand, include phylogenetic and functional explanations which deal with *why* traits have been selected, and account for their historical and ecological drivers. Although these questions could inform each other, they should mostly be kept separate to avoid confusion. As an example, one may account for male starlings' singing in spring from different angles (Davies et al., 2012): With respect to function, starlings appear to sing in order to attract mates for breeding. Regarding phylogeny, one could say that the singing has evolved from simpler ancestral calls. In terms of mechanism, starlings sing partly because the increasing length of day triggers changes in their hormones. And with respect to ontogeny, one can say that starlings sing because they have a genetic disposition to learn the song of their own species and they have learned the songs from their parents and neighbors.

I reviewed some of the proximate explanations for ostensive communication above, and I will address ontogenetic and phylogenetic questions in later chapters. Now the question is: what is the function of ostension? If ostension in intentionalist accounts serves to express a communicative, and not only an informative, intention, then one of its crucial effects is distinguishing communicative from non-communicative action. The same holds for Grice's higher-order intentions which are presented as distinguishing features of communicative behavior. Thus, a hallmark of human communication is that we not only communicate a content, but we also communicate *that* we are communicating. In other words, human communication often signals its own signalhood (Scott-Phillips et al., 2009). After Grice's notion of nonnatural meaning, I call this process marking$_{NN}$. Hence, I define marking$_{NN}$ as the act of flexibly marking one's action as communicative. "Action" is used here in a broad sense to include also the objects manipulated in action. This is important because, interestingly, humans can also use objects as symbolic, communicative artifacts (see chapter 5). Nonetheless, there seems to be an asymmetry here: intransitive actions can be turned communicative, but objects need to be stipulated as communicative *in action*. So, action appears to be a necessary component, even if there is only a trace of it (e.g., an object placed in a way to suggest it is communicative). I add the qualifier "flexibly" to the definition so as to exclude cases where some behavioral component is tied to a specific communicative act. And "communicative" is used here in a sense largely similar to how signaling theorists use the associated notions. Communication often refers to the use of signals which are, according to influence-based accounts

(Stegmann, 2013), traits that alter the behavior of other organisms, which evolved because of those effects, and are effective because the receivers' response has also evolved (Maynard Smith & Harper, 2003). As such, communication differs from *coercion* in which the receiver's response has not evolved to respond to that specific trait. For example, if one stag pushes another stag backwards, it is a case of coercion, rather than communication, because the response is somehow dictated by the actor's behavior. Communication also differs from *cues* which are features of the world that can be used by animals as a guide to future action. Cues, such as the size of an animal, have not evolved because of their effects on other organisms. Proponents of information-based approaches to communication (Scarantino, 2013; Seyfarth et al., 2010; Stegmann, 2013) emphasize that this definition is not complete unless it takes account of the information-transferring feature of communicative traits. Thus, according to another classic definition, signals are seen as "behavioral, physiological, or morphological characteristics fashioned or maintained by natural selection because they convey information to other organisms" (Otte, 1974, p. 385). Scarantino (2013) suggests that, although the influence-based account has its own merits, to distinguish communication more effectively from coercion one should note that a signal must be specialized for influencing recipients *by carrying information* to them. Otherwise, the response to coercion may also be designed by natural selection.

Hence, as a result of marking$_{NN}$, an otherwise non-communicative action will acquire the status of communicative action. Of course, since these action tokens have by definition not evolved for influencing receivers communicatively, they are not signals in the strict sense of the term. However, since our cognitive system has most likely evolved to use marked$_{NN}$ action, as a *type*, for influencing receivers through transmission of information, and the response to this type of action is also specialized to retrieving information from it, it would still make sense to classify this type of action as communicative in the evolutionary sense.

This formulation of ostensive communication corresponds functionally to Gricean formulations (for another functional account see Moore, 2018). The function of marking$_{NN}$ corresponds to Grice's second-order intention. And the marked$_{NN}$ action corresponds to the first-order intention to elicit a response or belief in the audience. For instance, when one shakes an object or makes eye contact before presenting an object, the shaking and eye contact function as markers$_{NN}$ rendering the presentation communicative. In chapter 6, I will discuss the function of conveying an informational

content, or simply informing$_{NN}$, in more detail. The marked$_{NN}$ action is typically not a communicative action. But an already communicative action could also be marked$_{NN}$. This is not problematic for the current account, since, although the ultimate function of marking$_{NN}$ is to render action communicative, it must be implemented using proximate means. These proximate means can then be applied outside their ultimate domain. Moreover, the way marked$_{NN}$ actions transmit their informational content need not be similar to other, non-marked$_{NN}$ communicative means. Thus, the proximate means of marking$_{NN}$ can be exploited also to alter the way biological signals (e.g., laughter) convey information, in order, say, to confer a novel social use to them or to tap into distinct cognitive processes in the audience.

Marking$_{NN}$ contributes significantly to the flexibility of human communication. Part of the flexibility of our communication is due to the open-ended range of information that communicative (i.e., marked$_{NN}$) acts can potentially convey. This we can call *informational open-endedness*. Informational open-endedness is to some extent due to the ability in humans to attribute mental states to communicators as the causal drivers of their utterances. However, as I will argue in chapter 6, this flexible quality also, and more crucially, derives from the capacity to use perceptual stimuli as representational media in communication. Another related feature contributing to the flexibility of ostensive communication is that an open-ended range of stimuli can be utilized for communicating information. This feature, which could be called *formal open-endedness*, is due chiefly to marking$_{NN}$ and allows humans to generate novel communicative means on the fly and also to establish various communicative channels.

Understood as marking$_{NN}$, the evolution of ostension should not be seen as a novelty in the ways that attention is manipulated or a novelty in the representational sophistication with which communication is produced and interpreted. Rather, ostension presents a true conceptual novelty. In the present account, ostension implies that interlocutors have a concept of communication (a “naïve signaling theory”, if you like) dedicated to interpreting marked$_{NN}$ actions. Marking$_{NN}$ can not only distinguish ostensive communication from non-ostensive communication, but it accounts also for the distinct types of informative behavior. Although information is transmitted in hidden authorship, too, this type of behavior is not marked$_{NN}$ and so the action is not conceptualized as communication. The social consequences of communication, like commitment to the truth of the message, are therefore not due

(only) to whether the purported informative intentions are made overt or not. They could, instead, be understood as arising from the norms associated with action that is marked and, complementarily, conceptualized as communicative—norms that do not follow non-communicative, however informative, behavior.

2.6. How is Marking$_{NN}$ Achieved?

According to Sperber and Wilson (1995), to focus the attention on the communicator's intention, "the assumption that the stimulus is ostensive must be both manifest enough and relevant enough to lead to optimal processing. This condition is generally met by stimuli which both pre-empt the attention and are irrelevant unless treated as ostensive stimuli" (p. 154). Such stimuli can be coded or non-coded. Linguistic utterances are examples of the former which are suggested to be irrelevant noises unless taken as ostensive stimuli. Non-coded ostensive stimuli include, among other things, movements that are performed with attention-arresting rigidity and are not intrinsically very relevant. For example, when Peter leans back in a stylized fashion to let Mary see William coming, Peter's action is best understood as ostensive. Otherwise, his action would be irrelevant. That is, the only relevant assumptions manifested by such action are assumptions about the individual's informative intention (see also Scott-Phillips, 2014).

The functional account of ostension defended here allows both that the two functions of marking$_{NN}$ and informing$_{NN}$ be enacted separately or by the same act as in Sperber and Wilson's examples. In the latter case, the communicative stimulus signals its own signalhood (Scott-Phillips et al., 2009). These *self-ostensive* behaviors acquire their communicative status either because they have been somehow established as communicative in development (e.g., through learning and/or innate dispositions) or because they have some quality in themselves which invites the audience to infer that they are produced for a communicative purpose. Similarly to Sperber and Wilson, Royka et al. (2022) suggest that communicative inferences are structured around an expectation that communicative actions will efficiently reveal that they are not world-directed. Drawing on the result of their experiments, they argue that people do not depend solely on direct cues, such as ostensive signals, for recognizing communicative actions. Instead, people analyze the probable goals behind other's actions through an inferential mechanism. As a result, communicative actions should be shaped so that they are unlikely to be produced while pursuing world-directed

goals. Hence, communicative movements should be rare under the distribution of movements that people produce when acting on the world. Likewise, Bohn and Frank (2019) suggest that, in recognizing communication, children follow a logic that could be paraphrased as "Why else would they have done that—other than to tell me something?" Thus, the communicative value of an action itself might serve as prima facie evidence for it being communicative.

In a study demonstrating humans' ability to provide evidence for, and infer, the occurrence of communication, Scott-Phillips et al. (2009) devised an "embodied communication game" for pairs of adult participants. In this game, the participants scored points based on whether they managed to finish on identically colored squares. Importantly, they had no means of communicating to each other except for their movement patterns. The results of the study show that successful pairs managed to create communicative "conventions" in which the communicative nature of the movement and what color it denoted was signaled through unexpected behaviors such as oscillations and loops.

Similarly, Moore and colleagues (2015) conducted studies which suggested that two-year-old children, but not domestic dogs, may be able to infer communicative acts. To test this, an experimenter lifted and shook one of two buckets via a centrally pulled rope to indicate that the bucket contained some reward. This was performed either intentionally or accidentally, and it was accompanied or not accompanied by ostensive signals (i.e., eye contact and motherese). The results revealed that only children understood the experimenter's act in intentional conditions, and that ostensive pulling of the rope made no difference to their success. The authors concluded that human children might be capable of inferring communicative intent from a wide variety of intentional actions without necessarily requiring ostensive signals.

Beside these self-ostensive communicative behaviors, the functions of marking$_{NN}$ and informing$_{NN}$ may be enacted and interpreted separately. Csibra notes that "from the perspective of cognitive mechanisms, the attribution of communicative and informative intentions can be temporally and procedurally (but not conceptually) separated" (Csibra, 2010, p. 143). This means that one can recognize the communicative nature of an action without having access to its informational content—a content which may be inferred in a successive process. This possibility to separate the two functions allows that marking$_{NN}$ be performed via specialized *ostensive signals* (Csibra & Gergely, 2011). These signals then render accompanying or following

actions communicative. Because this system is largely based on coded signals, it is less dependent on inferential processes compared to the examples above. Rather, it conveys the communicative status of behavior more directly and perhaps more reliably. Among other things, these limited signals specify the infant as the addressee of communication, are discriminable by newborns, and evoke preferential orientation towards their source. The signals include the intonation characteristic of child-directed speech, direct gaze leading to eye contact, and contingent responsivity to the infant's behavior in a turn-taking manner. There may still be other signals which perform a similar function either because they have a matching status in development or due to their co-occurrence with ostensive signals. For example, Hirai and Kanakogi (2019) demonstrated that, for infants, horizontal, as opposed to vertical, hand-waving gestures induce identity encoding for the cued objects. Based on such findings they suggest that hand-waving may also serve as an ostensive signal.

Specialized ostensive signals can actively mark ad hoc actions as communicative. They may, moreover, establish channels of communication in development, in which case the actions in the established modality acquire a communicative status whether or not they are accompanied by ostensive signals. For example, the gestural modality may be established as a channel due to its co-occurrence with ostensive signals. Likewise, the spoken modality might be established because it is produced in a child-directed intonation, preceded by eye contact, and produced in a contingent manner. Alternatively, speech might be self-ostensive, signaling itself that whichever stimulus is produced with the characteristic auditory and articulatory features is communicative and carries a relevant informational content. The seemingly innate sensitivity of newborns to speech stimuli strengthens the latter possibility (Vouloumanos & Werker, 2007).

There are several developmental studies that illustrate infants' sensitivity to ostensive signals and their disposition to interpret accompanying actions as communicative. In a study by Pomiechowska and Csibra (2022), 12-month-old infants interpreted pointing gestures, but not object grasping, as referring to a subsequently labeled object. However, if the child-directed labeling preceded the grasping action, infants took the target object as the referent of the novel word, as they did in pointing events. This suggests that while pointing has a communicative status for humans early in development (either because it is learned early or it is privileged as an innately-specified communicative signal), grasping is conceptualized as an instrumental, goal-

directed action (Woodward, 1998), unless marked ostensively as a communicative action.

However, Behne et al. (2005) present studies which indicate that infants interpreted pointing and gazing as referring to a baited container only when they were marked by signals such as eye contact. But infants performed at chance when those deictic gestures were produced absent-mindedly and without any communicative cue. These results could be construed as showing that ostensive signals serve here as markers$_{NN}$ of the deictic gestures. Nevertheless, another interpretation could be that the signals only disambiguate the nature of the already communicative gestures. In this reading, infants perform at chance in non-communicative conditions, not because they do not understand the gestures as communicative, but rather because there is evidence against such a default interpretation (e.g., looking away).

Schulze and Tomasello (2015) tested whether 18- and 26-month-old children can comprehend indirect communicative acts. In their experiments, children were encouraged to engage in a game with the experimenter. When the child needed a missing piece, the experimenter held up a key that could be used to open a container holding the piece. This was done either using ostensive signals, inadvertently, or intentionally but non-ostensively. Children of both ages interpreted only the ostensive showing as an indirect request to take the key and retrieve the missing toy.

Ferguson and Waxman (2016) show that when sine-wave tone sequences are presented with objects, 6-month-old infants fail to use these for categorizing the objects. In similar setups, young infants had categorized objects based on the spoken (Waxman & Markow, 1995) and signed (Novack et al., 2021) utterances. Nevertheless, when the same tone sequences were embedded in a communicative context (i.e., synchronized with the mouth movement of one experimenter who makes eye contact with the infant and another experimenter) the infants subsequently used these artificially produced sounds for categorizing the objects—implying that the ostensive signals were taken as marking$_{NN}$ the auditory stimuli and promoting them to a communicative channel.

Moreover, there are some studies that test infants' communicative interpretations of contingent action. Deligianni et al. (2011) showed 8-month-old infants animated objects without human features that did or did not respond contingently to the infants' gaze fixations. Their results demonstrated that the object's orientation to other stimuli elicited (like a deictic gesture) a congruent visual orientation from infants only if the

object had previously shown a contingent response to the gaze. Thus, even the behavior of a non-human, animated object seems to be understood as communicative when presented with an ostensive signal (see also Tauzin & Gergely, 2018).

Therefore, humans seem to be able both to infer the communicative nature of actions based on various cues and heuristics and also to use more direct evidence via coded ostensive signals. Following Sperber and Wilson's terminology, we can term any stimulus that $marks_{NN}$ behavior an ostensive stimulus. These stimuli, then, include inference-oriented ad hoc stimuli (e.g., instrumentally non-efficient movements), conventional stimuli that serve as $markers_{NN}$ (potentially, waving), and the specialized ostensive signals. Is the developmental emergence of ostensive communication due to purely inferential processes or coded ostensive signals? There are reasons to believe that the latter is the case. First, there are many studies that show early recognition of communication based on ostensive signals, some of which were mentioned. However, there is much less evidence in support of such recognition with the help of inferential means in young infants. (This could, of course, be simply because the question has not been explored enough.) Second, many communicative actions, such as action demonstrations which are crucial in teaching, are often well-formed instrumental actions aimed at achieving a goal. As a result, instrumentally sub-efficient or rare movements cannot always be a reliable source for recognizing communication. Third, young infants' world knowledge about the range of instrumental goals is limited. As a result, it would be a costly heuristic for them to rely on the inefficiency or rarity of actions. Ostensive signals, on the other hand, provide reliable and minimal cues to the occurrence of communication that can bootstrap the development of the communicative system. Fourth, communicative actions are rarely produced in absence of ostensive signals or other established means. We often communicate in face-to-face interactions where eye contact and other signals are abundant. Indeed, as argued before, the absence of ostensive signals may be taken as evidence against communication. Such inferential heuristics might, instead, facilitate the recognition of the content-bearing action. Even in the presence of ostensive signals, one is left with the problem of figuring out which parts of the behavior are more likely to be meant as the $marked_{NN}$ behavior. Inferential heuristics regarding what behaviors tend to be used as, say, gestures, could narrow the hypothesis space. Thus, based on the observations above, ostensive signals may provide more efficient means of rendering actions communicative for infants.

There are still many outstanding questions that have not been addressed here. As the focus in this chapter is mainly providing a functional characterization of ostension and ultimately an explanandum for evolutionary accounts of communication, how this process is exactly represented in the mind is not fleshed out. As mentioned, one crucial question is how ostension is connected to the content-bearing action in the mind of the receiver, that is, how the receiver recognizes which part of a stream of actions is to be taken as marked$_{NN}$ by the ostensive stimulus. As will be discussed in chapter 5, a core feature of most linguistic as well as non-linguistic communication is that it has a predicate-argument structure: an entity is identified in communication, to which a property is ascribed. In parent-child interactions, which are argued to be the cradle of communication both in development and evolution, the argument or referent is typically an object of which something is predicated. If so, then the manipulation of, and reference to, objects in communication may provide a useful clue that the action accompanying ostensive stimuli is indeed meant to convey a content. This way the infant can bind the marking$_{NN}$ to the action that is to be marked$_{NN}$. Of course, communication is not restricted to such object-oriented interactions. But such a heuristic may help both to narrow the hypothesis space and establish a communicative channel (e.g., speech or gesture)—a channel which can then independently be taken as communicating information.

2.7. Is There Marking$_{NN}$ in Non-Human Animals?

There are no convincing examples of ostension in non-human animals in the sense used here. This may partly be because the question has not been investigated. Moreover, the proximate formulation of the problem makes the discovery of similar traits less straightforward due to the inherent difficulty of identifying mental states. And the relatively complex cognitive processes that are often attributed to ostensive communicators are unlikely to be found in other animals. However, some similar traits have been recorded that may eventually lead to finding both homologous and analogous traits in animals.

Gómez (2010) suggests that enculturated apes could learn to use eye contact in human-like ways. For instance, he reports the case of an infant gorilla that appeared to transform manipulative behaviors, like taking a human by the hand, into communicative acts through making eye contact with humans and schematizing the behavior. Albeit interesting, further systematic studies are needed to investigate

whether actions are indeed marked$_{NN}$ by primates, that is, whether the action becomes communicative as a result of eye contact. Receptive communication might provide a clearer setting in which to explore primates' grasp of ostensive signals. Kano et al. (2018) tested this in chimpanzees by conducting studies that asked whether chimpanzees, like human infants (Hernik & Broesch, 2019; Senju & Csibra, 2008; but see Gredebäck et al., 2018), follow a human experimenter's referential gaze to objects after the production of ostensive signals (i.e., eye contact and name-calling). The results revealed that, while human ostensive signals did not enhance chimpanzees' gaze following to the target object compared to controls, they evoked longer looking at both objects to the sides of the experimenter. Thus, they concluded that chimpanzees appear to expect that the ostensive signals precede information about the objects. These results may be due to the chimpanzees' experience with humans. Alternatively, they may show traits homologous with the more flexible use of ostensive signals in humans due to shared ancestry.

Interestingly, domestic dogs present some of the most human-like communicative behaviors among animals. Thanks possibly to a relatively long history of domestication, they have evolved a disposition to receiving communication from their human companions (Kaminski, 2009; Kaminski & Piotti, 2016). Similarly to human children, they distinguish between a communicative (i.e., preceded by ostensive signals) and a non-communicative extended index finger (Kaminski et al., 2012). Moreover, like human infants, they are more likely to follow the direction of gaze to peripheral objects if this follows ostensive signals (Téglás et al., 2012). Thus, dogs may be more receptive to human communication than the phylogenetically closer great apes. This emphasizes the significance of ecology in the evolution of ostensive communication. Nonetheless, more research is required to understand whether dogs use perceived ostensive signals only for disambiguating some human communication in a relatively inflexible manner, or, like humans, they understand it as open-endedly turning actions into communicative ones. One could test, say, whether dogs treat an unfamiliar action as an object-directed, referential gesture when it is produced in an ostensive context.

On the productive side, too, dogs exhibit various human-like informative behaviors, such as directing attention to objects (Kaminski & Piotti, 2016). Moreover, Fitch (2015) has suggested that the play bow of dogs and other canids may be taken as communicative markers. Indeed, many animal species use such communicative

behaviors to signal the playful nature of their actions in order to avoid unwanted aggression (Palagi et al., 2016). While social play may be a fruitful domain to look for potential marking$_{NN}$ behavior, as it provides a context in which formal open-endedness can be adaptive, such "metacommunication" (Mitchell, 1991) seems to lack the two-tiered communication characteristic of ostensive communication. That is, although play signals communicatively mark behavior, they do not seem to mark the playful behavior *as communicative*: whether for individual (e.g., learning) or social (e.g., bonding) gains, social play need not necessarily be communicative in itself. As a result, play markers may not be functionally very different from other expressive behaviors.

2.8. Conclusion

In this chapter, I surveyed some of the notions that have been used in characterizing ostension and suggested that they are mostly proximate, mechanistic notions and, as such, they may not be ideal for a definition that can encompass different examples of the same phenomenon across taxa in the animal kingdom, and across age ranges in humans. If, instead, one offers an ultimate, functional account that is targeted at the adaptive problem that ostension solves, one may find (or at least investigate) different proximate implementations of varying complexity. I suggested that defining ostension as the act of marking$_{NN}$, that is, flexibly marking actions as communicative, may provide such an account. Higher-order intentionality is one alternative notion that is often used to account for ostensive communication. Yet, while it is available to adult humans, it might not always be necessary, and it might not be sufficient for explaining crucial features of ostension. Importantly, it does not in itself elucidate the human ability to produce novel communicative means. Manipulation of attention, as another notion used in expounding ostension, also comes short of accounting for this ability. Although ostensive stimuli often do possess an attention-grabbing feature, the communicative nature of ostensive actions is not a perceptual phenomenon. Hence, its abstractness necessitates a conceptual understanding of communicative behavior that goes beyond mere manipulation of attention. Lastly, whereas the specification of audience facilitates communicative interactions, it also fails to explain the distinctive, open-ended properties of human communication.

The present formulation permits both relatively complex inferential mechanisms which likely rely on attributing mental states and simpler mechanisms that are largely

dependent on coded signals. Moreover, defining ostension in terms of the adaptive problems enables and facilitates research into the ecologies in which such behavior could be beneficial. Thus, it allows for comparison between the various selective scenarios proposed for the evolution of human-specific communication. In the next chapter, I discuss and assess the strength of these scenarios in explaining the unique design features of ostensive communication. In chapter 4, I argue extensively why I think teaching technological and cultural knowledge may present the most suitable environment in which ostensive communication and marking$_{NN}$ can be of adaptive value. Specifically, I suggest that formal and informational open-endedness are most beneficial if the communicated content cannot be determined through natural selection. Humans' variable technology and other types of socially transmitted knowledge offer exactly such a context, where the content is by definition not fixed in space and time.

Chapter 3. Selective Scenarios for the Emergence of Ostensive Communication

3.1. Introduction

The previous chapter provided an explanandum for evolutionary explanations of human communication. I suggested that ours is a communicative system that involves ostension, and ostension is best understood in terms of marking$_{NN}$—that is, flexibly marking actions as communicative. This enables formal open-endedness in producing novel communicative means and contributes also to informational open-endedness, both of which features distinguish ostensive communication from other signaling systems in nature. Now one can ask: what were the selection pressures that led to the evolution of ostensive communication? Alternatively: what was the original function of ostensive communication? Since the function of ostension was proposed to be marking$_{NN}$, other ways of understanding these questions are to ask what content marked$_{NN}$ stimuli were originally used to transmit or, relatedly, what environment or evolutionary forces supported the emergence of this communicative complex.

There is certainly no shortage of hypothesized selective scenarios for the evolution of linguistic communication (see Számadó & Szathmáry, 2006). These include notoriously diverse scenarios including pair bonding (Deacon, 1998), grooming (Dunbar, 1998a), sexual (Miller, 2000) or social (Dessalles, 2007) display, confrontational scavenging (Bickerton, 2009), hunting (Számadó, 2010), and teaching (Laland, 2017a), *inter alia*. Fortunately, there are much fewer hypotheses about the more basic pragmatic capacities that enabled the emergence of linguistic, as well as non-linguistic, communication. As we saw in the previous chapter, non-linguistic communication itself already shows qualities that are most likely unique to humans. Thus, ostensive communication calls for evolutionary explanations that can account for this distinctive capacity. However, although the scientific study of language evolution has attracted interest at least since Darwin, only recently has pragmatics come to the fore.

In this chapter, I will review the major theories of the evolution of ostensive communication. An account of all or most of the selective scenarios for the evolution of full-fledged linguistic capacities is beyond the scope of the present thesis. But I will discuss some of the prominent accounts that may also be relevant to debates on the

origin of ostensive communication. I will leave the discussion of teaching and related kin-based scenarios for the next chapter, where I will argue extensively that pedagogical selection pressures, particularly those associated with the transmission of cognitively opaque technological and cultural knowledge, may provide the most convincing explanation for the evolution of human-specific communication.

3.2. Criteria for Assessing Selective Scenarios

The diversity of scenarios suggested for the evolution of human communication might lead one to conclude that the question is entirely unscientific and not worth pursuing. The absence of direct paleoanthropological evidence of communicative behavior strengthens such a view. However, like geology and cosmology, evolutionary enquiries can make learned inferences about the past based on behavioral, comparative, paleoarcheological, genetic, and neuroscientific evidence (Fitch, 2017b). Our understanding of general evolutionary processes is also conducive to better judgments about evolutionary history.

Thus, faced with the numerous hypotheses as to the evolutionary origin of language, scholars have compiled a list of criteria that provide a strict benchmark against which the plausibility of different accounts can be assessed (Bickerton, 2009; Laland, 2017a; Morrison, 2020; Számadó & Szathmáry, 2006). Although individually these are not difficult to account for, together they discard the majority of proposed hypotheses. These criteria include, among others: uniqueness, immediate utility, generality, honesty, and cooperativeness. While these criteria are suggested primarily for language evolution, with some modification, they can be applied to scenarios for ostensive communication too.

- *Uniqueness*: Perhaps most important of all criteria, any account of the evolution of ostensive communication should explain why it evolved in humans and no other animal (Hurford, 1998). Traits or adaptations as complex as eusociality or vision have convergently evolved several times in evolutionary history. However, ostensive communication has seemingly evolved only once. This demands a special explanation. Specifically, any selective scenario for the origin of ostensive communication must offer unique selection pressures that our hominin ancestors encountered but other animals, including non-human primates, did not. Of course, all communicative systems in nature are unique in some sense. Surely, vervet

monkey calls differ from the honeybee dance in interesting ways. Yet, our communication is *anomalous* in nature, in that its core features arguably do not resemble any other signaling system (Bickerton, 2009). Formal and informational open-endedness are among these features. This criterion alone refutes a majority of proposals, as they involve pressures that are observable in other animals too.

- *Immediate utility*: Any theory of the origin of both the language faculty and ostensive communication should show how the communicative system was useful or adaptive from the outset. As natural selection does not have foresight, it cannot first produce words and find a use for them only when the lexicon reaches a certain size. Thus, a protolanguage of, say, ten or fewer symbols should already have been able to convey useful and meaningful content (Bickerton, 2009). A related, crucial question is the evolutionary *bootstrapping problem* (Origgi & Sperber, 2000). Since languages are cultural artifacts rather than innate codes, learning them requires some kind of a Language Acquisition Device (Chomsky, 2006) or at least, as I will argue, a concept of communication. Such a learning device, however, is of no use in absence of an ambient language. And such a language cannot exist without individuals possessing the learning device. This chicken-and-egg problem creates an issue for most theories.
- *Generality*: Our communication is distinguished by its capacity to be used for multiple purposes. It enables, as we saw in the previous chapter, the production of an open-ended range of communicative means and their use for an open-ended range of contents. Another design feature of language and ostensive communication, which contributes to its power of generalization, is *displaced reference*, that is, the possibility to communicate about entities that are absent in space and/or time (Hockett, 1959). Displacement and immediate utility together pose a significant hurdle to selective scenarios: how can a communicative system be of immediate utility and about absent entities without presupposing an established code?
- *Honesty*: If in a communicative system a signaler can gain benefits by producing signals dishonestly, such signaling is likely to be selected. Receivers, on the other hand, may evolve to ignore the signals, for

responding to them becomes costly. As receivers stop responding to the signals, there is no payoff to producing them either. As a result, the communicative system collapses. For a communicative system to evolve, then, there must be mechanisms in place to ensure its evolutionary stability. Known as the problem of honesty or reliability, this is the central question in signaling theory (Maynard Smith & Harper, 2003; Scott-Phillips, 2011; Searcy & Nowicki, 2010). Multiple mechanisms are proposed for guaranteeing the reliability of communicative systems: *Handicaps* (Zahavi & Zahavi, 1999) are signals in which some strategic cost, paid by honest signalers, is tied to signal form, working against dishonest signaling. The peacock's tail is a famous example, where the longer tail reliably indicates that the peacock is fit enough to overcome the cost of carrying such a tail. In *indices*, the signal meaning is tied to the signal form and so the signal becomes unfakeable. A possible example is tigers' scratching of trees as high as they can, used as an index of their size (Maynard Smith & Harper, 2003). If there are *deterrents* in place, including punishment and reputational costs, honest communication can be (unlike in handicaps) relatively cost-free; in this case, the cost is paid instead by dishonest signalers (Scott-Phillips, 2014). Finally, signaler and receiver may have a *common interest* in the outcome of the interaction due, for instance, to genetic relatedness—low-cost signals may arise in such a condition (Maynard Smith & Harper, 2003). Any theory of the evolutionary origin of ostensive communication should be able to explain which of these mechanisms accounts for the stability of early human communication. The problem is exacerbated by the fact that in many instances of ostensive communication deceivers cannot be immediately detected due to a displaced content.

- *Cooperativeness*: In many contexts, humans communicate content that is beneficial to the receiver. Since communication could be costly, not least due to opportunity costs, there must be an explanation for why we communicate cooperatively. One should be cautious about what is meant by "cooperation" in communication. *Informative cooperation* (Scott-Phillips, 2014) refers to the honest use of communication, discussed above. Here, however, the concern is *material cooperation* where communication is benevolently used to benefit the receiver (Hurford, 2007; Scott-Phillips, 2014). Thus, the evolution of

> communicative systems generally necessitates informative cooperation but not material cooperation—communication can also evolve in contexts with a conflict of interests (Maynard Smith & Harper, 2003). The question is then why early human communicators cooperatively provided information to others (Laland, 2017a).

In light of these criteria, we can now evaluate the existing evolutionary theories of ostensive communication and, where applicable, see whether and how they manage to satisfy them. And, in the next chapter, I will argue that teaching may indeed be the most convincing scenario with respect to these criteria.

3.3. Relevance Theory

I discussed some of the central notions of relevance theory in the previous chapter and I will return to it in chapter 6, where I will talk about its representational requirements. To repeat, relevance theory offers, among other things, a cognitively more realistic account of Grice's main insights. Grice proposed that, to distinguish between natural and nonnatural meaning, we must look at the intentional structure with which nonnatural (i.e., communicative) meaning is produced. As formulated by advocates of relevance theory, this involves a metarepresentational intention directed at the belief state of an audience. Although Grice was more concerned with a philosophical conceptual analysis of communication and meaning than with cognitive processes, his approach seemed to imply that, in inferring meaning, a sophisticated rational reasoning is at work. Such reasoning may appear in conflict with the ease with which we communicate and the early development of the communicative system. For relevance theory, on the other hand, the job of processing communicative acts is done by a dedicated mindreading submodule and through "fast and frugal heuristics" (Sperber & Wilson, 2002). Moreover, the intended meaning is inferred not based on several principles and maxims, as suggested by Grice (1975), but mainly by a communicative principle of relevance according to which ostensive stimuli convey a presumption of their own optimal relevance (Wilson & Sperber, 2006).

Relevance theorists have had a major role in bringing pragmatics to the fore in evolutionary linguistics. Since relevance theory was largely a cognitive, rather than a philosophical, approach to communication from the outset, it has proven more amenable than most other pragmatic theories to evolutionary-psychological theorizing.

As an intentionalist account, that is, one that places the expression and attribution of intentions at the center of communication, it portrays the evolutionary trajectory of communication as involving ever more complex metarepresentational capacities until a tipping point was reached, where full-fledged ostensive communication was possible (Scott-Phillips, 2014; Sperber, 2000). This evolutionary process was largely driven by Machiavellian forces arising from the highly social life of primates (Sperber, 2000). Recently, however, the focus of the theory is shifting more towards the way in which ostensive behaviors manipulate attention (Scott-Phillips & Heintz, 2023; Sperber, 2019). Thus, although ostensive communication is possibly not a more sophisticated version of primate communication, there may be evolutionary continuity between the ways primates manipulate attention and the enhanced human-specific manipulation of attention toward informative intentions. As a result, great apes might possess the basic cognitive requirements of ostension, yet, due to the absence of similar selection pressures in evolutionary history, they do not typically develop human-like communication. These selection pressures were generated mostly by living in a partner choice social ecology (Heintz & Scott-Phillips, 2022).

Relevance theorists deny a narrow function for language use and communication. Just as the function of organs of locomotion is best understood as locomotion itself, rather than finding food or fleeing predators, the function of our language faculty should be seen, foremost, as enabling the acquisition of languages (Origgi & Sperber, 2000). Languages, then, add to the expressive power of ostensive communication, which could be used for gaining various benefits. Ostension serves a rather broad function of manipulating others' mental states, while the audience's inferential capacities serve the function of inferring mental states. Overall, then, ostensive communication evolved as an adaptation for social navigation (Scott-Phillips, 2014).

Relevance theory provides two complementary explanations for the uniqueness of ostensive communication. According to one explanation, humans' recursive mindreading has reached a level of sophistication that permits ostensive communication. This quality, which evolved independently of communication, is not present in other primates (Scott-Phillips, 2015). Therefore, ostensive communication is unique to humans partly because humans alone are capable of its representational demands. According to the other, compatible explanation, humans evolved ostensive communication in response to selection pressures arising from a unique ecology. This is a partner choice ecology (Heintz & Scott-Phillips, 2022). In this ecology, individuals

can find many opportunities for win-win cooperation. Thus, they seek partners for participating in mutually beneficial enterprises and behave such that their chances of being chosen as partners are enhanced.

Immediate utility and bootstrapping do not pose a problem for this theory because, thanks to a general communicative capacity, communication can be possible even before a conventional code is developed (Origgi & Sperber, 2000). Symbolic means of communication may have started off as icons, where there is a resemblance in meaning between the form and content, and eventually developed into full-blown conventional symbols without any form-content correspondence (Scott-Phillips, 2014). Regardless, by drawing attention to objects and events and the ability to attribute mental states, communicators could convey information open-endedly—going beyond the limited communication of non-human primates. In such a context, even the addition of one conventional symbol could augment the expressive power of the already functional communicative system.

The metarepresentational structure of human communication, as understood in relevance theory, can account for its generality. As mentioned above, there are constraints on the evolution of communicative systems that determine their honesty and stability. These constraints dictate that communication be tied to narrow domains of statistical mutual benefit. Otherwise, the system would collapse. The domain generality of human communication, however, appears to present an explanatory challenge. How can communication be open-ended and yet tied to a narrow domain of mutual benefit? Heintz and Scott-Phillips (2022) believe that metarepresentation in communication creates the possibility of *virtual* domain generality (Sperber, 2001). The actual domain of communication is narrow and specific: the informative intention. However, the informative intention can be about almost anything. Thus, communication can be open-ended while conforming to the usual evolutionary constraints. Although not elaborated in the theory, displaced reference, too, can perhaps be explained by the role of informative intentions. As the content of intentions can take anything as its referent, communicators can exploit that to inform about absent entities by providing evidence for their informative intention.

Since the question of honesty and stability is posed at the ultimate level of explanation, proximate explanations such as metarepresentations may not be the right solution. A partner choice ecology, however, can provide a more convincing solution to the problem. As it is in the interest of individuals in this ecology to be chosen as

partners for joint enterprises, what others think of them becomes crucial. Vigilance causes others to select partners based on information from past behavior (Heintz & Scott-Phillips, 2022). Therefore, one's reputation is vital for their success. Reputation can thus work as a deterrent against the drive for deception (Scott-Phillips, 2014). Hence, the honesty of human communication is guaranteed by the reputational costs incurred by liars. Living in this highly cooperative environment may also create situations of common interest in which communicators have little incentive to cheat.

Likewise, this ecology encourages cooperation between partners. When it is advantageous to take part in joint enterprises, individuals would also gain from informing others, since through expanding the common ground one can facilitate win-win cooperation. In such a context, interlocutors can rely on a presumption of cooperativeness to understand one another (Heintz & Scott-Phillips, 2022). By being relevant informers, cooperative partners enable many activities, like hunting and building shelters, that would otherwise be difficult or impossible. Moreover, through teaching opaque knowledge to partners and kin, communicators can gain adaptive benefits both directly and indirectly.

In chapter 6, I will present my arguments against intentionalist accounts of communication and explain why I think they cannot account for our concept of communication. Specifically, intention and belief attribution are insufficient for interpreting communicative, as opposed to goal-directed instrumental, actions—for example, because the epistemic effect of communication in others is unobservable for explaining the intention causing it. Instead, an explanation is required for informative behavior as perhaps an evolutionarily novel concept which is arguably present in humans but no other species. If so, then enhanced metarepresentational capacities of a mentalistic nature are unlikely to have led to humans' uniquely open-ended communication. Moreover, the same uniqueness criterion that is applied to evolutionary theories of communication may be required of theories of metarepresentation. Why did only humans acquire a representational theory of mind? A partner choice ecology could be part of the answer. However, this scenario raises a further question: what caused this ecology? One possibility is that a third factor, like technology, should be invoked to explain both the evolution of hyper-cooperativeness and ostensive communication (Mussavifard & Csibra, in press). Moreover, human-specific forms of cooperation are often enabled thanks to human communication (Brinck & Gärdenfors, 2003). Collaboration for a displaced goal, that is, a goal in the

future or distant in space, is made possible mostly through expression and coordination in communication (Geurts, 2022). If true, this creates issues for the idea that a cooperative, partner choice context drove the evolution of communication.

Since contemporary adult communicators use communication in all possible contexts, it might seem reasonable to assume that early communication was also a general-purpose system used with kin and non-kin for a variety of goals. However, just as we look for specific selection pressures leading to hominid bipedalism (say, carrying tools or walking on branches; Boyd & Silk, 2014), unique among primates, it makes sense to seek the original function of ostensive communication. After all, current utility may diverge from the initial biological function (Bateson & Laland, 2013). But even bipedalism is found in other taxa, if not in primates. Ostensive communication, however, has occurred only once in evolutionary history, despite the manifold apparent benefits. This fact prompts one to seek a unique selective scenario for this anomaly.

Finally, this account provides mostly negative explanations for why open-ended communication ("unleashed expression"; Heintz & Scott-Phillips, 2022) was evolutionarily and cognitively possible for humans. However, the removal of ecological and representational obstacles may not be enough for the emergence of a complex trait. We may additionally need a positive account that offers special selection pressures demanding the evolution of open-ended communication. More specifically, one must ask why marking behavior as communicative was required for humans. Or in other words, why did humans, but not other apes, need to generate novel or ad hoc communicative means that open-endedly conveyed novel messages? A partner choice ecology might guarantee the stability of ostensive communication, but, on the face of it, an enhanced form of the largely innate and limited communicative system of great apes (Byrne et al., 2017) could also serve the task of informing potential partners. An account of the evolution of human communication should ideally suggest a scenario in which the design features of ostensive communication are well suited to the function for which they were presumably selected.

3.4. Tomasello

In Tomasello's account (2008, 2019), human communication is a special case of human collaborative cooperation. His account is interesting in that he does not suggest a currently extinct form of communication that initiated the evolution of more sophisticated human communication. Rather, he sees the early developing

communicative behavior of children as reflecting, to a large extent, that of our hominin ancestors. These are our use of “natural” means of communication, namely pointing and pantomiming, which likely emerged before fully-fledged conventional languages. The skills that enable such communication include, on the one hand, cognitive mechanisms related to joint attention and joint intentionality, and, on the other hand, prosocial motivations for helping and sharing. From around the first birthday, humans start producing the pointing gesture with the goal of establishing joint attention to some entity. On the receiving side, moreover, humans make recursive inferences about what the communicator intends for them to think. Thus, as also suggested by relevance theorists, communication relies largely on mindreading abilities. At approximately the same time in development, infants also begin to join others in collaboration involving joint goals and coordinated plans. This suggests that, by design, communication rests on, and facilitates, mutualistic collaboration.

As an adaptation for cooperation, Tomasello (2008) maintains, human communication arose in the context of mutualistic collaborative activities where individuals helping others were also helping themselves. In these activities, multiple individuals pursue a joint goal together. Hence, at some point in our evolutionary history individuals who could take part in collaborative cooperation involving joint goals, joint attention, and cooperative motives were at an adaptive advantage. Our cooperative communication arose then to enable and facilitate these activities. Initially, the communicative system was possibly used only in such contexts and mainly for requesting and providing help. Communication is of course not specific to contexts of mutual benefit. Tomasello argues that both direct and indirect reciprocity can account for cooperative informing outside of mutualistic contexts. Like Heintz and Scott-Phillips, he notes that informing others prosocially enhances one’s reputation as a cooperator, so that others would wish to take them as partners for cooperation and offer them help in return. Finally, human-specific sharing of emotions and attitudes can be explained by the process of cultural group selection. Rather than transmitting information, such expressive use of communication serves an affiliative, group identity function for us which helps to expand common ground between individuals.

For Tomasello, the uniqueness of human communication seems to be explained partly by the cognitive and motivational resources that are available to humans but not other great apes. Non-human primates lack the skills to form joint goals with others in activities that involve shared intentionality. Furthermore, only humans, from early in

development, represent joint actions from a “bird’s-eye view”, in a format that includes both the goal and the complementary roles of participants. Humans’ mindreading capacities also far surpass that of other primates. These with the species-unique prosocial motivation for information provision enable our distinctive forms of communication. The underpinnings of our communication suggest that they might have evolved in a context where individuals showed more tolerance and generosity, and less competition with one another. The first step in this direction was taken when humans, as contrasted with other primates, became obligate collaborative foragers (Tomasello et al., 2012). According to a Stag Hunt model of human cooperation, individuals cooperate to gain a greater benefit compared to solo alternatives (the “Hare”). This is argued to have happened particularly in hunting where individuals had to cooperate using various coordination strategies.

Basing the evolution of human communication in non-linguistic, non-conventional (i.e., deictic and iconic) gestures, Tomasello’s theory also accounts for the immediate utility problem. Once the cognitive skills for this type of communication were in place, conventional communication could also emerge by piggybacking on the same infrastructure—replacing the “natural” feature of communicative behavior with a shared learning history. Iconic gestures are particularly important in this process as they, like conventional symbols, can refer to displaced entities. But since it is unlikely that individuals explicitly agree on conventions, the process should occur slowly and naturally. So, as individuals use iconic gestures to convey information, others possessing the capacity for simulation and role reversal imitation may start repeating that same gesture to solve a similar communicative problem. But the connection between the gesture and the meaning may be opaque to some observers. This might eventually lead to an insight that the relation between the signs and their referents can be arbitrary, thus permitting the emergence of devices entirely based on a shared history rather than an iconic character.

The generality of our communication, like other intentionalist accounts, could perhaps be explained by the reliance on inferring the intention of the communicator. Since intentions can be open-ended, communicative intentions can be too (although originally in a specific, collaborative domain). Moreover, as mentioned, displaced reference arose initially in iconic communication. Iconic communication, according to Tomasello, requires considering the communicative intention of the actor. Otherwise, a pantomiming gesture would be seen as an odd behavior. However, in ultra-helpful

contexts based on reciprocity, the recursive, cooperative reasoning needed for Gricean communication becomes possible. To enhance their reputation, individuals inform each other freely. And, to fulfil their desire, it is enough that they simply let others know of it. Thus, in hyper-cooperative environments, communication can, through the expression of communicative intentions, go beyond the here and now to refer to spatiotemporally displaced entities.

The cooperativeness and honesty of our communication are readily explained in this account. As the very act of communication is seen as an exercise in collaboration, evolved initially to enable mutually beneficial activities, it is no surprise that it is used cooperatively. Similarly, the honesty of early communication is explained by the lack of a conflict of interest in these mutualistic scenarios. Later, as in Heintz and Scott-Phillips' approach, reputational effects can guarantee the stability of ostensive communication beyond contexts of mutual benefit.

Perhaps the main question that Tomasello's evolutionary account leaves unanswered is also what caused the unique context in which human communication evolved. Granting that it originated in a mutualistic context, there must have been an ecology in which collaboration was especially advantageous to humans—an ecology that was not present for other primates. Again, even if such an ecology is ultimately spelled out (e.g., scarcity of non-meat nutrients), it must be one where open-ended communication, specific to humans, was required, that is, the limited but relatively flexible primate-typical communication was insufficient for the purpose. Furthermore, cooperative assumptions may not provide the best explanation to why other great apes do not communicate in human-like ways (contra Hare & Tomasello, 2004). While they fail to understand pointing the way human children do, they do communicate prosocially in other modalities (Crockford et al., 2012). Their failure might instead be because pointing is simply not part of their communicative repertoire. Thus, whereas they can learn to instrumentally use a pointing gesture to obtain food through humans, they do not appreciate the referential quality of pointing when addressed—despite otherwise being able to communicate prosocially. In sum, although my account also postulates a highly cooperative context for the emergence of communication (i.e., teaching), cooperation in itself does not seem to account for the peculiarities of ostensive communication.

3.5. Geurts

Geurts is an avid critic of standard intentionalist accounts of communication (Geurts, 2019a). Like many others, he points out that the evidence from developmental psychology does not seem to support the idea that communication is based on attribution of complex mental states. On the contrary, he believes that it is public communicative practices that enable mental state attribution both in development and evolution (Geurts, 2021). Thus, although mindreading contributes to communication, the explanation should be sought elsewhere. His main proposal is that, rather than expressing intentions, communication is a practice in commitment-sharing. According to him, making commitments is a form of managing expectations. It is a way of allowing others to rely on us so that they can coordinate their actions with ours. When someone is committed to someone else to act on a proposition *p*, she is committed to act in a way that is consistent with the truth of *p*. Therefore, when we communicate, rather than expressing our intentions, we share our commitment to act consistently with the truth of the proposition we are expressing. This permits us to coordinate our activities. Thus, communication is to be understood as a kind of coordinated action that is practiced for the purpose of action coordination (Geurts, 2019b). The reliance on commitment as a normative, relational notion may reduce the cognitive requirements for the development of communication, offering a perhaps more parsimonious account compared to theories that draw on individualistic notions of intentions and beliefs (Rakoczy & Behne, 2019).

Geurts' evolutionary account also derives from his commitment-based theory (Geurts, 2022). For him, one of the core functions of both chimpanzees' and humans' communicative systems is that they coordinate interactions between individuals. But, while chimpanzees communicate to coordinate their activities on the spot, humans can coordinate their future interactions as well. He mentions two main reasons for this. One is that chimpanzees are not sufficiently responsive, particularly on the receiving side. This is discernible in their inability to signal agreement in response to communication. Second, the agreement requires enduring commitments that chimpanzees do not appear to practice. Human conversations, in contrast, are marked by moment-by-moment collaborations between interlocutors in which they respond to one another's utterances by expressing various states including agreement—enabling future action coordination. To explain the transition from "chimp-style" communication to human-specific communication, Geurts puts forward four evolutionary stages. The

first stage is represented by the communication of modern chimpanzees, lacking the ability to coordinate future actions. In the second stage, however, agreement is enabled by simple means such as echoing. That is, just as humans commonly express agreement by simply repeating the speakers' utterances, hominins might have taken a similar, initial step towards responsive communication. But this likely supports coordination only over short time spans. An appointment, for example, may be possible only as long as interlocutors can see each other. To work over longer time spans, it must acquire a normative nature. This takes place at stage three. However, since most of our normative behavior is rooted in language, our pragmatic abilities should be explained evolutionarily via more basic, non-linguistic practices such as sanctioning. Sanctioning is argued to be an effective measure for upholding normativity, especially in small groups. Yet, sanctioning might not always be transparent, and it could moreover be harmful to social relationships in such groups. Thus, in stage four and with the development of linguistic devices, meta-talk (i.e., meta-linguistic reference to utterances) can be invoked to make and uphold commitments. For instance, reminding ("You said that…") is a useful tool for preserving various commitments including appointments. Therefore, through a co-evolutionary process, communication itself eventually becomes sophisticated enough to permit flexible ways of sharing commitments and carrying out joint, collaborative activities like group hunts.

Geurts' evolutionary theory is focused more on the processes rather than selection pressures that led to the evolution of our pragmatic abilities. Nonetheless, similarly to other collaborative accounts, his must also be complemented by offering unique conditions in which the design features of our communication were of adaptive value. Moreover, his normative approach seems to take for granted the symbolic and/or representational devices we use in communication. While it may explain our distinctive *uses* of communication, as compared with chimpanzees', the fact that this use is predicated on open-ended means cannot be neglected. Our collaborative and interactive conversations, targeted at joint communicative goals, utilize a wide range of flexible verbal and non-verbal devices—devices which should be the primary explanandum for an evolutionary theory of human communication. An account of how and why adult communicators use these devices in elaborate ways should arguably be the next step.

3.6. Incrementalist Accounts

While most theories suggest an incremental evolutionary trajectory for ostensive communication, there are accounts that emphasize this by proposing communicative systems of intermediate complexity which likely provided an evolutionary platform for the emergence of more modern forms. The presence of such intermediate systems can make the evolutionary scenario for human communication a more gradual one and, as a result, the question might become more tractable (Bar-On, 2013).

Moore's functional reading of Grice's formulation of meaning$_{NN}$ permits attributing rudimentary forms of ostensive communication to non-human primates. As his model of communication is discussed in chapters 2 and 6, here I will not get into its details. Nonetheless, if great apes are already Gricean or ostensive communicators, one does not need an extra account of why such communication is present in humans. Rather, one should offer an explanation of how and why ostensive communication became more sophisticated to reach its human level of complexity. Here, Moore (2017) thinks, Tomasello's Stag Hunt scenario may be relevant. He believes that, while great apes are ostensive communicators capable of expressing and inferring communicative intentions, they have not evolved human-like linguistic communication because they do not depend, as humans did, on communication for foraging. Supporting this claim, chimpanzees do not use their communication in collaborative tasks, although they appear to possess the basic skills. The reason humans started using communication in more collaborative ways was a consequence of their reliance on meat for nutrition. Since communication in collaborative contexts seems to require more creative use and an enhanced social attention, chimpanzees' limitations might be due to their less developed general intelligence. As brains are metabolically expensive to maintain, there was selection pressure for more effective foraging strategies. Therefore, collaborative hunting and the meat it made available provided the energy for larger brains. Then, rather than social pressures arising from group living, domain general inferential abilities enabled by a rich diet resulted in more human-like ostensive communication.

Planer (2017a, 2017b), however, does not believe that great apes possess the representational requirements of ostensive communication. Indeed, he suggests that ostensive communication, as described in standard approaches, is cognitively too demanding to explain the emergence of human communication and language. Instead, he proposes that early Pleistocene hominins might have had a protolanguage

before any ostensive communication. Rather than requiring a full-blown belief-desire psychology, this protolanguage could have gotten off the ground with a simpler cognitive package. These include: an informational theory of mind, without false-belief understanding, enabling the attribution of goal states and information-carrying states, and an understanding of how these states can interact to generate behavior; mental model construction involving the ability to bring together a variety of information into a single mental representation of a situation; and the ability to imitate interpersonal acts. Together these traits can make possible the production and interpretation of an expandable lexicon. Evidence from Pleistocene stone toolmaking shows that early hominins may have possessed the requisite cognitive abilities and used a protolanguage to transmit their technological skills to others (Planer, 2017b).

Sterelny (2017), too, rejects the sharp distinction, maintained by relevance theorists, between the coded signals of non-human animals and the ostensive-inferential system of humans. As seen in the gestural communication of non-human primates, with the cognitive sophistication of agents a code can be productively expanded. According to Sterelny, iconic gestures might have been one way that our repertoire was augmented without requiring the full Gricean higher-order intentions. With some theory of mind capacity, causal reasoning, and executive control over motor routines, early hominins might have mimed, say, hunting to recruit others for hunting. Here, the resemblance between the gesture and its meaning could have supported signal comprehension.

Bar-On (2013, 2017) similarly postulates that the dichotomy between reflexive animal signals conveying information about biologically significant attributes of the producer, on the one hand, and humans' communication relying on reflective, metarepresentational intentions, on the other, may not be exhaustive. She argues, instead, that there is a third class of communicative signals, namely expressive behaviors, which lies between the two extremes (see also Green, 2019). This kind of communication, like other animal signals, does not arise from intentions to inform or draw attention to things in the environment. However, like the expression of intentions in Gricean communication, expressive signals reveal the animal's psychological state, while at the same time drawing attention to some object or event at which the state is directed. These signals include vocalizations such as alarm and distress calls as well as non-vocal behaviors such as teeth-baring, tail-wagging, and food-begging gestures. Thus, for instance, a vervet alarm call indicates both the presence of a

predator and the monkey's fear of the predator, prompting others to hide from the danger. Since explaining the emergence of Gricean intentions is no less difficult than explaining linguistic communication itself, focusing on the less demanding expressive signals can make the task easier by positing that, before language, evolution had already solved the problem of information transmission. This way, theories of language evolution can concentrate, instead, on the more tractable problem of how linguistic expressive vehicles replaced or altered animal non-linguistic expressive behaviors.

As most of these incrementalist theories do not offer specific selective scenarios, they cannot be assessed against the five criteria. Nevertheless, the mere existence of intermediate cases between animal signals and full-fledged ostensive communication suggests that the problem for evolutionary accounts may be less critical than it is otherwise assumed. Uniqueness, for example, would be less of an issue, for ostensive communication is a sophistication of simpler types of communication—so, uniqueness becomes a question of degree rather than kind. The same goes for immediate utility, since human communication originates from other, already functional systems.

With respect to incrementalism, it should be noted that gradual evolution does not necessitate the emergence of biological traits from earlier traits with a similar function: gradualism applies to biological form rather than function (Scott-Phillips, 2015). So, while the human communicative system evolved plausibly on the back of ancestral cognitive and behavioral mechanisms, it could have acquired a communicative function relatively recently. There is thus no a priori reason to assume that it evolved out of primate-like communication. Indeed, as I will argue in chapter 5, postulating that ostensive communication emerged from animal signals, complicates, rather than simplifies, the hypothesized evolutionary trajectory.

3.7. Language for Food

Besides theories dealing with the emergence of ostensive communication as a whole, there are those that focus narrowly on the evolution of the language faculty. Prominent among these are theories that suggest language evolved as a means of recruitment for obtaining food, either through confrontational scavenging or group hunting.

According to Bickerton, from around 2 Ma (million years ago), hominins occupied a new niche that lasted for several hundred thousand years (Bickerton, 2009; Bickerton & Szathmáry, 2011; B. Clark, 2011). In east Africa, the climate had become drier and more variable. This resulted in vast areas of savannah that lacked the traditional

primate food sources (e.g., fruits). However, scavenged meat from herbivorous megafauna inhabiting this environment offered a highly calorific substitute. But this came at a cost. The process of penetrating the hides of the large carcasses could take up to several hours accompanied by attacks from rival species. Thus, solitary scavenging was not an option. Rather, large groups had to cooperate both to butcher the carcass and fend off competing scavengers. Although more dangerous than other foraging methods, cooperative defense assisted by stone projectiles rendered confrontational scavenging a fruitful strategy. Such cooperation necessitated the recruitment of non-kin and convincing them to leave other activities for a potential reward outside their sensory range. These recruitment signals characterized the first step towards human language.

Számadó's (2010) hunting scenario is very similar to Bickerton's account. Also emphasizing the centrality of meat to the diet of human ancestors, he interprets the abundance of stone tools as indicating regular access to large carcasses. Nevertheless, he finds hunting to create stronger pressure for novel skills (e.g., tool making and long-distance running) for obtaining food compared to scavenging. Hunting large, prime-adult prey required unprecedented cooperation and coordination between a large number of individuals and new levels of technological advancement. This was possible only through novel communicative means. Thus, the first proto-linguistic signals served the function of pre-hunt recruitment. This communicative system developed further for coordination of hunting roles and transfer of hunting know-how. The need to plan the hunt and negotiate tactics provided extra pressure for complexity of the system.

Both accounts offer explicit answers to the five evolutionary criteria. The uniqueness of linguistic communication is suggested to be explained by the unique dependence on recruitment signals for subsistence. Recruiting others for procuring food that is not currently visible presents new pressures, seen perhaps only among bees and ants. Cooperation and coordinating roles, moreover, require novel communicative means. Other great apes do not rely to this extent on meat. Consequently, the problems associated with hunting or aggressive scavenging do not arise for them. Chimpanzees hunt only occasionally and possibly without explicit coordination. Nor do they require cutting tools for their small-sized prey which can be torn apart. There exist pre-hunt calls among other animals such as wolves and hunting dogs. But, unlike humans,

these are specialized hunters which do not specify in their communication the type of prey or hunting roles (Számadó, 2010).

Recruitment signals are also of immediate utility. Even one word or gesture would be useful, say, for informing about the presence of prey or a carcass. Both theories, like many others, agree that iconic signals might have emerged before arbitrary signals. For example, by mimicking prominent features of the dead animal such as its shape or sound, or by drawing the prey, individuals could convey the species of the spotted animal (Számadó, 2010). Számadó thinks, however, that some form of indexical communication, with a causal relation to the referent, might have preceded iconic communication. Just as some bees bring to the hive the odor of the flower they have encountered, so too hominins could use the body parts of the prey species for recruitment.

Recruitment signals are also inherently displaced from what they denote. Informing about an animal seen some time ago at a long distance was possibly the first clear case of displacement outside the insect order *Hymenoptera*, and the first major step toward language. Coordinating the group and transmitting knowledge later provided opportunities for innovation in communication—accounting thus for its generality.

As in Tomasello's account, the shared interest of the group members prevents them from lying. Misleading information undermines the success of the hunt or scavenging, an outcome which is not desirable for any of the members. Moreover, recruitment involves relatively rapid confirmation (Bickerton, 2009). The promised prey or carcass is either present or not, with potential consequences for the informer. Cooperation too contributes to the mutual benefit of the individuals involved, which can act as a positive incentive. The foraging is successful only if all members cooperate. Otherwise, they all lose. This leads to a coevolutionary process between communication and cooperation: the more cooperation, the more language required for enacting it; and the more language, the more opportunities for cooperation.

Three points should be made about these scenarios. One is that, while iconic communication may facilitate interpreting the meaning of communication in absence of a conventional code, it does not solve the problem of signalhood; that is, iconic devices are not transparently communicative. Therefore, although it might have preceded other types of communication, its use cannot be taken for granted. Rather, it is founded upon a (not necessarily simpler) system of communication that marks$_{NN}$ actions and uses them as representational media for conveying information—a

conceptual underpinning that is likely absent in non-human primates. It thus already involves a leap from other systems, demanding explanation. Secondly and relatedly, the emergence of these signals requires grounding them to their referent. This cannot happen since the context is, as mentioned, inherently displaced from the referent. So, although over long evolutionary time natural selection can create such signals (like the bee dance) in the innate repertoire of a species, it is unclear how this can happen in the case of a flexible iconic system. The signals had to somehow be immediately useful and yet conjure up an absent referent. The first individuals using iconic recruitment signals could not achieve this unless the conceptual requirements of the system were already in place through other evolutionary forces. Thirdly, both theories use evidence from stone toolmaking to support the need for procuring meat, either through hunting or scavenging. However, as I will argue in the next chapter, this technology is itself a unique human feat whose propagation in an artifact culture might depend on productive communication. The same goes for hunting know-how which implies a rich cultural knowledge acquired via teaching through various means including communication (Lombard, 2015).

3.8. Social Functions

Another class of theories imply that, rather than providing directly beneficial semantic or propositional content (Pinker & Bloom, 1990), language evolved due to other social functions. That is, these theories propose functions that, at least initially, have more to do with achieving a sociopolitical purpose than, as other theories suggest, vital information independently contributing to survival.

In Dunbar's (1997, 1998a, 2017) view, language evolved as a social bonding mechanism. According to the *social brain hypothesis* (Dunbar, 1998b), the evolution of the primate brain is primarily due to pressures for solving social problems (but see DeCasien et al., 2017). Support for this comes from the finding that neocortex volume is correlated with group size in primates. Thus, primate groups, expanded due to ecological pressures, imposed information-processing demands for creating and maintaining such groups, leading ultimately to larger brains. The primary means of bonding in these groups is through grooming, so that grooming cliques serve as alliances mitigating the cost of primates' highly social life. As the size of these cliques is predictable by the size of the species' neocortex volume, it appears that humans must have been adapted to group sizes of 150 individuals. This number, Dunbar

argues, corresponds aptly to the number of people each individual knows well. It is also a common number for villages and communities in traditional societies. However, this is three times larger than the largest group sizes observed in other primates. Bonding the groups using the same grooming method, then, would exert a pressure on the time budget of humans. This would take 40% of total day time—much more than any extant primate spends on grooming. Therefore, to manage their livelihood, our ancestors had to evolve a more efficient bonding mechanism. Here, spoken language comes into the picture. While grooming is a one-on-one interaction, conversations are often held simultaneously between several people. Moreover, by allowing the spread of social information, language increases the efficiency of social interactions—unavailable to primates relying solely on direct observation. Studies show that indeed the majority of conversations revolve around social topics such as personal experiences and relationships of both the interlocutors and third parties—rather than, say, technological topics. All in all, language evolved as a form of "grooming-at-a-distance".

Dunbar's theory suggests unique conditions for the evolution of language, but these do not seem to explain why "traditional" animal signals could not have resolved the bonding problem—as seen, for example, in gelada vocalizations (Gustison et al., 2019). Indeed, as Dunbar notes himself, initially a more primate-like communicative system such as laughing could have served the grooming function just as well. Singing, moreover, could have enabled bonding between larger groups (see also Savage et al., 2021). Both laughter and singing trigger the endorphin mechanism similarly to grooming, providing an incentive for the individuals involved. It is, furthermore, not evident how the social gossiping function could be immediately useful (see also Számadó & Szathmáry, 2006). Gossip seems to require some sophistication of language before it can be sufficiently meaningful. Such a language, however, presents a bootstrapping problem. How could individuals without a language faculty create and learn this protolanguage? And although social topics are often displaced, they seem to presuppose conventions established prior to their use in conversation. Without a basic communicative system already in place, this is hard to spell out. Overall, then, Dunbar's theory appears more successful in accounting for the spoken modality and its articulatory development than its productive amodal features.

Dessalles' account (2007, 2014) may be seen as complementary to Dunbar's for its emphasis on the role of language in establishing social relations. Yet, he proposes a

scenario in which productive communication confers adaptive value. He points out that examination of human verbal behavior does not demonstrate the cooperative features that most theories attribute to it. Rather than involving cautious cooperation where valuable information is offered parsimoniously as a gift, humans communicate readily and extensively. They spend one third of their waking hours on verbal activities and struggle to hold the floor in conversations (see also Miller, 2000). So, contrary to cooperative models, human communication exhibits features of competitive advertising (see also Locke & Bogin, 2006). Regarding casual conversation as the main context of communication, he considers two primary modes of verbal activity: conversational narrative and argumentative discussion. During narratives, speakers draw attention to current or past maximally unexpected events. During discussion, on the other hand, speakers indicate inconsistencies or suggest ways of increasing logical consistency. Thus, debatable issues concern contradictory situations that need logical solutions. Besides being very talkative, another feature of human communicators is that they speak about futile matters that do not seem to have any serious consequences for their survival. This begs the question: why do humans communicate, whether in narrative or argumentative form, if not to convey vital information? According to Dessalles, humans choose their friends based on their conversational skills. If we see verbal behavior from this perspective, then more important than the content conveyed in utterances is the quality signaled by them. By being relevant, speakers advertise their ability to detect unexpectedness and inconsistency. For this purpose, even futile topics are adequate.

Despite their dependence on alliances, the reproductive success of chimpanzees is strongly correlated with their physical strength and their ability to exert their supremacy through physical coercion. But with the advent of weapons among our ancestors, easy killing became possible. This had immediate consequences for the preexisting social order: as top-ranking individuals became the target of subordinates, an inverted hierarchy was established in which individuals submit to the group and refrain from dominance (see also Whiten & Erdal, 2012). In this context, criteria for the selection of allies changes too. Here, information replaces muscle as the primary social asset. Since easy killing by ambush is a possibility, you are better off choosing your friends based on their capacity to anticipate danger and their willingness to share it with you. Unexpectedness being the most reliable indicator of such dangers, you seek allies who are adroit at this, and you try to advertise your own ability as well.

Thus, even futile events can be used in communication to display the sensitivity to spot unexpectedness. Moreover, those who can find logical inconsistency in others' testimonies and can demonstrate this in their speech are unlikely to be deceived or taken by surprise. Argumentative discussions are then displays of this quality in order to be chosen as allies. Hence, the uniqueness of human language is explainable by the species-unique reliance on the information-processing skills of individuals for their selection as allies.

Dessalles argues that the display function permits the immediate utility of even one word or gesture. Citing developmental studies (Carpenter et al., 1998), he sees infants' pointing to unexpected things as already revealing their propensity to refer to unexpected events in narratives. A similar phase might have occurred in evolution, providing the foundation for more advanced ways of communicating unexpectedness. Reference to displaced or abstract entities in discussions is also useful in pointing out logical abilities. Although, as Dessalles contends, information-transmission may be explained in this scenario by the selfish need of individuals to be chosen as allies, there is room for conflict of interests between the communicator and addressee. This possibility together with the advertising nature of his account make costly handicaps or unfakeable indices better candidates for explaining the reliability of the signal. But, while such qualities may be observed in language *use* (e.g., verbal competence, vocabulary breadth, being to the point, etc.) they may not explain the building blocks of human communication, that is, the productive generation of words, gestures, and ad hoc means. Were the early one-word utterances honest because they were impossible or hard to fake? Given the negligible cost of speech (Fitch, 2007), this sounds implausible. Thus, the theory seems to presuppose the existence of communicative means that were later used in a costly or unfakeable fashion to display the cognitive capacity of communicators. This is also true for theories based on sexual display of fitness (Locke & Bogin, 2006; Miller, 2000). Besides, even if the utterance of words is somehow tied to costs or causally constrained, selection need not favor truthful reality-grounded communication, since a fictional topic may display the requisite qualities just as well. Eventually, the language as a world-directed communicative system may collapse, making room for other means of advertising one's cognitive skills.

Although the ethological methods used both by Dunbar and Dessalles are useful in understanding verbal behavior, the focus on how language is used most of the time

may not necessarily provide the best measure for deciding its original function. Our communication has certainly gone through several stages in its evolutionary history to acquire new functions. Its use in bonding and coalition building might have been among relatively recent developments. Moreover, if a signal such as an alarm call saves you only once from fatal danger, this may outweigh any less significant, if frequent, uses. Lastly, whereas narratives taking up around half of our linguistic activities appear to entail immediately futile matters, they may serve a broader function: as observed among hunter-gatherers, narratives can be used pedagogically to transmit social and ecological knowledge to novices (Scalise Sugiyama, 2021).

The last scenario to consider is Deacon's (1998) suggestion that the first symbols were established for enforcing primitive forms of marriage agreement. He notes that, while the benefits of symbolic communication might appear obvious to us, the earliest symbols were likely not nearly as flexible as modern ones. The question is then why such inefficient symbols were needed on top of the already functional primate communicative systems. His answer is that there must have been a shift in communicative strategy, rather than a mere improvement on existing signals. This shift was driven by a unique organization of the mating relationship, necessitated by a unique foraging style. Contrary to most mammals, caring for offspring in humans involves contribution from both sexes, leading to the formation of cooperative pair bonds between the sexes. Notwithstanding, the interests of the pair are not entirely symmetric. The male can improve his reproduction by copulating with other females, leaving the resulting offspring to be raised by another male. And the female can copulate with several males to maximize care giving from them. In mammals with a similar social arrangement, pair bonds often isolate themselves from other conspecifics to maintain exclusive sexual access. However, the problem is amplified in humans due to group living and the dependence on hunting for meat—which demands the males to leave care-giving females behind. This rare combination of male provisioning and social cooperation creates a potentially highly volatile system. The evidence from tool-assisted access to meat and the reduction in sexual dimorphism suggests that this system was likely in place with the emergence of the genus *Homo*. Thus, females depended on male provision of meat for themselves and their children, while males needed guarantees that they were provisioning their own offspring. How was such a system stabilized? According to Deacon, there must have been ways to mark exclusive sexual relationships. This came in the form of a symbolic

marriage agreement. As pair bonding in humans is a prescription or promise directed at future activities, Deacon's account, like Geurts', is rooted in normative uses of communication. Such social contracts do not only determine the obligations in the relationship between the productive pair, but also those between the broader community and the pair. The tool-dependent hunting and provisioning subsistence strategy of hominins was then responsible for the emergence of a symbolic regulation of reproductive relationships. These reproductive social contracts were established and maintained through ritualized construction processes (see also Knight, 1998).

In Deacon's account, tool-dependent, cooperative procurement of meat provided the unique conditions for the emergence of the first symbols. Although these are suggested to be useful without the need for an extensive "lexicon", it is not clear how they could have been effective in enforcing the associated norms without language (see § 3.5) and crucially how they would have developed into ordinary means of communication. Furthermore, despite being inherently displaced because of their future-oriented nature, they do not seem to explain the productive, open-ended feature of language or communication. Hence, as noted by Számadó and Szathmáry (2006), the theory fails to meet most of the criteria for selective scenarios. Both Dessalles' and Deacon's accounts, like many others, presuppose the existence of advanced tools—namely tools used in confrontation with conspecifics or for obtaining meat. Since these tools were unique to humans, one must also clarify why they were available to humans but not other primates. Otherwise, both theories may be right in how the use of technology could have established novel, more egalitarian social structures and driven the evolution of ever more complex communication. Overall, theories that suggest functions which require the productive transmission of propositional content might be more successful in explaining the design features of language and ostensive communication than those relying on less direct sociopolitical benefits involved in novel ways of interacting with conspecifics.

3.9. Conclusion

The above discussion of various selective scenarios, although not exhaustive, might already show that, thanks to the limiting criteria and our understanding of evolutionary processes, theories of the origin of human communication are not to be seen as equally plausible. This hopefully helps us to go beyond mere storytelling in our evolutionary accounts and, with a better grasp of how ostensive communication works

and develops, it might in the future yield fairly good approximations of how the system actually evolved.

One possibility not explored above is of course that the cognitive underpinnings of language and ostensive communication emerged, not as adaptations, but as spandrels or exaptations of other cognitive traits (Hauser et al., 2002; Reboul, 2015). But, while many if not most features of our communicative system may derive as by-products of some earlier non-communicative traits, their distinctive use in communication requires a no less elaborate explanation, for other species, despite their rich conceptual life, do not show similar capacities in their signaling (Fitch, 2017a).

A few generalizations can be made based on the foregoing discussions. First, theories based on basic communicative skills seem to fare better than language-based accounts, due to the difficulties associated with the establishment of conventions and evolutionary bootstrapping. Secondly, the majority of these accounts propose an episodic use for early communication, such as organizing and managing ongoing collaborations, rather than a context in which communication of generic contents is useful. This can be contrasted with pedagogical scenarios, discussed in the next chapter. Thirdly, as mentioned, human-specific tools are often used as evidence for the hypothesized scenario. However, this might call for explanation itself and cannot be readily used as a premise in evolutionary accounts. Fourthly, most of the pragmatic accounts are based on mutualistic, collaborative contexts. As perhaps the strongest alternative, this can be pitted against the hypothesis that ostensive communication evolved primarily for teaching generic knowledge—an inherently asymmetric phenomenon. I will next argue that the latter may provide a stronger case for the emergence of our communicative system and its unique features.

Chapter 4. The Evolutionary Origin of Ostensive Communication in Teaching

4.1. Introduction

In this chapter, I discuss the proposal that ostensive communication evolved primarily to facilitate and enable the pedagogical transfer of technological knowledge from caregivers to children. In light of my previous discussion on the core features of ostensive communication, the main focus will be on whether this selective scenario can positively explain open-endedness. Of course, language and ostensive communication are flexible general-purpose systems that can be applied in various domains. Their selective advantage for teaching or indeed any other purpose is thus not in question. To choose between the multiple proposals, however, one must show that the scenario best characterizes the distinctive properties of the communicative system as compared with the alternatives.

Thus, after placing human teaching within the wider research on the evolution of teaching in animals, I will provide my arguments in favor of the hypothesis that ostensive communication emerged to enable pedagogy, and I will examine whether it can satisfy the five criteria introduced in the previous chapter. I will then address potential objections to the hypothesis, how it relates to other accounts, and the subsequent expansion of ostensive communication across other domains of adaptive value. The arguments in this chapter target the original *function* of ostensive communication. How this function is proximately achieved and the general representational requirements of ostensive communication will be discussed, respectively, in the next two chapters.

4.2. Teaching in Animals

Although teaching as an institutional practice is a modern cultural innovation of humans, understood in functional evolutionary terms it is attributed to a wide range of animal taxa. The biological interest in teaching is largely thanks to Caro and Hauser's (1992) seminal paper. Distinguishing it from other forms of social learning, they offer an operational definition of teaching that emphasizes the active participation of the source of information: "An individual actor **A** can be said to teach if it modifies its behavior only in the presence of a naive observer, **B**, at some cost or at least without

obtaining an immediate benefit for itself. **A**'s behavior thereby encourages or punishes **B**'s behavior, or provides **B** with experience, or sets an example for **B**. As a result, **B** acquires knowledge or learns a skill earlier in life or more rapidly or efficiently than it might otherwise do, or that it would not learn at all" (p. 153). Various modifications have been suggested for this definition. Thornton and Raihani (2008), for example, argue that it should be amended with the conditions that the behavior be cooperative and involve coordinated interaction between the donor and receiver of information. However, such operational definitions should be supplemented with conceptual definitions highlighting the adaptive value of teaching for transmitting knowledge and skills (Hoppitt et al., 2008). This is important to rule out cases that superficially resemble teaching but serve different purposes. For instance, one species of ant uses a technique known as tandem running to lead naïve nest-mates to a source of food—a behavior seen as an example of teaching (Franks & Richardson, 2006). Nonetheless, despite fulfilling the conditions of the operational definition, the episodic ("here-and-now") nature of the information provided makes it a good case of communication, as defined in the previous chapter, rather than teaching (Csibra, 2007b; Leadbeater et al., 2006). A good conceptual account of teaching might be to characterize it as behavior with the function of facilitating long-term learning in others (see also Kline, 2015). Emphasizing function rules out cases where the behavior is explicable in terms of some local benefit (e.g., food provisioning) that may also result in learning as a by-product; and emphasizing long-term learning, as opposed to mere transfer of information, excludes cases of non-pedagogical signaling.

Several classifications of teaching behavior have been put forward. Caro and Hauser distinguish between *opportunity teaching*, where the teacher puts the pupil in a situation that is conducive to acquiring skills or knowledge, and *coaching*, in which the pupil's behavior is altered through encouragement or punishment. Another distinction is made between *progressive* and *fixed* teaching (Thornton & Raihani, 2008). The former is aimed at the acquisition of procedural knowledge for developing skills (e.g., hunting). Procedural knowledge requires the costly investment of teachers in modifying their behavior in accordance with the pupils' stage of development. Fixed teaching, however, is targeted at the acquisition of declarative knowledge. This has a binary nature, as the pupil either knows or does not know a fact. Still more extensive classifications have been offered (Hoppitt et al., 2008; Kline, 2015). Kline, for instance, proposes five separate classes: teaching by social tolerance, opportunity provisioning,

stimulus enhancement, evaluative feedback, and direct active teaching. Of these, she believes, the latter type occurs only in humans.

Despite the benefits to both pupils and teachers, teaching is relatively rare in non-human animals. This is because the high costs of teaching permit its evolution only in specific conditions. Teaching can evolve where the costs of teaching are outweighed by the benefits gained as a result of learning in pupils (Caro & Hauser, 1992; Fogarty et al., 2011; Thornton & Raihani, 2008). This may happen more directly when teaching leads to a reduction in the period of offspring dependence and its provisioning demands (Hoppitt et al., 2008; Thornton & Raihani, 2008). The benefits can also accrue from the contribution to the teacher's inclusive fitness through increased offspring survival. Teaching is, therefore, most common between closely related individuals (Thornton & Raihani, 2008). The utility of teaching is highest when there are few opportunities for individual and inadvertent social learning or when the costs of learning are particularly high. These costs arise, for instance, due to a need for specialized skills or dependence on dangerous prey (Thornton & Raihani, 2008). Teaching seems to be disproportionately observed in cooperative breeders including humans (Burkart et al., 2009; Hoppitt et al., 2008; Kline, 2015). This breeding system involves care and provisioning from alloparents, that is, individuals other than the biological parents. Cooperative breeding is believed to foster prosocial attitudes that range from donation of food to donation of information in teaching (Burkart et al., 2009).

The most convincing cases of teaching are seen in predatory species. For example, meerkat, a cooperative breeder, appears to provide opportunity for learning by bringing live prey to pups. As hunting scorpions requires skills that the young pups lack, adults gradually introduce the pups to scorpions. The scorpions are incapacitated to various degrees based on the developmental stage of the pups. This modification of prey depends not on direct assessment of the pups' skill level but rather on the change in their begging calls (Thornton & McAuliffe, 2006). Besides the dangers involved in hunting scorpions, the pups require teaching because adults consume the scorpions rapidly—making inadvertent social learning impossible (Hoppitt et al., 2008).

Although tool use is recorded in a diversity of taxa (Biro et al., 2013; Seed & Byrne, 2010), it is noteworthy that teaching is often not needed in the acquisition of tool use or toolmaking (e.g., Lonsdorf, 2006). Despite some anecdotal evidence (Boesch,

1991), great apes do not seem to regularly facilitate tool use in their young, or the purported teaching behavior can be explained in terms of local benefits such as cooperative transfer of tools (e.g., Musgrave et al., 2016). Teaching is generally rare among non-human primates. This is likely due to the abundance of opportunities for individual and social learning—rendering active intervention unnecessary (Byrne & Rapaport, 2011; Hoppitt et al., 2008; Thornton & Raihani, 2008). Thus, humans have evolved various teaching mechanisms despite, rather than because of, having other learning means at their disposal (Fogarty et al., 2011).

4.3. Teaching in Humans

Although local traditions that may be classed as culture have been observed in various species, *cumulative culture* is believed to be absent or rare in non-human animals (Dean et al., 2014). Cumulative culture refers to the modification, over multiple transmission episodes, of socially transmitted behavior patterns, leading to an increase in the complexity or efficiency of those patterns. This accumulation of modifications is known as the "ratchet effect" (Tennie et al., 2009). Humans' livelihood depends to a large extent on cumulative culture (R. Boyd et al., 2011). Possibly due to environmental fluctuations and changes in the rearing system, humans developed systems of knowledge involving both sophisticated technology and non-material cultural know-how (Henrich & McElreath, 2003; van Schaik et al., 2019). The later geographical expansion of humans out of Africa also required the creation of novel tools, knowledge, and social arrangements. This heavy reliance on cumulative culture could mean that human subsistence is possible only through the use of complex skills and technology that no single individual can invent during their lifetime (Fogarty et al., 2011). Despite their intelligence, then, humans often do not understand why (or whether) certain behaviors, tools, and social norms are necessary for their survival. Thus, humans seem to have evolved in a "cultural niche" (R. Boyd et al., 2011). This mode of survival demands especial adaptations for the efficient acquisition of cultural information. These adaptations include imitation as well as various content and context biases (e.g., imitating successful people), enabling the selective learning of fitness-relevant knowledge (Henrich & McElreath, 2003).

Human teaching is among these adaptations. In addition to other teaching methods, humans distinctively use active teaching to transfer knowledge to their children. Direct active teaching is useful when the content to be conveyed presents a

“frame problem”, that is, the relevant aspects of the behavior are not evident to the pupil. In this case, the teacher makes the relevant information manifest through intervention or communication, and the pupil interprets it as knowledge content (Kline, 2015). The reason this kind of teaching is seen only in humans is suggested to be that ours is the only species in which it would be adaptive. This is in turn because our evolution in the cultural niche involves traits too complex for any individual to develop on their own. Hence, cumulative culture creates a suitable environment for the selection of efficient and flexible teaching (Burdett et al., 2018). Studies modeling the evolution of teaching have shown that, when cumulative culture results in a high frequency of difficult-to-acquire information, faithful means of knowledge transmission are promoted (Castro & Toro, 2014; Fogarty et al., 2011). Simple mechanisms to approve or disapprove of the learned behavior can greatly enhance the reliability and accuracy of imitation (Castro & Toro, 2014). Moreover, teaching and culture reinforce each other, as teaching is more adaptive in a cultural context and cumulative culture often depends on teaching (Lucas et al., 2020). Teaching also fosters the retention and transmission of innovation through preferential transmission from innovators to pupils (Castro & Toro, 2014; van Schaik et al., 2019). In addition, the exceptionally diverse teaching toolkit that humans have at their disposal enables teachers to select their pedagogical strategy based on whether higher or lower levels of fidelity are needed—permitting various degrees of flexibility and innovation in the development of skills (Burdett et al., 2018). In sum, teaching supports efficient, high-fidelity cultural transmission, and it may be necessary for the development of complex knowledge and technology, suggesting that teaching and cumulative culture coevolved in our lineage.

Teaching seems to be a universal phenomenon in human societies. Contemporary hunter-gatherers, whose lifeway is most representative of the conditions in which humans evolved, teach to a much lesser extent than industrialized societies, relying mostly on individual and observational learning. Moreover, they may not engage in institutional or culturally explicit methods of teaching. Nonetheless, they do practice subtle ways of teaching to transmit knowledge (Boyette & Hewlett, 2018; Hewlett & Roulette, 2016). As also suggested by theoretical and modeling studies cited above, hunter-gatherers use teaching mainly in domains that are difficult to acquire individually or critical for survival. These include social norms and rituals, as well as ecological knowledge and subsistence skills such as digging for roots, basket making, and the use of spears for hunting. Furthermore, teaching is believed to be a specialized

natural cognitive ability, developing reliably without any conscious effort or awareness of its underlying logic (Strauss et al., 2014; Strauss & Ziv, 2012). From 3 years of age and perhaps earlier in infancy, children use means such as demonstration to resolve a perceived knowledge gap in others. And on the receiving side, infants seem well-adapted to acquire knowledge from teaching episodes, making use of diverse learning mechanisms (Skerry et al., 2013). All in all, teaching is practiced much more ubiquitously and flexibly in humans than in other species and is used particularly for conveying complex cultural and technological knowledge that individuals cannot gain on their own.

4.4. The Evolution of Ostensive Communication for Teaching

The idea that the transmission of technology and cultural knowledge more broadly is dependent on some form of communication is of course not new. Nor is the idea that human communication originally evolved for teaching. Laland (2017a, 2017b), for instance, has argued extensively that the coevolution of teaching and cumulative culture produced language to reduce the costs, and increase the accuracy and scope, of teaching. Similarly, Fitch (2004, 2007) believes that the evolution of language as a cheap, honest communicative system is best explained by postulating that language evolved to enable communication among kin, and particularly between parents and their offspring. The topic of this early system included mostly what would be classified as teaching: information about foraging techniques and the dangers and affordances of the environment.

Here, however, I would like to argue that before fully-fledged linguistic communication was developed through cultural and biological evolutionary processes, our ancestors had evolved a basic ostensive communicative system to enable the transmission of opaque technological knowledge—a system that later provided the cognitive structure for communication in various modalities and beyond teaching purposes. This hypothesis is grounded in work in the theory of *natural pedagogy* (Csibra & Gergely, 2009, 2011). According to this theory, hominins' extensive tool use and their dependence on an artifact culture presented novel evolutionary problems that did not exist before (Csibra & Gergely, 2006). Learning tool use in other great apes, and likely early hominids, is facilitated by the fact that the function of tools is largely transparent: there is a relatively immediate reward to tool use that can be leveraged in interpreting the function of the tool. For example, chimpanzees insert

tools made from surrounding vegetation into termite mounds to extract the termites that cling to it (Lonsdorf, 2006). The food reward can then be taken as the function of the tool through observation of the mother's behavior without the need for active teaching. In human tool use, however, there is often no obvious reward. In many cases, a tool is used to make another tool which, in turn, may be used recursively to create yet other tools. There is then a spatiotemporal decoupling of tools and their distal function that renders function attribution difficult if not impossible. Moreover, our tools can be used for infinite non-obvious purposes. This is observable in the difficulty in deciding what ancient artifacts were actually used for or indeed whether they were useless leftovers of the manufacture of some other functional artifacts. Observing others using the tools may partly help one to figure out their utility. If the tool has an effect on the environment that effect could be the purpose of tool use. Yet, the difficulty of choosing between the many outcomes of tool use notwithstanding, one may still struggle to find out which aspects of an action were responsible for the effect—sometimes even experts are unaware of how they habitually manage to produce their desired effect. Thus, as a result of the long chains of action and the means-end decoupling, human cultural and technological practices are cognitively opaque to learners: they are *teleologically* opaque in that the ultimate goal of practices is not obvious; and they are *causally* opaque in that the goal-relevant elements in a sequence of actions are not obvious (Csibra & Gergely, 2011).

Hence, natural pedagogy theory argues, humans have evolved a set of related cognitive adaptations that serve to enable the transmission of knowledge, made necessary by the opacity of human technology. Humans from early in infancy are receptive to ostensive signals such as eye contact and child-directed speech (see chapter 2). Ostensive signals evoke various default expectations and interpretive biases in encoding actions that facilitate and enable the acquisition of cognitively opaque knowledge (Csibra, 2010; Csibra & Gergely, 2011). Children take adult communicators as benevolent teachers and interpret the actions accompanying ostensive signals as providing *for them* relevant, generic information—that is, knowledge that is generalizable across objects and events and shared by members of the cultural group. Furthermore, deictic gestures such as eye gaze (Senju & Csibra, 2008) and pointing help identify and highlight the referent about which knowledge is to be transferred. Children attend to ostensive signals preferentially and, above all, interpret accompanying actions as communicative and referential (Csibra, 2010).

Numerous experimental studies (elaborated in chapter 5) have been conducted which suggest that, in ostensive contexts, infants imitate causally opaque actions such as turning on a light with the head (Király et al., 2013), understand emotional expressions as conveying shared knowledge about object kinds rather than personal preferences (Egyed et al., 2013; Gergely et al., 2007), and encode kind-relevant information at the expense of episodic features such as object location (Yoon et al., 2008). These findings imply that even preverbal infants are equipped with cognitive and attentional adaptations that enhance the learning of complex cultural knowledge including artifact functions, rituals, and conventions.

Thus, to use the terminology from chapter 2, ostensive signals serve to $mark_{NN}$ demonstrations, creating a communicative channel to transmit information open-endedly. This information has a generic nature and, like language, possesses a predicate-argument structure (see chapter 5). While the argument in demonstrations is represented by the manipulated object, the predicate is conveyed by action performed on the object and drawn typically from the spatiotemporal features of the manipulation. Moreover, by conveying a kind-relevant content that is inherently detached from the proximal medium used in communication (i.e., the action and object), these pedagogical demonstrations already involve public representations that can be exploited for linguistic communication, too (see chapter 6). Natural pedagogy has sometimes been characterized as one of the many specializations that ostensive communication acquired in the course of its evolution (e.g., Scott-Phillips, 2014). Here, however, I wish to argue that natural pedagogy is in fact the cradle of ostensive communication and all other uses are more recent exaptations of the communicative system that evolved for the function of teaching. This hypothesis, as I will try to demonstrate, not only sheds light on the difficult problems that any evolutionary theory of ostensive communication must address, but it is more consistent than the alternatives with the evidence we have of the development of this trait.

We obviously do not have access to the actual processes that hominins underwent to evolve full-blown ostensive communication, but a sound speculation still helps to check if the necessary transitions are at least feasible. Extant great apes attempt to make eye contact with their addressee when they gesture (Gómez, 1996). This might serve to monitor the gaze of the addressee to make sure that they are attending to the communicative event. Early hominins might have done the same thing to ensure that the child was observing fitness-relevant actions. The adaptive value of such actions

may have then led to the evolution of preferential attention to action preceded by direct gaze. Parents could, on the other hand, actively exploit this preference to secure the attention of their child during demonstrations—effectively promoting eye contact from a *cue* to a genuine *signal*. Moreover, child-directed vocalizations might have had the function of reassuring physically removed infants at first, compensating for their inability to cling to their mothers like other apes (Falk, 2004). But, similarly to eye contact, these vocalizations were later utilized to draw attention to important actions, eventually rendering these into communicative demonstrations. Although initially mere attention to these episodes might have provided opportunity for learning, the evolution of dedicated mechanisms that interpreted them as entailing kind-relevant information would add to the expressive power of the communicative/teaching system. In this way, ostensive signals were used, and interpreted, to mark demonstrations as communicative. As what is true about the functionality of one artifact is often generalizable to the whole kind, objects manipulated in demonstration could be taken as exemplifying (like symbols) their respective kind (Csibra & Shamsudheen, 2015), while the demonstrated action predicated a property about the objects. This action-based iconic communication was ultimately replaced by conventional means, although the conceptual foundation remained in place.

One argument in favor of a pedagogical origin for ostensive communication is then that it became necessary for the transmission of complex and opaque technological knowledge. Various pieces of paleoarcheological evidence have been suggested to reveal the evolutionary period in which pedagogical communication became necessary (Tehrani & Riede, 2008). Gärdenfors and Högberg (2017), for example, have argued that the Oldowan industry (~2.6 - 1.7 Ma), associated with early *Homo*, features skills that could not be transmitted without teaching via demonstration. This industry involves the production of flakes through hitting one rock, the hammerstone, against another, the core—a process called knapping (R. Boyd & Silk, 2014; Roche et al., 2009; Toth & Schick, 2015). Cut marks on bones suggest that, with the help of lithic tools, hominins consumed the meat of animals much larger than themselves—a behavior not seen in other apes (Roche et al., 2009). Apart from obtaining meat, the flakes may have been used for preparing leather from animal hide and possibly woodworking (Roche et al., 2009; Toth & Schick, 2015). Thus, Oldowan technology already requires the recursive use of tools to make further tools. The lengthening of the chain of necessary actions to achieve an ultimate goal may render the process

opaque to observers, which could explain the rarity of stone toolmaking in non-human primates (but see Motes-Rodrigo et al., 2022). Moreover, core maintenance and procuring raw material from specific locations may involve knowledge transmitted through communication. Nevertheless, some researchers maintain that early lithic tools were likely within the species' "zone of latent solutions" (i.e., individuals could create them independently in their lifetime; Tennie et al., 2017), while others view inadvertent social learning as sufficient for transmitting the necessary skills (Morgan et al., 2015). The long stasis in this technology supports these possibilities.

The subsequent Acheulian industry (~1.7 - 0.13 Ma) may provide stronger pressure for the coevolution of culture and teaching. Associated with *Homo erectus* and *Homo heidelbergensis*, Acheulian technology is characterized by bifacial stone tools (e.g., symmetrical cleavers and handaxes) fashioned from large flakes and cobbles (R. Boyd & Silk, 2014; Roche et al., 2009; Toth & Schick, 2015). The complexity of this technology is hypothesized to require active transfer of displaced concepts (e.g., to teach platform preparation; Gärdenfors & Högberg, 2017). Accordingly, it is argued that linguistic communication might have been present in this period (Gärdenfors & Högberg, 2017; Morgan et al., 2015). Schaik et al. (2019) propose that it is only around 500,000 years ago that hominin culture became truly cumulative and reached outside the primate range. At this point, hominins started producing composite tools—for instance, by hafting stone points onto wooden handles using adhesives. The opaque skills involved in the production of such tools are unlikely to be discovered by single individuals, and so means of cultural transmission including a relatively sophisticated communication were perhaps necessary to share them. Similarly, the early use of fire and cooking (Wrangham et al., 1999; Zohar et al., 2022) may have involved tool use and other types of opaque know-how that called for instruction. Specifying the precise period in which ostensive communication became necessary is beyond the scope of the present thesis and may be impossible based on existing evidence. The main point is nonetheless that the dependence on increasingly complex technology made the associated skills and knowledge opaque to a degree that the traditional means of social and individual learning were no longer adequate. These circumstances then provided selection pressure for the emergence of a flexible communicative system for efficient and high-fidelity transmission of information from parents to children.

Besides these mostly theoretical observations, various experimental studies have attempted to elucidate the role of pedagogical communication in the transmission of

technology and cumulative cultural knowledge. Results from experimental archeology suggest that, whereas simple stone-toolmaking skills characteristic of Oldowan technology *can* be acquired without teaching, verbal and nonverbal communication augment the effectiveness and efficiency of the process (Lombao et al., 2017; Morgan et al., 2015; but see Putt et al., 2014). So, flexible communicative means may have evolved to facilitate transfer of early toolmaking, driving a coevolutionary process between culture and communication of increasing complexity. This would eventually lead to a protolinguistic communicative system capable of producing the skills required for more sophisticated Acheulian tools. Similarly, Lucas and colleague's (2020) transmission chain experiments show that simple toolmaking can be learned non-pedagogically, yet the manufacture of complex, causally opaque tools significantly benefits from teaching through communication. They conclude that the initial reliance on cumulative culture may have generated species-unique selection pressures for teaching in humans.

There is, additionally, some neurocognitive evidence in support of the pedagogical origin of ostensive communication. Stout and Chaminade (2009, 2012) present neuroscientific evidence indicating that language-relevant regions of the brain such as the inferior frontal gyrus contribute not only to the comprehension and production of syntactical structures but are also activated for the processing of object manipulation. This is likely due to the similar role of hierarchically structured information in the domains of language and tool use. Their *technological pedagogy* hypothesis proposes that intentional pedagogical demonstration could have provided a scaffold for the evolution of intentional vocal communication. Likewise, Kolodny and Edelman (2018) note the overlap between the neural underpinnings of language and tool use, emanating from hierarchical planning and sequential control. They argue that most evolutionary theories of language have focused primarily on ultimate explanations for language, ignoring the proximate mechanism that supported its evolution. According to their *Cognitive Coupling* hypothesis, language relies on the coupling of pre-existing neural mechanisms for communication and serial behavior. Therefore, they believe, the most likely ecological context for the emergence of language that incorporated this coupling was the teaching of tool use and production.

I argued in the second chapter that ostensive signals seem to develop earlier than other types of ostensive stimuli and are responsible for bootstrapping the communicative system. Moreover, as mentioned here, these signals appear to evoke

interpretive biases (e.g., genericity assumption) that are suited to teaching—implying that children take adults who address them with ostensive signals as teachers. If indeed ostensive signals are central in the development of our communicative concept and they initially serve as teaching-initiators, one may conclude that for infants the teaching function is the primary function of communication early in development. Of course, ontogeny does not necessarily recapitulate phylogeny. But the tentative developmental primacy of teaching may be taken as providing indirect evidence for the hypothesis that teaching was the primary function of ostensive communication in phylogeny as well, retaining this prominent place despite being coopted for other communicative purposes in evolutionary time. This could, for instance, make the emergence of this function more reliable ontogenetically. (Alternatively, the early receptiveness to communicative teaching could be explained by its potential role in facilitating language acquisition.)

This conjecture about the primacy of teaching is also backed by the status of generics in language development. As noted in the previous chapter, scenarios other than teaching mostly emphasize contexts involving communication about episodic features such as information about the presence of prey, coordinating joint action, socially relevant events, etc. This predicts that episodic communication (i.e., about transient, accidental, context-bound properties) has a privileged status in development. The pedagogical scenario, however, proposes that initially children were the pupils in teaching events and so they have come to be especially receptive to pedagogical communication. As a result, the theory predicts that children be more prepared to acquire generic information (i.e., about essential and enduring properties) than episodic information. As already mentioned, if addressed by ostensive signals, infants seem to encode generic properties (e.g., object kind) at the expense of transient, episodic properties (e.g., object location). This of course does not mean that children are unable to receive episodic information in communication. As a general-purpose communicative system, ostensive communication has evolved various functions that go beyond teaching. However, if its primary function was to enable teaching, pitted against each other, a generic interpretation may override an episodic one. The claimed generic bias appears to be present in linguistic communication, too. Children show motivational, encoding, and memory biases for category-related information compared to quantified information—misremembering, for instance, a quantified statement as a generic statement (Cimpian, 2016). This has led some

scholars to attribute a default status to generics in comprehension (Gelman, 2004; S.-J. Leslie, 2008). Gelman notes that, unlike generics, non-generic statements are marked linguistically by devices such as determiners, number, and tense. Therefore, generics seem to be marked more by the absence of episodic markers than by the presence of specialized generic markers. Language users then assume that a statement is generic unless this interpretation is somehow blocked. Accordingly, language learners should learn the ways particulars are marked linguistically and contextually to filter out the specific from the generic. Thus, these patterns of language use seem more consistent with the present pedagogical hypothesis than the alternatives. If ostensive communication evolved for collaboration, children would perhaps show a bias for episodic interpretations. Coordinating, say, a scavenging may require communicating how many people are required, what the specifics of the location are, and what needs to be done to achieve the goal—all involving episodic features.

In sum, the hypothesis is that ostensive communication evolved to enable teaching open-endedly about various kinds of technological knowledge that individuals could not acquire on their own—facilitated by interpretive biases such as a generic construal of the transmitted information. This hypothesis is backed by several empirical and theoretical observations regarding the development and evolution of humans. Next, I turn to the five criteria introduced in the previous chapter to see if the present account can satisfy them convincingly.

4.5. Uniqueness

The first criterion to address is uniqueness. A sound theory of the evolution of ostensive communication must explain why it evolved in humans but not in other animals including great apes. This is where most theories fail to offer a convincing proposal because most suggestions involve pressures that are likely present for other species as well (Számadó & Szathmáry, 2006). Although ostensive communication and language do create greater opportunities for hunting, scavenging, advertising, bonding, and cooperation, other animals would arguably benefit from the expansion of topics for communication in those domains too. Thus, the scenario must be somehow evolutionarily novel. Besides, even if the aforementioned scenarios involve novel elements (e.g., hunting a diverse range of megafauna, advertising intelligence, bonding bigger groups, cooperating more abundantly, etc.), one must demonstrate

that the proposed conditions called for a new, flexible communicative system. This is, as argued in chapter 2, a system that permits formal and informational open-endedness—enabled by marking$_{NN}$ and the representational use of external stimuli (such as objects and actions). Thus, the scenario should indicate why natural selection did not solve the evolutionary problem using the "traditional" limited signaling systems of other animals. Although miming the prey species before the hunt may facilitate planning, an expansion of the great ape vocal or gestural repertoire could have solved the task. That ostensive communication makes this easier is not a sufficient explanation because all domains of use could profit from the addition of an open-ended system on top of the already existing innate repertoire. Metarepresentation could perhaps enable such a system through virtual domain generality (Heintz & Scott-Phillips, 2022), but it only provides a negative explanation (i.e., removal of a cognitive constraint) for why open-endedness became possible, not why it was uniquely required in our lineage. A positive account should therefore be one that suggests problems which are evolutionarily novel *and* could be resolved only through open-ended communication.

The argument so far has been that the complexity of human technology and its relative opacity made a communicative transmission of knowledge *necessary*. That is, although early toolmaking may have been possible without active skill transmission, the increasing dependence on, and complexity of, such technology necessitated flexible communication. Pedagogical demonstration already (i.e., before conventional language) enables open-ended communication of opaque skills. By marking$_{NN}$ your object-directed actions, you tap into the addressee's cognitive mechanisms for interpreting ostensive communication. As a result, the addressee understands the act as relevant, generalizable, and shared in the community—enabling thus a more faithful transmission of knowledge. However, this purported necessity of ostensive communication is insufficient to account for its evolutionary emergence. It could be the case that ostensive communication evolved for another (narrower or broader) function and once it was in place it was also exploited for conveying technology-related information. Hence, to offer a convincing proposal, teaching should not only benefit from more efficient communication, but it should also provide stronger and more specific pressure for ostensive communication compared to the alternative scenarios. In other words, one must show that ostensive communication is indeed well-suited to teaching, such that teaching explains its design features. I believe that this is the case.

The fact that we can generate novel communicative means and inform others open-endedly suggests that our communicative system is adapted to an open-ended evolutionary domain. Of course, all domains that can profitably be the topic of communication have open-ended aspects. No two hunting travels are identical with respect to the location, type of prey, and specific events. Nor are social topics regarding "who did what to whom" all alike. Nonetheless, natural selection has solved similar problems by producing the typical animal signaling systems which are limited but useful for the type of content they convey. Although snakes are not identical and their presence always involves new features (e.g., where exactly they are spotted, what species they belong to, etc.), alarm calls of a limited nature have been adequate for alleviating the danger. This is because such topics, despite the variance in the specifics of the situation, concern recurring evolutionary problems. Yet, dependence on an artifact culture poses problems that natural selection cannot keep up with (see also Laland, 2017a). As our technology varies both in space and time according to diverse local needs, it is inherently open-ended. And the body of knowledge that is useful in one environment, may be useless or even maladaptive in another environment. Thus, a communicative system for the transmission of technology-related knowledge (and perhaps later, also cumulative culture) had to be open-ended. To simplify, the recurring problem that such communication should solve is open-endedness itself.

Ostensive communication is then unique because it is targeted at a unique problem, that is, to teach opaque technological knowledge. Other great apes do not communicate ostensively (at least, not to the extent that we do) due to lesser reliance on such technology and cumulative culture, especially ones that involves relatively opaque elements. Hence, chimpanzees need not mark$_{NN}$ their actions to inform open-endedly. Solving the range of problems that they need to communicate about is done by typical, although more flexible, innate signals (Byrne et al., 2017; Fischer & Price, 2017). And what they need to learn socially, they acquire via observing their mothers, without requiring active transfer of information (Hoppitt et al., 2008). Moreover, humans seem to be the only species that transmits generic information in communication (Csibra & Gergely, 2011). Besides, infant-directed communication is relatively rare in non-human great apes (Schick et al., 2022). Thus, the lack of ostensive communication in other apes is coupled with an absence or rarity of opaque technology, cumulative culture, active teaching, and generic and infant-directed

communication. In humans, on the other hand, dependence on an artifact culture necessitated active teaching, specifically one that involves open-ended communication of generic information to infants.

4.6. Immediate Utility

The question is now how ostensive communication was of immediate utility before it was fully developed to its modern state. Particularly, the system had to be functional before the establishment of a large number of conventional symbols and grammatical rules (Bickerton, 2009). Since evolution lacks foresight, the early communicative means could not evolve such that they be put to use only when their number reaches a significant level. Moreover, the evolutionary development of sophisticated symbolic means rests on the existence of a (more or less specialized) acquisition device that enables their ontogenetic development, and the evolution of this device relies, in turn, on the existence of symbols that justify its selection (Origgi & Sperber, 2000). As discussed in the previous chapter, the best answer to these dilemmas seems to be that initially there existed a functional non-verbal communicative system that did not require conventions and depended instead on pragmatic principles and "natural" means of communication.

The emergence of communication in teaching is also consistent with this conclusion. As argued, the earliest instances of human communication involved ostensive demonstrations of object-directed actions. If true, these interactions required that the action be marked and understood as communicative by participants. This would solve the problem of signaling signalhood (Scott-Phillips et al., 2009). However, the informational content needed to be somehow transmitted without a conventional lexicon and syntax. This was possible because the medium of communication utilized already meaningful entities: objects and actions. Communication without conventions was therefore enabled by exploiting the more ancient systems of object cognition and goal-directed action understanding. One interpretation of infants' early understanding of demonstrations is that they interpret objects as symbols exemplifying the kind they belong to (Csibra & Shamsudheen, 2015). This follows from a parsimonious reading of infants' generalization of demonstrated properties to other members of the object kind (e.g., Butler & Markman, 2012, 2014). If so, the ostensive manipulation of objects permits proto-symbolic communication of various predicates about the object kind before the emergence of full linguistic capacities. Hence, like many others, the present

theory considers iconic communication as the initial mode of form-content mapping. Although demonstrations are cognitively opaque with respect to their functionality, they are transparent with respect to their form-content mapping (as compared with arbitrary conventions). That is, objects in demonstrations typically exemplify their kind (unless used in pretense) and their manipulation manifests the intended predicate (e.g., their function). Therefore, pedagogical demonstration enables communication that is adaptive from the outset without any conventional lexicon and syntax.

But how did conventional symbols emerge? One common way of thinking about this is that they originated as emergent solutions to a coordination problem (e.g., Scott-Phillips et al., 2009). Competent adults try to communicate with one another and as a result conventions arise, facilitating future interactions. However, this is not the only function of conventional labels. Labels also allow one to acquire categories vicariously. While other animals rely on induction and trial-and-error to learn about categories, symbolic communication enables learning through other individuals. The advantage of this way of acquiring categories is that it is more efficient and it does not entail the risks involved in individual experience (Harnad, 2011). Several developmental studies suggest that this function is indeed available to infants in comprehension, long before labels can be produced by them to facilitate linguistic interactions. For instance, verbal labeling changes the way infants categorize objects along a perceptual continuum, forming one category when presented with one or no label and forming two categories when presented with two labels (Althaus & Westermann, 2016; Havy & Waxman, 2016). More dramatically, 10-month-old infants expect different-looking objects to share non-obvious properties (e.g., making a sound) when they are assigned the same label, while they expect similar-looking objects with distinct labels to possess different properties (Dewar & Xu, 2009). Thus, from early in infancy, labels seem to work as invitations to construct categories (Waxman & Markow, 1995)—a process that supports learning about the world. This is done both positively by emphasizing commonalities between identically labeled objects and negatively by emphasizing dissimilarities between differently labeled objects.

Parents can thus make use of this capacity in infants to pedagogically specify the extension of the kinds. This is advantageous independently of whether the labels will subsequently be produced for referring to particular objects outside the categorizing context. While labeling one fruit may not convey any information in itself, realizing that it shares a label with another fruit can inform you that the two share hidden properties

and help you in deciding whether to eat it or avoid it. Thus, although the generation of labels as emergent by-products of interpersonal communication is also consistent with the present account, another possibility is that labels were initially introduced in an asymmetrical pedagogical context for their more direct benefit in transmitting kind-relevant knowledge, similarly to action demonstrations. This would resolve the explanatory problem of the immediate utility of early words. Furthermore, this possibility is evolutionarily more parsimonious in that it explains many features of conventions without postulating extra processes. The abovementioned categorizing capacity does not seem to be unique to verbal labeling. For example, ostensive demonstration of artifact function has been shown to facilitate object individuation similarly to words (Futó et al., 2010)—suggesting that infants may categorize artifacts according to their demonstrated function (see also Booth et al., 2010). As demonstrations were associated with object kinds, this likely opened up the way for exploiting the existing disposition in order to assign arbitrary action-object pairings that helped specify the extension of object kinds. Besides this categorizing feature, object-directed actions are, as mentioned, interpreted as shared in the group and generalizable. Additionally, the generic nature of demonstrated manipulations already involves normative stipulations ("This is how one uses this tool!"; see also Birch, 2021). Therefore, these features did not have to evolve separately for linguistic conventions. In conclusion, the pedagogical origin of ostensive communication meets the immediate utility criterion, while also providing explanation for other properties of human communicative means.

4.6. Generality

The teaching scenario should be able to account for the system's power of generalization. Of course, specifying a domain for the early emergence of ostensive communication restricts the range of topics that communication can inform about. But teaching technological knowledge creates strong pressure for open-ended communication within the specified domain—a feat that may not be as available to other proposed functions. As mentioned in the discussion on uniqueness, the open-endedness of communication was dictated by the inherent open-endedness of technology and other types of socially acquired knowledge. The breadth and variability of our artifact culture was thus matched by a similarly productive communication that

was amenable to change. In ultimate terms, then, teaching provides the necessary selection pressure for generality of communication.

But were the initial means of communication capable of such open-ended communication? Marking$_{NN}$ allowed rendering any object-directed action communicative and so capable of conveying pedagogical information. In addition, if objects indeed came to be represented as symbols exemplifying the kind, these could be manipulated in various ways so that the spatiotemporal relation between objects (e.g., A should be inside B) and the action performed on them (e.g., A should be struck repeatedly) could convey an open range of messages, before any language was created (see next chapter). Therefore, this system was limited mainly by the objects that could be manipulated (the arguments) and the possible actions on them (the predicates). Demonstrations are thus a generative means of communication that could be used for flexible teaching and provide a cognitive platform for other uses including in a general-purpose language.

Displacement, as a factor contributing to the generality of ostensive communication, is difficult to explain for most theories, especially when combined with immediate utility. How can the communicative system be immediately useful and yet capable of denoting spatiotemporally absent entities? The pedagogical scenario can account for this as well. Generic communication about objects meant that what was proximally demonstrated about a particular object referred distally to the object kind. And kind is an abstract entity that is essentially displaced from the here and now. Relatedly, demonstrations are cases of depiction: a physical scene is staged so as to inform about other entities to which the object-props refer (H. H. Clark, 2016). If so, demonstrations already involved representational communication in an iconic format that could be utilized later for arbitrary symbolic communication (see chapter 6). Because representational use of perceptual stimuli necessarily conveys a content that is detached from the medium used for informing (i.e., a prop is distinct from what it represents), this detachment supports increasing use of communication about absent referents. Consequently, representational communication about kinds through demonstration creates the conceptual foundation required for other uses of public representations, including the utterance of words and gestures to denote spatiotemporally absent phenomena. Overall, pedagogy not only accounts for the ultimate forces behind the generality of ostensive communication, but it also explains the proximate mechanisms needed for this.

4.7. Honesty

How was the honesty of early communication guaranteed despite potential incentives to cheat? If the formulation of ostensive communication in the second chapter is accepted, we are in fact faced with two problems. Ostensive communication involves two types of communication: marking$_{NN}$ and informing$_{NN}$. With respect to the latter, the problem is how the communication of the informational content was stabilized. Regarding the former, on the other hand, the problem is how the marking of action as communicative was stabilized. Although these two layers of communication may be enacted in a single action (e.g., shaking an object) rather than two (e.g., making eye contact and showing an object), the fact that they can be separated suggests that they must be given distinct, however interrelated, evolutionary explanations.

As my proposed scenario concerns mainly teaching between parent and child, the most obvious account of the reliability of informing in communication is inclusive fitness through kin selection (Hamilton, 1964; Maynard Smith, 1964). Thus, the reason early ostensive communication was honest regarding what it conveyed was that, due to common adaptive interests, dishonesty would be detrimental to the inclusive fitness of the communicator. Given that our ancestors likely lived in kin-structured groups (Laland, 2017b), honest communication would affect both “direct” (by teaching offspring) and “indirect” (by teaching relatives) fitness (Griffin & West, 2002). It is, of course, another question how much a parent is willing to invest in teaching; but once the parent engages in teaching, it is likely to be honest (Laland, 2017a). If indeed the early communicative behavior of humans involved demonstration, there was no strong incentive, say, to demonstrate an incorrect or less optimal use of artifacts. Kin selection also explains why ostensive communication is a relatively cost-free system: due to a lack of conflict of interests, there is no need for paying production-related costs—the cost is paid instead by the harm to inclusive fitness in the case of deception.

This proposal is consistent with Fitch’s (2004, 2007) *mother tongue* hypothesis of language evolution. According to this hypothesis, language evolved as a “mother tongue”, that is, a system used among kin and especially parents and their offspring. The long period of children’s dependence on their parents and humans’ reliance on regionally-variable, complex, learned extractive foraging techniques created selection pressure for the evolution of a semantic communicative system that provided the necessary information for survival in such ecological conditions. This kin-based

scenario, as mentioned, also accounts for why language is a cheap signaling system (de Boer, 2011).

The second layer of communication conveys that communication is taking place, without necessarily transmitting any further content. In theory, one could deceptively make eye contact or produce other ostensive stimuli, while there is indeed no valuable content to be transferred or the content is misleading. Yet, provided that the reliability of informing$_{NN}$ is guaranteed through kin selection (or any other process, for that matter), then markers$_{NN}$ could work as unfakeable signals. When you signal that communication is taking place, this is proof in itself that it is indeed taking place—communication is evidence of communication. For this reason, ostension or marking$_{NN}$ may work like an *index* (Maynard Smith & Harper 2003): it cannot be faked because there is a causal relation between the signal and its function.

Nonetheless, even if communication was stabilized through kin selection, it has evolved to be used outside of kin-based teaching contexts. So, these arguably novel uses may have required other processes to remain evolutionarily stable strategies. Thus, because ostensive communication is a general-purpose system with multiple important functions, its reliability may also have been ensured by multiple processes. Teaching by alloparents is one context in which shared interests can stabilize communication between non-kin. Reputational consequences of cheating, such as not being believed in future interactions, is another candidate for low-cost communication across domains (Maynard Smith & Harper, 2003; Scott-Phillips, 2014). However, for reputation to work, individuals must be able to recognize each other and, crucially, remember past behavior. The latter is particularly important for humans, since often the topic of communication is displaced and not immediately verifiable. Therefore, reputation may necessitate some level of episodic memory for keeping track of what each individual communicates and verifying its reliability in the future. The existence of such a capacity for episodic memory is, however, not a trivial assumption and may require metarepresentation (Cosmides & Tooby, 2000; Mahr & Csibra, 2018).

4.8. Cooperativeness

Cooperative use of communication is not as major a problem as honesty. Whereas all signaling systems must guarantee a degree of honesty to remain stable, communication can arise even in antagonistic contexts (e.g., threat signals). Thus, in principle, it is possible that ostensive communication also emerged in a non-

cooperative context. That is, although there had to be informative cooperation (i.e., honesty), material cooperation (i.e., prosocial communication) was not necessary.

In some theories, cooperation is even considered to be essential to the proximate working of the system (Heintz & Scott-Phillips, 2022; Tomasello, 2008). In a sense, all signaling systems in nature are cooperative: on average there must be a benefit to the receiver for the system to evolve, and ostensive communication is no exception. Yet, as long as there are ultimate guarantees for this (e.g., a kin-based social organization), such cooperation need not be proximately "assumed" by the interactants. So, while it is of general importance to know what ensured reliability, the cooperativeness of human communication is not of unique or central importance in explaining its evolution (contra Knight, 1998).

Despite this proviso, there are some indications that infants expect participants in communication to be generally prosocial. For example, 13-month-olds take a third-party interaction in an unfamiliar communicative channel to involve information relevant to the goal of a naïve agent (Tauzin & Gergely, 2018). Infants point cooperatively to help others (Tomasello, 2008; Tomasello et al., 2007) and expect recipients of gestures to prosocially fulfil a request (Thorgrimsson et al., 2014). These may suggest that ostensive communication indeed evolved in a highly cooperative context. This expectation is in line with the teaching scenario. If ostensive communication arose in a pedagogical context involving kin, the expectation that it be prosocial is evolutionarily well-founded.

As mentioned before, many cooperative enterprises, especially those concerning a spatiotemporally displaced goal, might depend on the presence of a flexible communication and as such are driven by, rather than drivers of, ostensive communication. Once there is a sophisticated communication in place and there are opportunities for cooperation, the system can be coopted for various new joint goals. Nonetheless, although prosocial communication beyond kin-based contexts is not readily explained by the teaching scenario, the *pattern* of material cooperation seems to be exactly what would be expected from a kin-directed communication: where information is valued highly, it is not freely given to non-kin. For instance, pre-contact fishermen in Hawaii shared the location of productive fishing grounds only with family members (Kamakau, 1976, as cited in Burdett et al., 2018). Similarly, among lobstermen in Maine's Middle Harbor, information about the location of lobsters was kept from people other than kin and those with a similar skill level (Palmer, 1991).

Where new information is scarce due to the public availability of people's social life, even apparently futile information is not provided to non-family members. Thus, in rural Madagascar (as contrasted with western cultures), if A asks B, "Where is your mother?" and B answers, "She is either in the house or at the market", this answer is usually not interpreted as implicating that B cannot provide more specific information (E. L. Keenan & Ochs, 1979; E. O. Keenan, 1976). There, information appears to be a valuable commodity only shared with kin and neighbors.

To sum up, material cooperation is not a priori necessary for selective scenarios and may not be as crucial as informative cooperation. However, the pattern of cooperation observed in human interactions may need explanation. Both children's expectation that interlocutors be prosocial and the exclusive sharing of vital information among family members are well explained by the hypothesis that ostensive communication evolved to support teaching among kin.

4.9. Possible Objections

Above we saw that the pedagogical scenario can indeed satisfy the five criteria for a sound evolutionary theory of communication. Specifically, teaching seems to provide more straightforward answers than alternatives to the uniqueness and generality of ostensive communication. While open-ended communication is useful for all proposed functions, transmitting variable technological knowledge through communication is arguably possible only if the communicative system is similarly flexible and responsive to change. Moreover, pedagogical demonstration, as a means used also by contemporary humans, is immediately useful in the absence of a linguistic code and, through its representational nature, can denote displaced entities. Honesty and cooperativeness, too, are explained well by positing a kin-selected communication. Collaboration-based accounts, however, do not seem able to explain all these features simultaneously. Despite participating in joint activities to a greater extent than other apes, why this is so for humans and why this necessitated open-ended communication, rather than an enhanced primate-typical gestural or vocal system, is less clear. Coordinating future action calls for displaced reference but presupposes developed representational devices. The initial collaborative contexts were likely tied to the here and now and so did not require flexible public representations. Therefore, the emergence of representational devices does not follow readily from these contexts.

And honesty of early communication beyond these immediately verifiable contexts may need episodic memory—making the scenario less parsimonious.

Despite the advantages of teaching in explaining ostensive communication, several objections might be raised against its feasibility. One objection is that teaching is not as widespread among traditional societies as in industrialized cultures (Tallerman, 2013). However, as mentioned before, traditional societies including hunter-gatherers do practice subtle means of teaching, although their respect for autonomy and egalitarianism often inhibits active intervention (Boyette & Hewlett, 2018). Moreover, the prevalence of human teaching should be compared against other animals and especially great apes, rather than industrialized cultures (Csibra & Gergely, 2009, 2011). But teaching is rare, if not absent, in other apes—suggesting that humans engage in teaching much more extensively.

Another potential objection might be that early in development infants do not produce communication for teaching. As mentioned, children develop teaching skills relatively early (Strauss et al., 2014; Strauss & Ziv, 2012). Nonetheless, in the proposed evolutionary scenario, infants have the role of pupils rather than teachers. As a result, they should be receptive to teaching, while adults provide them with learning opportunities. Both of these requirements are, as discussed, supported by developmental and ethnographic studies. The early pedagogy-relevant productive communication of infants should instead *elicit* teaching from caregivers. Pointing as one of the earliest-emerging adult-like communicative skills (Tomasello, 2008) is indeed used for this purpose. Besides informing and sharing (Tomasello et al., 2007), early pointing is suggested to serve an interrogative function (Southgate et al., 2007). This hypothesis is backed by empirical studies suggesting that infants point more for adults than for peers (Franco et al., 2009; Kachel et al., 2018), and more for competent adults than ignorant ones (Begus & Southgate, 2012). Additionally, informative feedback elicits more pointing from infants than sharing interest or attention does (Kovács et al., 2014). Prelinguistic vocalizations seem to have a similar function, as suggested by evidence that informative response to these vocalizations promotes learning in infants (Goldstein et al., 2010; Matthews, 2020). Aside from its direct epistemic benefit for infants, interrogative pointing can also solve a related problem for teachers. Flexible teaching requires a means of identifying infants' level of knowledge. This can be done by an advanced theory of mind that enables (meta)representing others' knowledge. Nevertheless, behavioral cues (e.g.,

incompetence) indicative of the child's skill can guide adults lacking sophisticated mental state attribution (Csibra & Gergely, 2006). Likewise, producing gestures and vocalizations with an interrogative force potentially provides direct evidence of knowledge. If infants routinely point when they observe an unfamiliar stimulus (Begus & Southgate, 2012), this can show to the caregiver that infants likely lack knowledge about the stimulus and are in need of instruction.

It has been proposed that many of the species-unique receptive abilities of infants are due to cultural evolution rather than innate, genetically coded dispositions, while some of the traits are shared with other animals (Heyes, 2016). The central idea that open-ended communication emerged for teaching is not incompatible with this proposal. However, it seems unlikely that the entire set of interrelated, early-developing cognitive dispositions suggested for natural pedagogy are the result of cultural evolution. Furthermore, although many of these traits such as a preference for eyes may have ancient phylogenetic roots, revealing thus features homologous with other animals, they have nonetheless acquired specific functions in our species, namely marking action as communicative which is, in turn, used to teach generic knowledge. As such, even if other animals are eventually shown to exhibit preference to ostensive signals, they are unlikely to interpret action following these as communicative and conveying generalizable knowledge. Future comparative and cross-cultural research can shed light on how much of our pedagogical abilities are indeed due to innate, human-specific adaptations.

That men and women communicate equally well may be seen as inconsistent with the central role of mothers in teaching. Nonetheless, although linguistic abilities are slightly biased towards females, paternal care is a distinctive feature of humans compared with other great apes (Fitch, 2007). This unique involvement of fathers and cooperative breeding in general point to the possibility that both mothers and fathers participated in the provision of adaptive information for offspring and evolved the ability to communicate ostensively. Theories based on sexual display, hunting, and confrontational scavenging, however, seem to predict better communicative capacities in men than in women. Thus, compared with these, the pedagogical scenario is more compatible with the equally distributed (and slightly female-biased) skills in humans.

4.10. Subsequent Evolution and Relation to Other Theories

Both ostensive and linguistic communication are general-purpose systems that enable transfer of information in a diversity of domains. A theory of their evolution should then account for the transition from specific domains of use to the capacity for generalized communication. More specific to the teaching scenario, the theory should explain how a system for transmitting generic information could also be extended to convey information restricted to particular episodes involving specified entities, agents, times, and scopes of generalization.

Once an open-ended system of communication emerges, it becomes available to various pressures that may not have driven its emergence. Vertical teaching by biological parents can be followed by teaching by alloparents and eventually horizontal teaching between related peers. With the increasing complexity of toolmaking, experts can also demonstrate and advertise their specialized skills to others and gain in status (Hiscock, 2014). This allows for teaching between more distant individuals. The reliance on tools also enables collaboration for transfer of raw material and joint toolmaking (Planer, 2017b). These activities create opportunities for applying the already existing communication in new contexts involving episodic features. Moreover, although teaching itself requires generic information transmission, coordinating teaching events may benefit from communication in the here and now. The coevolutionary emergence of manufactured weapons and flexible communicative devices can facilitate hunting and confrontational scavenging, which are both dependent on planning, coordination, and a body of knowledge about the necessary skills (e.g., tracking animal spoor) and animal behavior (Lombard, 2015). Complex toolmaking additionally necessitates a degree of skill standardization. Teaching standard toolmaking techniques may therefore have led to the earliest forms of norm enforcement, involving the necessary processing and affective components of normative cognition. These tool-directed norms could then give rise to social norms for managing pair bonds and cooperative activities (Birch, 2021). Thus, although teaching may be preferred in accounting for the early emergence of ostensive communication, other scenarios (collaboration, advertisement, hunting, social norms, etc.) are also valid for explaining the further sophistication of the system and its application across adaptive domains.

One outstanding question is how conventional language, including its symbolic devices for episodic communication (e.g., tense, number, quantification, determiners,

and modalities), arose from generic demonstrations involving the physical manipulation of object-symbols. Storytelling can perhaps present such a link. Demonstration, like any other means of communication, must be reliable to be worth attending to. However, as a generic medium resilient to counterexamples, its reliability is determined by the usefulness, rather than the truth value, of the content. Moreover, early demonstrations may have transmitted information that is also locally applicable, meaning that the medium of communication (i.e., action on objects) was itself an instance of the type it denoted. However, as a depictive medium, demonstration also permitted the communication of hypothetical or modal knowledge. Arrangements of objects were not only to be understood as observable physical relations, but also as symbolic relations between kind exemplars containing relatively abstract knowledge (e.g., “A generally goes with B” or “A is handled this way”). If such a depictive way of communication was available to our ancestors, one can also envisage that these object-symbols were replaced by their verbal predicates, namely words, to yield proto-stories. Storytelling, like demonstration, uses hypothetical as well as true events to create narratives. These narratives are moreover believed to encode generic knowledge, accompanied by various ostensive stimuli (Scalise Sugiyama, 2021). Accordingly, it has been proposed that teaching is one of the primary functions of storytelling (B. Boyd, 2018; Scalise Sugiyama, 2021). By listening to stories children can acquire knowledge about a vast range of domains including animal behavior, capture techniques, and social and demographic information (Lombard, 2015). This similarity between demonstration and storytelling makes the latter a good candidate for the transition to language. Interestingly, despite conveying generic information, stories make frequent use of episodic devices. One possibility is then that episodic markers emerged originally to facilitate keeping track of the elements within a story (the co-text), rather than in reference to external situations (the context). While in depiction physical objects and the sequence of events are identifiable by perceptual features, one needs symbolic means (e.g., tense, pronouns, demonstratives, proper names, etc.) for making a similar distinction between elements in a story. If this transition from depictive demonstrations to storytelling is plausible, the latter provides a link between communicating based on actions and objects to full-blown linguistic communication.

4.11. Conclusion

In the second chapter, I suggested that ostensive communication is best characterized by marking$_{NN}$—a distinctive property enabling formal open-endedness in creating novel communicative means. These novel means, in turn, enable informational open-endedness, that is, transmitting an open range of messages. Designating open-ended communication as an explanandum for evolutionary accounts, one can then ask which ecology fosters this type of communication, as opposed to the limited and largely innate signals of other species. The answer here has been that teaching cognitively opaque knowledge, generated by dependence on an advanced technology, provides a context in which one needs to communicate variable contents that fixed signals, established by natural selection, cannot accommodate. Such a context is arguably not provided by other scenarios. Teaching also satisfies the five criteria introduced in the previous chapter, offering more straightforward answers than alternatives.

The present chapter is by no means intended as an exhaustive account of the origin of ostensive communication or language. There are many important properties of language that were not addressed here. The purpose of this chapter was rather to explain the evolutionary emergence of the more basic pragmatic building blocks needed for ostensive communication. Nevertheless, it could still be argued that although human communication involves manifold properties that are interesting in their own right, they would not arise or be nearly as productive if the pragmatic foundation was not in place. Features such as conversational turn-taking, phonology, and syntax, *inter alia*, are all predicated on the existence of flexible communicative means. Thus, although their origin may be distinct from ostensive communication (say, in a language of thought), their productive use in communication is enabled by the presence of an open-ended system. Without such a system, conversational turn-taking as a collaborative process would be very limited, phonology could not make as many semantic distinctions as it does, and grammar could not operate over the numerous symbolic units generated thanks to ostension. Therefore, explaining open-endedness as a core property has explanatory priority in accounting for human communication over other, arguably more recent, developments of the system. Notwithstanding, in the next chapter I will try to show that some of the semantic and syntactic properties of language may already be present in pedagogical demonstrations.

Chapter 5. Speaking in Action: Predicate-Argument Structure as an Adaptation for Interpreting Action Demonstrations

5.1. Introduction

In the previous chapter, I suggested that ostensive communication evolved to enable teaching between parents and their offspring. Hence, our communicative system had a largely pedagogical function. In the remainder of the dissertation, I aim to discuss whether and how this system can indeed permit the open-ended transfer of information. The next chapter will be focused on the contentious role of metarepresentation in all instances of ostensive communication, whether used for teaching or not. Here, however, I will try to show that pedagogical demonstrations, although simpler and more limited than language, are actually capable of conveying an open range of contents. Particularly, I will enquire into the ontogenetic and phylogenetic origin of the predicate-argument (henceforth P-A) structure in human communication as a property contributing to the productivity of the system. While this structure may have had more ancient roots in individual cognition, its expression and interpretation in communication calls for explanation. Thus, I argue that this usage is a cognitive adaptation for children's interpretation of, and learning from, instances of pedagogical demonstration. More specifically, objects in demonstrations are suggested to play the role of arguments, while actions performed on the objects act as predicates—together communicating a proposition-like content. As a result, nonverbal communication already involves P-A structure before the mastery of a full-blown linguistic system.

The question of P-A structure is of crucial importance in linguistics and the evolutionary study of human communication. It has been suggested that P-A structure constitutes the semantic foundation of language (Hurford, 2003), and that "the fundamental operation of meaning composition is the application of predicate terms to argument terms" (Rocci & Luciani, 2016). If we do not understand this process, Davidson (2009) claims, we cannot account for the structure of the simplest thoughts expressible in language. Dessalles (2008), on the other hand, argues that syntax is a tool that is devoted to the expression of predicates. And in a different theoretical framework, the P-A distinction has been characterized as a linguistic universal, accounting for the distinction between verbs and nouns (Luuk, 2009). Propositions, as

the outcomes of combining predicates and arguments, are commonly assumed to provide the semantic content of sentences. It is thus no wonder that, in their seminal paper on language evolution, Pinker and Bloom (1990) see grammar as a mechanism tailored to the transmission of propositional structures. Despite its significance, however, the literature on the origin of P-A structure in communication is rather scant.

The notion of predication goes as far back as Aristotle's *On Interpretation*, where he defines it as asserting or denying something of something. In the Aristotelian or *concatenative* approach to semantics, predication is seen as the relation between the subject and the predicate (Stalmaszczyk, 2014). Dominant in traditional grammar, this conception of predicates includes not only a verb but also other elements such as a direct object. Thus, in the sentence *John saw a deer*, *John* is the subject and *saw a deer* is the predicate. Frege's *functional* approach to predication in the late 19th century was a major breakthrough in the study of logic. Frege (1884) asserted that it is only in the context of a proposition that words have any meaning. Predication, in his view, was the application of a function to one or more arguments. However, this operation is not a mere juxtaposition of the elements. The predicate (or function) combines with the arguments into a self-contained whole, as the predicate contains logical slots or argument placeholders that require filling (Stalmaszczyk, 2014). Therefore, the predicate is seen as an "unsaturated" element that is saturated by its arguments.

The Fregean approach to logic and meaning has become the dominant approach in modern semantics in which propositions are assumed to be structured asymmetrically—with one element functioning as the predicate and one or more elements functioning as arguments. As a result, the example above would often be logically represented as *saw (John, a deer)*. The argument structure of predicates specifies the number and type of semantic dependents cooccurring with predicates of different syntactic categories (i.e., verbs, nouns, adpositions, etc.; Napoli, 1989) in grammatical sentences (Comrie, 2008). Arguments are also believed to take distinct thematic roles, such as *agent* and *patient*, which specify their semantic nature. As necessary complements to the sense of predicates, arguments are often contrasted with adjuncts which only provide additional information.

Predication also features in linguistic pragmatics, particularly in *speech act theory*. Predication and reference are viewed as two species of speech acts which respectively yield predicates and referring expressions (Reboul, 2001). These speech acts are in turn varieties of *propositional acts*, conjointly producing propositions in

utterances. Searle (1969) sets out the conditions and rules under which, he believes, these speech acts are accomplished. In short, while referring, the speaker picks out or identifies some object for the hearer which she then goes on to say something about. Predication, on the other hand, is performed when the speaker utters an expression of a type or category about an object such that the expression can be true or false of the object. Predication thus raises the question of the truth or falsity of the expression about the object.

The notions of predication and P-A structure are utilized in other contexts, such as syntax and metaphysics (Stalmaszczyk, 2014) that are not directly related to meaning. Furthermore, one could think of mental predication (e.g., in thoughts and judgements) which does not have a public nature. Here, however, I am mainly concerned with the expression of P-A structure in verbal and non-verbal communication. In this chapter, I will mostly draw on the *ascription view* of predication (Liebesman, 2015). According to this view, what is special about predicates is *how* they designate meaning, rather than *what* they designate. Whereas arguments contribute their meaning through referring to something, predicates do so through ascribing a property to something. So, reference is a dyadic relation between an argument and its referent, while predication is a triadic relation between the predicate, the ascribed property, and the referent. My task therefore would be to investigate the ontogenetic and phylogenetic origin of this asymmetrical semantic distinction between predicates and arguments in human communication.

The P-A structure appears to be a unique property of human communication (Bickerton, 2007; Krifka, 2007). Non-human animal signals are best seen as holistic with regards to their “meaning” (but see Zuberbühler, 2020), which makes them semantically equivalent to whole propositions (Bickerton, 2007). This means that instances of animal signaling cannot be decomposed into predicates and arguments. Besides, animal signals are largely directed at adaptive responses, as opposed to inducing any representation in the mind of the receiver. Thus, whether this proposition-like function is cognitively represented is not obvious. In linguistic communication, on the other hand, P-A structure seems inevitable. We could imagine a “language” in which utterances would immediately result in a behavioral response in the recipient (e.g., uttering “Food!” would result in the provision of food) and/or utterances would work similarly to interjections (e.g., “Bye!”). This language would not require a P-A distinction, yet it will not be nearly as productive as human language. An obviously

non-adaptive language would be one in which either utterances are only referential (e.g., someone utters “This!” without any way of figuring out what is predicated of the referent), or there is no way to determine the reference (e.g., uttering “Good!” without any evident referent). But clearly in natural language, even in cases where either the argument or the predicate is omitted in utterance, predicates and arguments could be identified based on the context. Thus, P-A structure is a prerequisite of linguistic communication—so much so, it has been suggested, that even one-word utterances of the protolinguistic hominins may have involved P-A structure (Dessalles, 2007; Hurford, 2012).

While the question of reference has been researched and debated rather extensively in the literature on the evolution and development of language and other communicative systems (e.g., in the study of *functional reference* in animal communication; Wheeler & Fischer, 2012), predication has received much less attention. However, given the complementary nature of these components of communication, they should be studied together. Indeed, the productiveness of our communication is largely due to *what* we can say about various things rather than *what things* can be the target (i.e. the referent) of communication. Below, I will first overview the relevant work on the origins of predication and P-A structure and then set out my view on the matter.

5.2. Previous Work on the Origin of Predicate-Argument Structure

James Hurford (2003, 2006, 2007) has written extensively on the evolutionary precursors of P-A structure. For him, there is an analogy between the P-A distinction of the form *PREDICATE(x)* in first order predicate logic and the neural mechanisms involved in primate perception. He notes that there exist two independent neural pathways for vision: one mechanism, the dorsal or the “where” stream, provides information for directing attention to the egocentric location of an object and enabling some motor responses to it; the ventral or the “what” stream, on the other hand, provides more detailed information to enable a judgement about what kind of object the animal is dealing with (Mishkin et al., 1982). Thus, the dorsal stream assigns mental indices to objects, tracking them in the scene, while the ventral stream engages in the slower process of assigning properties to those objects. According to Hurford, this neural division of labor forms a natural evolutionary platform for the representation of meaning in which the working of the dorsal stream corresponds to the argument

variable in the above formula and the ventral stream serves the function of predication of categories. Beside these two systems, Hurford (2006, 2007) believes the typical upper limit on the degree of predicates (i.e., the number of arguments they receive) can be explained by the limit on the objects that can be tracked in visual attention at once and the items that can be held in working memory. The limit in both cases is around 4. Hurford's theory does not go very much into the details of how the hypothesized P-A structure in perception might evolve into the expression of predicates and arguments in human *communication*. After all, animals are endowed with rich perceptual and conceptual capacities that they share with humans. Nevertheless, these capacities are not automatically available to their communication. Thus, one must also explain how a structure as integral to human language as P-A structure is *expressed* in communication. This is also true for theories proposing that a P-A structure exists in prelinguistic cognition or a "language of thought" (Gallistel, 2011; Quilty-Dunn et al., 2022). Predication in human communication involves, on the one hand, the intentional (i.e., goal-directed and voluntary) ascription of properties to arguments in their behavior, and, on the other, the comprehension of such expressions. As a result, ascribing properties in cognition does not directly translate into the use of communicative means to predicate a property of an entity.

Seyfarth and Cheney (2011, 2018) appeal to the interaction of signal comprehension and social cognition in primates to account for the propositional content of human sentences. They note that the vocal production of non-human primates is very limited compared to that of humans. For instance, the acoustic features of their signals are adult-like in infancy and are only modified during development. Furthermore, the usage of these vocalizations is largely determined by genetics. Although infants initially make mistakes in their signaling, these mistakes are not entirely random—for example, an "eagle alarm" might be given to birds like pigeons that pose no danger, but not to terrestrial mammals. Based on these observations, Seyfarth and Cheney conclude that the flexible vocalizations typical of modern humans must have evolved after the human lineage diverged from the common ancestor of humans and other great apes. Unlike their production, the comprehension of non-human primates reveals similarities that, Seyfarth and Cheney claim, might be due to evolutionary continuity based on the principle of parsimony. For instance, whereas their calls are rather inflexible, they are capable of learning, and reacting to, the acoustically different alarm calls of other species. One compelling

example of comprehension capacities is that of baboons. Baboon groups are organized around a hierarchy of matrilines in which all members of a matriline outrank or are outranked by all members of another. Rank reversals within a matriline affect only the two individuals involved. However, rank reversals between two individuals from different matrilines affect other members of those matrilines as well, leading to a rise in the rank of most individuals from the lower-ranking matriline. As they are capable of recognizing the identity of individual signalers and each call occurs in a particular social context, baboons can acquire quite specific information from the calls they hear in their surroundings. The calls include threat-grunts which are vocalizations emitted by higher-ranking individuals and screams which are submissive calls emitted by lower-ranking individuals. Sequences of calls, like threat-grunt followed by scream, are therefore informative of which individual is threatening which. And as evidenced by playback experiments, baboons show surprise at sequences that violate the existing hierarchies. Baboons' ability to extract such a social narrative from a sequence of sounds reveals thus a rich cognitive system that can combine discrete elements in a rule-governed, open-ended manner to create a representation of social relations. And importantly, such social knowledge has a propositional format for it involves the recognition of intentions and causality. For example, they could interpret a stream of sounds as "Sylvia is mad at Hannah." Consequently, "the propositions that are expressed in language did not originate with language. They arose, instead, because to succeed in a social group of primates one must understand an elementary form of propositional relations" (Seyfarth & Cheney, 2011, p. 68). Thus, we could say that the earliest P-A constructions were chiefly due to a sophisticated social cognition. The main problem with their theory, as with any other theory that claims continuity between animal signaling and human communication, is that the mechanisms involved are so fundamentally different that such a claim would create more difficulties than it would solve. Non-human animal signals are not the equivalents of human words. Functionally, they resemble the propositions communicated by whole utterances. And developmentally, they are largely innate. Humans, on the other hand, possess a communicative system that is capable of learning and combining arbitrary, symbolic conventions of any community in which they are born. Thus, to argue for continuity in communication, one would need not only to explain how the human-specific capacity evolved, but additionally explain how the comprehension and production of distinct,

utterance-like, innate signals took the form of fully learned symbols processed and combined in a single cognitive system.

Nehaniv (2005) suggests a more specific road map for the emergence of predication. In his scenario, the process started with simple associations. Initially, signs (spoken or gestural), which were produced in temporal vicinity, had their referents associated in an unspecified way. In a subsequent step this association became more specific: one sign, for instance a deictic gesture, selected a target of joint attention and another commented something about it. So, a *topic-comment* structure emerged. Eventually, predication arose when not only there was an association between topic and comment, but the comment conferred to the topic a labelling category (e.g., "This - food"), a property label (e.g., "This - bad"), or a semantic action-role (e.g., "This - eat"). Agreeing with this scenario, Krifka (2007) puts forth a more detailed theory for the emergence of the topic-comment distinction in language use. This distinction, seen as a linguistic universal, concerns the information structure of sentences. Topics refer to entities that have already been mentioned in the discourse or are supposed to be part of the common background knowledge, while comments provide new information about the topic. Topics are often pronominalized and expressed in a prosodically weaker way, and they tend to occur earlier in sentences—commonly overlapping with subjects. Krifka finds a link between this distinction and the differential contribution of the dominant and non-dominant hands in the performance of manual actions. He notes that whereas often the non-dominant hand holds the object, the dominant hand acts upon it. And like topics, the contribution of the non-dominant hand starts earlier: it prehends the object before the dominant hand can start manipulating it. Additionally, the non-dominant hand is static, while the movements of the dominant hand expend more energy and are relatively on a more fine-grained scale. This difference is further reflected in sign language in which the non-dominant hand maintains the topic and the dominant hand signs new information. Following these, he concludes that "[t]he similarities between asymmetric bimanual coordination and topic/comment structuring, and the different roles of the two hands in gesturing, suggest that the manual coordination typical for humans and perhaps higher primates may be a preadaptation that facilitated the development of topic/comment structure in communication" (p. 88). However, although these similarities in the domains of communication and object manipulation are interesting, they will be merely of a metaphorical nature unless the scenario is extended to show

that the first-personal representation of *manual action* is also utilized in interpreting others' manual and non-manual *communication*. (The interpretation of manual action in demonstrations could make this link as I will later explain.) While analogies abound in nature, one must also indicate that they are not incidental but are rather the results of a developmental and/or phylogenetic causal link from the suggested preadaptation to the target adaptation.

This problem is avoided in Rizzolatti & Arbib's (1998) theory of a gestural protolanguage (see also Arbib, 2005). Protolanguage theories are primarily concerned with what language looked like before hominins acquired a full syntactic language. Based on archeological records of sophisticated toolmaking and evidence of brain growth, some scholars assume that the first protolanguage evolved around 2 million years ago (Tallerman, 2011a). Two important questions regarding protolanguage are whether it was compositional from the outset (the majority view) or it was holistic like animal signals, and in what modality language was first used. Most theories find a lexical, spoken language more likely and parsimonious. However, these theories often take too much for granted—such as the voluntary control of vocal expression. And they seem to assume that protolanguage has disappeared in modern humans (Fitch, 2010). Gestural protolanguage theories, on the other hand, posit (based on the "principle of the conservation of previous gains"; Donald, 1991) that protolanguage is still with us in the early-developing gestures which often accompany speech. What is more, intentional gestures seem to be within the reach of non-human great apes (Tomasello, 2008). Like Hurford, Rizzolatti & Arbib (1998) suggest a prelinguistic semantic representation rooted in the neural structure of primates. A class of neurons, named *mirror neurons*, in the F5 area of monkeys' brain has been found to discharge not only when the monkeys manipulate objects, but also when they observe others doing the same action. Area F5 is seen as the homolog of Broca's area in the human brain which is believed to be devoted to processing speech and language. Thus, these neurons serve to represent actions in self and others, thereby enabling the necessary link in communication between the sender and receiver. In normal circumstances the observer is inhibited from performing the observed motor behavior. Nevertheless, sometimes a brief prefix of the movement is produced, which might affect the behavior of the original actor. And noticing this change in the actor, the observer might then act upon it. Through this process, a "primitive dialogue" is established between the observer and the actor which, Rizzolatti and Arbib suggest, forms the core of

language. Of interest to the emergence of P-A structure, they propose the existence of a prelinguistic grammar of action. They believe that the activity of "canonical" F5 neurons provides a code for an imperative case structure (similar to an imperative sentence), in the form of *command: grasp-A(raisin)*, where grasp-A is a specific kind of grasping applied to a raisin. The firing of mirror F5 neurons, on the other hand, codes for a declarative case structure, say, *declaration: grasp-A(Luigi, raisin)*. The mimetic system of communication emerging out of these representations, may have provided a closed system with a fixed repertoire that evolved into an open gestural communicative system—paving the way for the later evolution of verbal communication. As a gestural theory, theirs faces the same questions as other gestural theories. Primarily, a hurdle to these theories is explaining why and how a gestural modality was largely substituted by an auditory modality, since for every argument in favor of spoken language (e.g., freeing the hands), one could come up with one in favor of gestural language (e.g., communicating without attracting predators; Fitch, 2010). Furthermore, non-human ape gestures are mostly innate (Byrne et al., 2017) and holistic, posing them to the same difficulties as other continuity theories face. Rizzolatti and Arbib's theory, is further undermined by the discovery of "auditory mirror neurons" which weaken the argument for a gestural modality based on the presence of mirror neurons in monkeys (Fitch, 2010). And besides, the contents that the mirror system can represent are limited to bodily actions, leaving the communication of more abstract concepts unexplained.

Luuk (2009, 2012) has a different approach to the origin of P-A structure. He takes nouns and verbs to serve as arguments and predicates[1], respectively, arguing that the P-A distinction is a universal among languages of the world. He speculates that there are two possibilities regarding the nature of the first grammatical categories. They could have involved word types, such as nouns and verbs (as in *man go*), or semantic roles (as in *man forest* meaning "a man goes to the forest"). However, due to the transparency of interpretation and the potential for combination, distinct word types might be a more plausible alternative. Luuk makes several arguments in favor of the proposal that linguistic arguments are evolutionarily more fundamental than predicates and evolved earlier (e.g., predicates presuppose arguments they act upon, nouns develop earlier in development, etc.). Yet, the complementary contribution of

[1] In many semantic theories nouns can be predicates too (see Napoli, 1989).

predicates and arguments in language and communication makes their independent evolution highly implausible. It must be noted, however, that Luuk's view is mainly concerned with P-A structure as represented in grammar, rather than a semantic-pragmatic feature that may or may not be grammaticalized in languages. A similar approach is that of Gil's (2012) who, in response to the question about the origin of predication, speaks of the possibility of a simple Isolating-Monocategorical-Associational (IMA) language. In this hypothetical language, words contain only one morpheme, there exist no distinct syntactic categories such as noun and verb, and there are no construction-specific rules of semantic interpretation, but rather the compositional semantics is determined via association between the constituents. This kind of language, he believes, must have been the original form of all human languages, and at least one language, Riau Indonesian, comes very close to it. He further notes that this kind of language involves no predication. Predication for him is a composite notion, rather than an atomistic one, consisting of the identification, within a single element, of the two notions of thematic role assigner and head—neither of which are present in IMA. As a result, the sentence *Ayam makan* in Riau Indonesian is roughly equivalent to the meaning CHICKEN EAT, but the thematic role assigned to chicken could be either agent or patient based on the context; and since there is no headedness it could either mean "the chicken that is eating" or "the chicken is eating". Predication in this sense is an emergent phenomenon that is not an integral part of language. This is again only a grammatical notion of predication and clearly does not imply that predication and thematic roles are not expressible in communication.

There has been even less work on the linguistic development of P-A structure in children. The most notable work is that of Özçalışkan & Goldin-Meadow (2005). They studied the productive communication of 40 children at various stages in development, with a focus on their gestural and speech combinations. They were particularly interested in children's gesture-speech combinations. Their analysis showed that infants as young as 14 months already produced gesture-speech combinations, although the nature of such combinations changed over time. By 18 months of age, children produced significantly more "supplementary" constructions in which gesture and speech together conveyed sentence-like meanings. In these constructions, one modality communicated the argument (e.g., pointing at a car or uttering "car"), while the other communicated the predicate (e.g., making an iconic gesture for "washing" or uttering "drive"). They also found that children produced these constructions before

they could produce them entirely within the spoken modality. Therefore, even before the full mastery of spoken language, children are capable of making use of P-A structure with the help of gestures. As I will argue, infants might entertain P-A structure in the *comprehension* of action demonstrations much earlier than they can produce them utilizing conventional languages.

5.3. The Origin of Predicate-Argument Structure in Pedagogical Demonstration

Our open-ended communicative system comes with special cognitive requirements. If communication through fixed signals, evolving independently over long evolutionary time for distinct adaptive problems, is no longer adequate, a cognitive system is to arise for dealing with the comprehension, acquisition, and production of diverse communicative means. This system has to process instances of communication in which both form and content might be novel. The apparent fulfillment of such requirements in humans suggests that likely for the first time in evolutionary history, instead of distinct signals, an animal was endowed with a *naïve signaling theory*[2]—that is, a general concept of communication dedicated to the use of various means and channels of transmitting information.

Experimental studies suggest that this concept emerges very early in infancy. As discussed throughout the dissertation, from birth on infants are attentive and responsive to ostensive signals that inform them of the occurrence of communication (Csibra, 2010). As young as 4 months, they can utilize the ostensively marked words and signs they receive to learn both about the world and the communicative channels themselves (Ferry et al., 2010; Novack et al., 2021). And by 6 months of age, they may understand that speech transmits relevant information (Vouloumanos et al., 2014). The early emergence of such a specialized system might imply that the mechanism behind our communication constitutes a *core cognitive system* (Carey, 2009; Spelke, 2003). Core systems are a set of learning mechanisms that are believed to provide the foundation for the acquisition of knowledge. They are largely dedicated, domain-specific input analyzers—that is, they represent, and compute the representations of, only a subset of entities in the world. The core system for agents, for instance, only deals with agents and not inanimate objects. The claimed

[2] I am not suggesting of course that this concept is theory-like. I use the term neutrally like "theory of mind". Just as we coin naïve physics or psychology in analogy to the scientific fields, we may coin naïve signaling theory in analogy to the field of signaling theory.

communicative concept would thus be dedicated to the analysis of instances of ostensive communication. This concept has at least three components. These include: *ostension* (or marking$_{NN}$) which is dedicated to the recognition of instances of ostensive communication (typically triggered by ostensive signals); *reference* which induces the recipient to expect and identify the referents (or arguments) for communication—that is, what is communicated about; and *predication* which induces the recipient to extract some property that is ascribed to the referents. Consequently, upon recognizing ostension, a placeholder with a P-A structure is generated in the mind of the receiver that leads to the assignment of a referent and a related predicate—together forming a proposition-like content (see also next chapter).

If I am right in claiming that ostensive communication originated in (prelinguistic) teaching, these components might have already been present in children's representation of pedagogical demonstrations. Demonstrations can be understood as a method of depicting, as contrasted with describing and indicating (Clark, 2016). In describing, things are denoted by arbitrary symbols (e.g., words), while indicating makes use of indexical signs (e.g., pointing) to locate and identify things. In depicting, instead, the communicator is proposed to create one physical scene to represent another scene. Thus, when depicting, people stage physical scenes for others to imagine the scenes depicted (Clark, 2016). Nevertheless, what is depicted or demonstrated is often not another physical instance of the same scene. As many studies, especially in the framework of natural pedagogy theory, suggest, preverbal infants can use demonstrations to acquire generalizable information about the world that goes beyond particular entities.

In a modified view on the transfer of generic knowledge in non-verbal communication, Csibra and Shamsudheen (2015) suggested that in object-directed demonstrations, infants might conceive the manipulated object as an exemplar of the whole class of objects of the same kind. This way, an object, similarly to a word, plays the role of a symbol that stands for something else. Indeed, we rarely use objects to symbolize other specific objects (cf. DeLoache, 2004; Tomasello et al., 1999). Rather, we use objects symbolically as tokens of their kind. For instance, we comment on a particular food to say something about the kind of food it belongs to. Or we present an object (say, an egg) as an ad hoc symbol to request more objects of that kind (i.e., eggs). Similarly, infants can potentially learn about the kind through what is communicated about a single exemplar of that kind. Here the particular object is only

the proximal referent of communication, linked to the distal referent—in other words, the object "serves as a sign that forwards the reference to the kind that the object belongs to" (Csibra & Shamsudheen, 2015, p. 696). In typical cases of demonstration, an object transparently represents its respective kind. But this stand-for relation, like in words, potentially permits arbitrary symbol assignment as well. This occurs in pretense, say, when a banana is stipulated as a symbol for the concept of telephone.

Butler & Markman (2012) introduced a novel object to children and demonstrated that it has a magnetic quality. They found that when this was done in a pedagogical context, 4-year-olds showed more perseverance in testing similar-looking non-magnetic objects. One possible interpretation of this finding is that children conceived the magnetism of the object to characterize the whole kind (Csibra & Shamsudheen, 2015). Children might therefore encode the demonstrated property as a property of the kind. In this sense, demonstrations communicate their semantic content in a similar fashion to generics. Generics (e.g., *tigers are striped*) are sentences that express generalizations without the use of explicit quantifiers (Carlson, 1977; S.-J. Leslie, 2008) and are often used in stating regularities, laws, and dispositions. They are contrasted with episodic sentences (e.g., *the tiger jumped*) which mention particular events and entities. Although there is ongoing debate as to the precise logical form of generic sentences (S.-J. Leslie, 2015; Liebesman, 2011), one can say that they express general propositions. If demonstrations do indeed convey general facts like generic sentences, they may contain comparable semantic elements, including predicates and arguments[3].

What aspects of an action demonstration play the roles of predicate and argument? Since demonstrations convey facts about object kinds through object exemplars, the object presented in a demonstration seems to play the role of the argument (or rather the argument term if we take the kind to be the referent). Thus, reference, as the act of identifying the target of communication, is twofold in ostensive demonstration. In one sense, the communicative episode refers to the particular object that is acted upon. Yet, in another sense, the communication refers, through exemplification, to the kind that the presented object belongs to. This is similar to nouns' putative reference

[3] The assumption here is that generics have more or less the same semantic structure as episodic sentences. This view is defended extensively by Liebesman (2011).

to kinds in generic sentences. Reference in non-verbal communication is mostly studied in the former sense, especially in research on pointing.

The claim that ostensive communication originates in demonstration could potentially imply that deictic gestures (including holding up or pointing to objects) originate in demonstrations as well. O'Madagain, Kachel, and Strickland's (2019) study could be viewed in line with this possibility. They investigated the use and interpretation of pointing gestures by children and adults. Their results indicated that in pointing the fingers are oriented as though the target is to be touched. Moreover, young children interpreted pointing gestures as if they were attempts to touch the objects, rather than "arrows" picking them out. The authors concluded that these results provide evidence for the hypothesis that pointing originates in touch. Accordingly, they postulated that pointing derives from infants' own explorations. But the very early emergence of the sensitivity to others' pointing in infancy (e.g., Rohlfing et al., 2012) could create a challenge for this proposal. Pointing might, instead, emerge out of their comprehension of action demonstrations, where the denoted object is grasped and/or manipulated by the demonstrator. Pomiechowska and Csibra's (2022) study suggests that a grasping action preceded by infant-directed speech is interpreted by one-year-olds as identifying the referent of communication. Therefore, as demonstrations usually involve ostensive grasping and manipulation of objects, deictic reference and the specification of arguments could emerge (ontogenetically and/or phylogenetically) as a by-product of demonstration.

What about the predicate? In the study by Butler and Markman (2012) the magnetic property of a novel object was shown by placing the object on some paperclips and picking these up. Similarly, Hernik and Csibra (2015) presented 13.5-month-old infants with video clips in which an actor's hands were shown to demonstrate the function of two novel objects, one of which ("banana-peeler") turned an intact banana into a peeled one, while the other ("banana-healer") turned a peeled banana into an intact one. They found that communicative demonstrations induced enduring object-function mappings in infants. What is common in these two studies is that the *action* of the demonstrator (including the action outcome) reveals non-obvious properties of objects. Therefore, the property of the object is "expressed" through the action. Accordingly, actions in demonstrations seem to serve as predicative acts that ascribe properties to object kinds. So, the action-object asymmetry in demonstrations would correspond to the semantic P-A asymmetry in sentences of language: one type of

element (i.e., objects) designates the referents (i.e., object kinds), and another (i.e., the action) designates a property that is ascribed to the arguments.

Although argument designation may depend on contextual inferences, it is a relatively straightforward process. In typical demonstrations, the object can simply be taken as an exemplar of its kind. In perhaps more sophisticated cases of demonstration, objects can also be stipulated to represent unrelated kinds (as in the banana-phone example). Predication is nonetheless a fully inferential process, for there is no fixed way to determine its content. Consequently, it would be beneficial for the receiver to store as many features of the action and the scene as possible, so that these can be later consulted for homing in on the more pertinent features. Moreover, considerations of relevance can assist the inferential process for selecting one feature over another (Csibra & Gergely, 2011; Sperber & Wilson, 1995). There are, however, properties of a demonstration that may be more useful in extracting the content.

Our interpretation of others' behavior is often guided by assuming that it is directed at achieving a goal state (Woodward, 1998). This teleological understanding (Csibra & Gergely, 2007; Gergely & Csibra, 2003) encodes actions in terms of their causal role in bringing about a perceivable outcome that justifies their performance. Moreover, the end state of an agent's (as opposed to non-agentive objects') action is encoded and remembered preferentially over its starting point (Lakusta & Carey, 2015; Lakusta & DiFabrizio, 2017). This interpretive bias may be exploited in demonstration to attribute the resulting state of objects as a relevant property. For example, if the manipulation of an object produces a change (e.g., shaking an object to make a sound), this is taken as stating something about the object kind ("this object kind produces this sound"). Or if an object A is positioned in a spatial relation with another object B, this resulting relation can be construed as a predicate of the two objects (e.g., "A goes *on top of* B"). These predicates, like linguistic ones, can additionally take multiple arguments based on the number of objects used.

As can be seen, the symbolic interpretation of objects enables a productive means of communicating information. Object-symbols can be physically manipulated to reveal an open range of properties and relations, similarly to the lexical ascription of properties and relations in language. If this way of communication was available to our ancestors, they may already have possessed a flexible system that enabled them to produce unlimited messages—constrained of course by the objects at hand and the possible actions on them. (As such, the current proposal permits a more productive

system than Rizzolatti and Arbib's.) Since this communication has a P-A structure, its semantics is compositional in nature (Szabó, 2008), that is, it involves complex expressions whose meaning is determined by the meanings of its constituents (i.e., the predicates and arguments extracted from actions and objects) and its structure (i.e., the relations between actions and objects). This possibility thus features a departure from the holistic communication of other animals and may have provided a platform for the evolution of our linguistic semantics.

5.4. Types of Predication in Demonstration

In natural language, most syntactic categories can play a predicative role. These include most verbs (e.g., *they **left***), predicative adjectives (e.g., *he is **nice***), predicative nominals (e.g., *she is a **doctor***), and adpositions (e.g., *it is **in** the bag*). With the help of these diverse categories, various properties and relations can be predicated of their arguments: one can express actions, attributes, categories, and spatiotemporal relations. In a study on imitation, Király, Csibra, & Gergely (2013) found that 14-month-old infants reenacted an arbitrary goal-directed action only in communicative contexts. The demonstration involved an experimenter turning on a "magic lamp" with the head rather than hands. Infants likely construed the arbitrary action as ascribing an action kind (i.e., the appropriate way of handling) to the object. Is the content of predication in demonstration restricted only to the types of action that can be performed on objects? The answer is no. Despite the physical limits of the means through which one can demonstrate, it can convey a wide range of contents.

In the examples above (Butler & Markman, 2012; Hernik & Csibra, 2015), we already saw how the function of a tool or a property of a novel object was predicated. In both studies, the final outcome of the action (i.e., picking paperclips and pealing/healing a banana) was understood as a kind-relevant property of the artifacts. A study by Kovács et al. (2017), moreover, provides evidence that one-year-old infants can construe action demonstration as predication of a category. In the training phase, infants saw videos of an actor sorting cups and plates into two locations (left or right) so that one category of objects was always transferred to the right and the other always to the left. Then, the actor demonstrated that an object that resembled a plate could be transformed into a cup. At test, the ambiguous object was presented in the plate format and infants' anticipatory looks to the cup or plate sides were measured. Infants who had observed the demonstration performed in a communicative context made

more anticipatory looks to the cup side. Thus, despite the superficial appearance of the object, infants had interpreted the demonstration as conveying that the object is indeed a cup. The authors conclude that ostension makes infants expect the demonstration to reveal kind-relevant information about the referent and so they interpret the final state of the act as the predicate of the message. Therefore, the action of turning a plate into a cup can be viewed as an act of predication, whose resulting outcome reveals an essential property of the object.

As previously mentioned, it has been postulated in natural pedagogy theory that, since ostensive communication is interpreted generically by infants, kind-relevant information might be encoded at the expense of episodic features that are extrinsic to objects. Yoon et al. (2008), for example, show that in ostensive contexts infants are sensitive to the change in the identity of a pointed-at object, ignoring location changes—a bias that is reversed in non-communicative contexts (see also Okumura et al., 2016, 2020; Thiele et al., 2021). However, as pointing mostly serves to pick out the argument, rather than predicate anything of it, it might be the case that location is not regarded as relevant by infants in this context and is treated as an "adjunct". Conversely, if the spatial information is part of the predicate, they might encode and store it as a property of the object kind. A study by Topál et al. (2008) could be informative in this regard. They investigated the well-studied perseverative search error (also known as the *A-not-B error*) in 10-month-old infants. This error occurs when children who have repeatedly retrieved an object from a location continue to search the same location even when they see the object being hidden at another location. Although often attributed to infants' deficit in inhibitory control, Topál et al. show that this error is significantly reduced when the hiding takes place in a non-communicative or non-social context, as opposed to a communicative context. Unlike in the pointing study by Yoon et al., here the location is the result of the action (i.e., hiding) on the object. Thus, if the present account is right, it should be represented as part of the predicate and stored as a property of the object kind (and so it is not dependent on the social partner who performs the hiding: Topál et al., 2009). In this light, it might be possible to convey spatial properties and relations in demonstration, similarly to adpositions in language.

Although they typically involve physical manipulation of the object tokens to communicate the content, demonstrations are not limited to manual actions. The presence of P-A structure allows one to target the already "meaningful" behaviors in

our repertoire at deictically identified entities to convey various related meanings. As such, there may be no need to assume the evolution of an intermediate gestural stage between demonstration and verbal communication, since the structure can be coopeted in various modalities. By emitting ostensive signals, an agent can in principle render any action that is normally seen as egocentric or in a transient relation with stimuli into public acts of generic communication. One such meaningful behavior is emotional expression. Egyed, Király, and Gergely (2013) tested whether infants would interpret ostensive, object-directed emotion displays as conveying culturally shared knowledge about objects. The experimenter showed a positive emotional expression towards one of two objects and a negative emotional expression towards the other. At the test phase a *different* experimenter requested infants to give her one of the objects. Infants chose the positively valenced object more than the negatively valenced one in the communicative condition, as compared with the non-communicative condition. This pattern of results shows that infants construed the emotional expression as predicating positive or negative attributes of the object kind—attributes which would remain unchanged from one context to the next. Thus adjective-like attributes are also expressible in ostensive demonstrations.

Aside from this content-based classification, predicates in demonstrations could also be classified based on the form of the action that is utilized in conveying the message. These various forms have already been mentioned in the discussion on content types. There is a good reason for this, as these predicative acts convey their content *iconically*. This means that the perceptual features of the medium of representation are mapped onto the content and so the meaning can be extracted from the form in the absence of any shared code. Thus, one type of predicates are spatiotemporal predicates, which carry their meaning through the demonstrator's positioning of objects in space and sequencing the actions in time. Through moving and manipulating objects, the demonstrator can show the functioning of a tool, the proper spatial relation of two objects ("object A goes inside object B"), or the proper sequence of actions ("action A is done before action B"). A second form of predication is through meaningful, coded behaviors. If the recipient already understands the function of certain behaviors, once they are embedded in ostensive communication, they could be conceived as conveying kind-relevant information about the object. Emotional expressions are among these. But it is likely that even the fully learned behavioral repertoire (e.g., coughing after eating spicy food) could be exploited by the

demonstrator to convey object-related knowledge (e.g., “this kind of food is spicy”). This form of predication has an indexical aspect, in that there is a causal relation between the behavior and its meaning: the predicate derives its meaning from preexisting relations between the behavior type and types of stimuli. However, since it is to be interpreted as a *demonstration* (rather than an ordinary incidence) of the behavior, this form of predication has an iconic nature too. That is, the demonstrated emotional expression is understood in terms of its resemblance to actual expressions and what they imply. Lastly, as suggested in the previous chapter, labels could be viewed as verbal predicates: despite carrying no information in themselves, these “empty” predicates can utilize the kind-generalizing disposition to specify the extension of the object kind. Hence, they can be used to group together objects whose category membership is not evident based on their perceptual features.

5.5. From Predicate-Argument Structure to a Spatiotemporal Syntax?

If the present account is correct in suggesting that simple demonstrations possess a P-A structure and this was available in early stages of the evolution of ostensive communication, hominins might have been capable of a compositional protolanguage that utilized the physical manipulation of object-symbols and the teleological understanding of instrumental actions to transmit generic predicates. This depictive system uses objects as props, whose properties and relations could carry various kinds of proposition-like content. Of course, what could be communicated through this system is limited compared to full-fledged language. For instance, the P-A content could not be negated or quantified. Thus, many features of linguistic communication are absent in such a system and may hinge on the biological and cultural evolution of more sophisticated symbolic and interpretive capacities. But the productive nature of the system provides the conceptual foundations for more enhanced methods of communication.

Did this system also possess syntactic properties? The hypothesized depictive communication has a prespecified way of assigning elements of a demonstration to the argument and predicate roles. Therefore, one may say that it is a rule-governed system. However, syntax does not comprise only rules of combining elements (Tallerman, 2011b). Most importantly, syntactic rules combine semantic units *hierarchically*. Although languages drastically diverge in their syntactic features, the hierarchical organization of elements is viewed as a core linguistic universal present

in all languages (Bickerton, 2011). Hierarchical representation is suggested to be enabled by the single operation *merge* that takes two elements *a* and *b*, and puts them together to yield the set {a, b}. The output of this operation can then be merged recursively with other elements, creating yet more complex expressions (Bolhuis et al., 2014). Thus, merge produces a *discrete infinity* (Hauser et al., 2002) of hierarchically structured expressions—believed to be at the center of our linguistic productivity. The hierarchical nature of language can be observed in phenomena such as structural ambiguity (i.e., the same sequence of words can be interpreted differently based on their syntactic relations), displacement (e.g., question-words like *what* appear to "move" within the sentence, leaving behind a trace), and long-distance dependency (e.g., in subject-verb agreement). It is commonly assumed that our protolanguage lacked a hierarchical syntax and relied instead on semantic-pragmatic rules (e.g., an agent-first rule) for identifying the contribution of elements to the meaning of the whole construction (Jackendoff & Wittenberg, 2017; Tallerman, 2011a). Nonetheless, this assumption may not be necessary if language is seen as emerging from an already hierarchical system. An action syntax (Maffongelli et al., 2019; Pulvermüller, 2014) is one such candidate.

The demonstrations discussed above involve relatively simple action sequences that do not necessarily demand the organization of elements beyond their semantic contribution. However, our goal-directed actions often require the achievement of multiple steps or sub-goals in the process of arriving at the final goal state (Csibra & Gergely, 2007; Maffongelli et al., 2019; Vicari & Adenzato, 2014). Reaching a goal such as making coffee or cooking a dish needs to be done in multiple complementary lower-order acts that not only follow a prespecified sequence but are understood hierarchically (Csibra, 2007a; Pulvermüller, 2014)—with each level in the action plan comprising of smaller necessary elements that are contingent on the preceding elements. Moreover, leaving aside competence constraints, these sequences can potentially embed an infinity of novel action elements (Pulvermüller, 2014; Vicari & Adenzato, 2014). For instance, one can add a new step in a cooking sequence involving new ingredients and processes. It is proposed that the Chomskyan minimalist syntax can be fruitfully applied in the analysis of complex action as well (Pastra & Aloimonos, 2012). The constituents in this syntax would include, among other things, action primitives (e.g., enclose hand), tool complements (e.g., knife), and object complements (e.g., apple). These elements, it is argued, can be combined or

merged recursively to yield increasingly complex sequences (e.g., “extend $hand_1$, grasp with $hand_2$ apple, grasp with $hand_1$ knife, reach with knife apple, slice with knife apple”). One may even talk of structural “ambiguity”, when the same action sequence could be conceptualized as a mere succession of simple actions or a structured complex action. The latter type of action may create difficulties for agrammatic children (Roy et al., 2013). And as discussed in the previous chapter, the link between action understanding and syntax is backed by evidence showing an overlap in their underlying neural mechanisms (Stout & Chaminade, 2009, 2012). Additionally, while the early hominin tools might have been possible to produce by successive application of processes, later tools employed skills (e.g., in the Levallois technique) that seem to have required hierarchical representations of the steps (M. Moore, 2010). For example, adjusting the platform for later removal of high mass may have involved a human-unique *subassembly* strategy (as opposed to the non-hierarchical *pairing* and *pot* strategies)—comprising “anticipatory” flakes for enabling subsequent “objective” flakes. Pain (2022) likewise argues that language coopts the mechanism for structured representations that initially evolved for computing complex lithic toolmaking. In sum, an action syntax might have generated the necessary preadaptions leading to linguistic syntax.

Moro (2014a, 2014b) notes that there are no equivalents of words, functional elements (e.g., *the*, *if*, *not*, etc.), and dependence in action representation. Thus, he criticizes the analogy between action and syntax as drawing on an evolutionary metaphor in the two domains. There have been attempts to find some of the seemingly language-specific elements in action representation too (Pastra & Aloimonos, 2012; Vicari & Adenzato, 2014). But some scholars argue that it is an unproductive approach to seek an identity between the domains: descent with modification might entail inevitable alterations in the mechanisms (Boeckx & Fujita, 2014; Pain, 2022). Notwithstanding, the evolutionary link between action and syntax could be made more directly if we postulate that human communication originated in action demonstrations (see also Kolodny & Edelman, 2018; Stout & Chaminade, 2012). The existence of a demonstration phase means that representing actions was done in the context of communication. If we have two domains with similar neural mechanisms (i.e., the Broca’s area), similar computational demands (i.e., hierarchical representation), and similar functions (i.e., open-ended communication), the link between the two is no more a mere metaphor. The relative complexity of toolmaking and the increasingly

longer sequences of action might have driven selection for more efficient means of representing the necessary processes. Hierarchical representation can utilize finite elements and rules to compress the data (Bolhuis et al., 2014; Dehaene et al., 2022). Acquiring skills from others' demonstration, on the other hand, enables the application of the computations in an ostensive communicative domain. The hypothesis that objects in demonstration can act as symbols provides the equivalents of lexical items. Both types of entities are in a symbolic, stand-for relation with their represented content, both are discrete elements, both can be manipulated in various combinations to produce complex expressions, and both can be ascribed predicates. This *spatiotemporal syntax* can exploit the spatial properties of object-symbols as well as the temporal sequence of events for transmitting kind-relevant knowledge. Dependences and some of the functional elements like determiners and pronouns are not to be found in demonstration because the iconic nature of demonstration relies on perceivable relations between items—not possible in a conventional language. The generic property of demonstration also does not necessitate many functional elements like quantification and tense. The hypothesized extension of communication from teaching to other domains involving episodic properties may have called for the establishment of arbitrary symbols both for non-manipulable entities and abstract relations that could not be conveyed iconically. If a spatiotemporal syntax is in place, the emerging conventional protolanguage need not necessarily be structureless as it can be embedded into the already existing structured representational devices.

5.6. Conclusion

As we saw, the division of labor between action and object in demonstration is akin to the semantic role of predicates and arguments in language. Human communication engages a concept that assigns a P-A structure in the interpretive process even in non-verbal modalities where the distinction is less evident. Triggered by ostension, this mechanism identifies a referent and a related predicate to form a proposition-like content. In demonstration, while objects constitute the referents, actions directed at them serve to ascribe an enduring property or relation. This creates a productive system of communication with the capacity to generate limitless messages using limited resources.

The current theory avoids many of the shortcomings of alternative theories that were discussed above. Importantly, it is based on empirical findings on the early

development of comprehension in ostensive communication. It relies, for instance, on capacities that emerge months earlier than the production of P-A structure in gesture-speech combinations as presented by Özçalışkan & Goldin-Meadow (2005). This early ontogenetic emergence could be an indicator that the P-A structure of full-fledged adult language taps into the already functional capacity for interpreting ostensive demonstrations. As an evolutionary theory, it has the merit of arguing for traits that are still present in modern humans, rather than speculative stages in the evolutionary past that have disappeared from the phenotype. Being grounded in the domain of communication, it is possible to link this cognitive capacity, both in ontogeny and phylogeny, to more sophisticated linguistic capacities. This is more difficult for theories that rely on non-communicative preadaptations. Consequently, the theory avoids the pitfall of evolutionary metaphors. Furthermore, some of the implications of the present theory can be tested empirically. For instance, one can test whether the presentation of objects leads to better encoding of the relations among them (e.g., “A goes on top of B” versus “B goes on top of A”) if it is done in an ostensive context. Likewise, one can investigate whether the hierarchical structure of complex actions performed ostensively is more efficiently encoded and remembered according to rule-like representations.

This theory could potentially shed light on some of the peculiarities of language to boot. One such peculiarity pertains to word order in languages. Of the six possible orderings of subject, object, and verb, subject-verb-object (SVO) and subject-object-verb (SOV) comprise the majority of languages of the world. Experiments by Langus and Nespor (2010) show that both in the production and comprehension of improvised gestures SOV is preferred, independently of the structure of the native language of the participants (see also Goldin-Meadow et al., 2008). The OV sequence is also predominant in homesigning (Goldin-Meadow & Mylander, 1998). In the spoken modality, on the other hand, SVO elicits a shorter response time. It is believed that SVO is related to computational requirements of language (Langus & Nespor, 2010). But the cognitive origin of the SOV word order is rather obscure. The present account can perhaps elucidate this. As described by Krifka (2007), in object manipulation first the object is prehended (with the non-dominant hand) and then the action is performed on it. If this holds for action demonstration as well, the audience will need to encode the object before the action. Assuming a common origin for language and communicative demonstration, the OV sequence could be seen as a bias for encoding

arguments before predicates in demonstration—which is manifested in languages as well.

Another potential "living fossil" (Bickerton, 1990) regards the developmental trajectory of the acquisition of different word types. Nouns often precede verbs in development (Waxman et al., 2013). There are different, albeit compatible, hypotheses as to why nouns are universally learned earlier and understood more easily than verbs. One theory asserts that objects are perceptually and conceptually more stable and less fleeting than actions or events which involve relations among those objects (Gentner, 1982). Alternatively, the dependance of verb meaning on the arguments might require infants to establish a repertoire of nouns before learning verbs. If the present theory of predication is correct, actions (including verbal utterances) are the main medium of predication, and objects prototypically function as arguments. Therefore, mapping a noun to an object kind falls neatly into this format—with the noun expressing the predicate and the object expressing the argument. However, the articulation of a verb (the form) as an action is typically mapped onto another action (the meaning) which the verb expression is supposed to denote. To simplify, a predicate (i.e., the verb) is ascribed to another potential predicate (i.e., the action). Thus, the comprehension of verbs may involve a second-order predication and, as a result, understanding the use of verbs to designate predicates could be relatively more cumbersome, as it deviates from the hardwired organization of the components of ostensive communication.

The current chapter's emphasis that non-verbal demonstration, like verbal utterance, possesses a P-A structure and uses perceptual stimuli (i.e., objects and actions) to represent and communicate displaced contents may have far-reaching consequences for our understanding of the human concept of communication. Contrary to dominant accounts, I will argue next that a metarepresentational capacity for dealing with such public or external representations may have explanatory priority over representation of mental representations.

Chapter 6. Metarepresenting in Communication

6.1. Introduction

Humans are unique in their metarepresentational capacities (Cosmides & Tooby, 2000; Sperber, 2000; Suddendorf, 1999). Metarepresentation allows humans not only to represent the world directly, as other animals do, but also to represent, reason about, and learn from other representations. These embedded representations could be of a mental nature. Metarepresenting mental representations enables predicting and explaining the behaviors of other organisms, typically by postulating beliefs and desires that could act as the causal drivers of those behaviors (Dennett, 1987). It was in the context of such "theory of mind" (Premack & Woodruff, 1978) capacities that the term metarepresentation gained currency (Pylyshyn, 1978). It is, thus, no wonder that metarepresentation is often taken as synonymous with theory of mind. Yet, undeniably, external representations, too, can be objects of representing minds, and humans are no less skillful at those. From markings (Ittelson, 1996), animations (Revencu & Csibra, 2021), and other types of depiction (H. H. Clark, 2016) to the more regularly used words and gestures, we exploit perceivable means to convey relevant information to one another. Both types of metarepresentation develop early and are instrumental in our social lives.

Another unique feat, ostensive communication is widely acknowledged to be linked to metarepresentation (Malle, 2002; Woensdregt & Smith, 2017). Since Grice's landmark paper (1957), *intentionalism* (Harris, 2019) in its various forms became the dominant account of human communication in philosophy, linguistics, and cognitive science. According to this account, communication largely consists in the expression and attribution of communicative intentions by interlocutors. Mental metarepresentation was then expectedly taken to be at play in communication, explaining both its development and its evolution (Scott-Phillips, 2014; Sperber & Wilson, 2002). But this approach has proved to be fraught with difficulties, mainly having to do with the complexity of the cognitive processes it accredits to infants and ancestral hominins. According to an opposing camp (e.g., Geurts, 2021), on the other hand, this picture takes things reversely: it is language and the capacities it affords that permit the development and evolution of mental metarepresentation.

Therefore, a central question in the study of human communication is the causal direction between mentalizing, on the one hand, and language and communication, on the other. What is missing in this debate is the role of representations of a public nature. Whichever stance we take on the ontogenetic and phylogenetic origins of ostensive communication, we must offer an account of the metarepresentation *in* communication, namely the necessary employment of external representations, whether conventional or ad hoc, for conveying information. This neglect, I will argue in this chapter, leads to problems deeper than mere cognitive load and complexity. After reviewing the predominantly Gricean approaches, the different roles they impute to metarepresentation in human communication, and the common problems that stem from them (§6.2), I discuss how in many contexts higher-order metarepresentation might not be necessary (§6.3) and, more crucially, why it is not sufficient in explaining the emergence of ostensive communication (§6.4). Then I sketch out an alternative account of the developmental and evolutionary origin of human communication according to which ostensive communication involves a primitive concept that is irreducible to a configuration of mentalistic propositional attitudes and requires instead a capacity to identify external representations and their detached content (§6.5). Lastly, I speculate that, if ostensive communication does not necessitate mentalizing, metarepresentation might have evolved to enable the use of communicative, external representations (§6.6).

6.2. Metarepresentation in the Gricean Approach

In his 1957 paper, Grice noted a fundamental difference between our uses of the word "meaning". For example, when we say,

(1) The smoke means fire.

we are concerned with a factive sense of the word, to the effect that "*x* means *p*" entails *p*. If smoke means fire, then what we have is indeed fire. He labeled this use of the word *natural meaning*. In contrast, when we say,

(2) Her utterance means "there is fire".

we are using the word in an entirely different sense. Here we cannot unproblematically conclude *p* from "*x* means that *p.*" Such non-factive uses involve what Grice termed *nonnatural meaning* (meaning$_{NN}$, for short), and this is the sense that theories of human communication must address. Grice's first proposal for analyzing meaning$_{NN}$ was that "*x* meant$_{NN}$ something" could be taken to be true if *x* was intended by the

utterer U to induce a belief in an audience. (Note that it is here that the Gricean reduction of the concept of communication to intentions and beliefs originates.) Grice notes immediately that this is not sufficient. For instance, if A leaves B's handkerchief by the scene of a murder in order to induce in a detective the belief that B is the murderer, we cannot say that the handkerchief $meant_{NN}$ anything. This is a case of hidden authorship, rather than communication (Grosse et al., 2013; Tomasello, 2008). For communication to take place, the intention behind the act must be overt (Strawson, 1964). (Note now that overtness, including the problems associated with it, applies only if one adheres to the above-mentioned reduction.) Thus, "*x* $means_{NN}$ something" would be true if somebody intended the utterance of *x* to produce a certain effect in an audience by means of the recognition of this intention. This seems to involve a self-referential intention (Bach, 1987). It sounds as though the audience should already know what U's intention is in order to recognize it. Likely in fear of such a "reflexive paradox" (Grice, 1957) and following Strawson's suggestion, Grice later (1969) formulated his analysis in terms of "iterative" intentions (Bach, 2012). Therefore,

> U meant something by uttering *x* if and only if, for some audience A, U uttered *x* intending
>
> (a) A to produce a particular response *r*
>
> (b) A to think (recognize) that U intends (a)
>
> (c) A to fulfil (a) on the basis of his fulfillment of (b).

Condition (c) is added to rule out what Grice took to be counterexamples to a simpler formulation: King Herod presents Salome with the head of Saint John the Baptist on a charger intending her to believe that Saint John is dead and intending her to recognize this intention; and someone shows his bandaged leg to convey that he has a bandaged leg (as opposed to, say, he cannot play squash). Grice wanted neither of these to be included by his concept of $meaning_{NN}$, for he believed that in such cases the attribution of intention is incidental to the response—that is, the evidence is enough to make the required inferences. Besides these, Grice's definition gave rise to an industry of counterexamples to the formulation, most notably by Strawson (1964) and Schiffer (1972). I will return to these below.

In developing *relevance theory*, Sperber and Wilson (1995; Wilson & Sperber, 2012) placed Gricean pragmatics within the framework of the emerging cognitive sciences and, in doing so, spelled out most explicitly the representational requirements of ostensive communication. Rejecting the necessity of sub-intention (c) in Grice's

formulation, they suggest that at least two intentions are involved in production and comprehension. The first intention, corresponding to Grice's sub-intention (a), is:

informative intention: to inform an audience of something.

This is embedded in a second-order intention, corresponding to Grice's sub-intention (b):

communicative intention: to inform the audience of one's informative intention.

They point out two potential difficulties in processing these intentions (Sperber, 2000; Sperber & Wilson, 2002): Firstly, the inferences required for interpreting the meaning of an utterance, as envisaged by standard Gricean accounts, seem to demand complex belief-desire reasoning on the part of the audience. In his other seminal work, Grice (1975) suggested that this is done by the expectation that interlocutors observe a *cooperative principle*, allowing the audience to eliminate any interpretation that is incompatible with the principle—a rational process which in turn necessitates sophisticated metapsychology. Secondly, specifying the processes involved in the recognition of communicative intentions reveals several layers of metarepresentation. As an example, when Mary is ostensively picking berries to convey to Peter that the berries are edible, for Peter to grasp fully the communicative act he must entertain a fourth-order metarepresentation (Sperber, 2000):

Mary intends$_4$
Peter should believe$_3$
Mary intends$_2$
Peter should believe$_1$
these berries are edible.

This metarepresentational format is what permits Mary's informative intention to be overt (or "mutually manifest"). Moreover, because of their inherent ambiguity, our utterances underdetermine the intended meaning, that is, there is a systematic gap between what is explicitly said and what is intended by the communicator. This underdetermination would be compensated if one takes the utterance to act only as a piece of evidence for the intended meaning. The audience ultimately forms a metarepresentation of the speaker's meaning in the interpretive process. And yet another reason for such high orders of metarepresentation is that, if there is doubt in the competence and benevolence of the communicator, the audience can still interpret the communication appropriately. In such cases, although the informative intention may not succeed and, as a result, the audience does not believe the content, he will

nevertheless recognize the communicative intention. That is, metarepresentation separates comprehension from acceptance (Heintz & Scott-Phillips, 2022). However, if the communicator is trustworthy, there would be no need for this metarepresentational distance and recognizing the informative intention indeed might give the audience an extra reason to believe the information (Sperber, 1994b).

Metarepresentation enjoys thus a central role in relevance theory; so much so, indeed, that "the very act of ostensive communication, in both production and comprehension, is an exercise in reading others' minds" (Scott-Phillips, 2014). Therefore, the causal direction clearly goes from (mentalistic) metarepresentation to ostensive communication. Consequently, the reason humans but no other great apes can communicate ostensively (but for a recent defense of the possibility of proto-ostension in non-human primates see Sperber, 2019) is mainly a metapsychological advantage (Scott-Phillips, 2015). Being more than a code-based signaling system, language could not have emerged in the absence of the metapsychological capability to attribute higher-order informative intentions. However, the apparent complexity of the necessary cognitive processes, on the one hand, and the ease and early developmental emergence of ostensive communication, on the other, suggest that a sub-module of our mindreading mechanism might have evolved to exploit the regularities in the domain of communication (Sperber, 2000; Sperber & Wilson, 2002). Particularly, this comprehension module is licensed to assume that the utterance is optimally relevant, that is, it is sufficiently relevant to be worth the audience's attention and is as relevant as is compatible with the utterer's abilities and preferences. Then, the audience can follow a path of least effort in constructing hypotheses as to the right interpretation and stop when his expectations of relevance are satisfied. As such, although the output of this process may be attributed as a metarepresentation of the speaker's meaning, the process need not be necessarily metarepresentational.

Recently there have been several other attempts at either simplifying or qualifying the standard formulations to allow that, both in ontogeny and phylogeny, prelinguistic human and non-human animals possess some form of ostensive communication. In one such attempt, Breheny (2006) states the developmental dilemma as a conflict between the assumption that communication involves attributing propositional attitudes like intentions and beliefs and the then widely acknowledged finding that, although capable of adult-like communication, children below the age of four are unable to represent propositional attitudes—as evidenced by their failure to appreciate

others' false beliefs (Wellman et al., 2001; Wimmer & Perner, 1983). This appears especially problematic for developmental accounts that place mindreading at the center of language development (Bloom, 2000; Tomasello, 2008). Drawing on some of the ideas in relevance theory, he proposes that basic communication might not require mentalizing, but rather the ability to recognize instances of communication based on an action concept, and the use of a relevance-guided procedure to identify the referent. But an alternative account of children's mindreading capacities, according to which their failure to pass standard false-belief tasks is due to processing demands rather than a conceptual shortcoming (Fodor, 1992; A. M. Leslie et al., 2004), has received more empirical support in recent years (Onishi & Baillargeon, 2005; Surian et al., 2007; but see Rakoczy & Behne, 2019). Thus, one strategy has been to use these results to deny that the complexity of the postulated communicative intentions poses a fatal blow to a Gricean approach to communication and meaning (Thompson, 2014).

Another strategy for mitigating the inferential complexity is to hold that, while the Gricean approach is generally correct, one can reduce the representations involved in the analysis without missing the main insights. Similarly to Breheny, Moore (2014, 2016, 2018) finds the dependence of the dominant views on a concept of belief problematic, as this could imply that neither infants nor non-human primates communicate ostensively. The issue is partially resolved in directive communication which is aimed mainly at producing an action in the audience. However, informative communication too need not always necessitate changing belief-states: communication can be directed at changing perceptual states such as seeing and attending. As for the metarepresentational complexity of communicative intentions, Moore remarks that even ten-year-old children have difficulty with entertaining fourth-order metarepresentations. Any Gricean account of the origin of human communication must therefore explain how creatures without such sophisticated capacities can nevertheless develop ostensive communication. Moore's response to this challenge is that, following Gómez (1994, 1996), a minimal Gricean account that is metarepresentationally much less demanding is possible if one admits a functional reading of Grice's formulation. According to such a reading, the first clause in the formulation can be enacted by *sign production*, which refers to communicative behavior (e.g., pointing) with the purpose of eliciting some response *r* from the audience. The second clause can then be enacted by functionally separate *acts of*

address. These are performed to direct the attention of the audience to the action (e.g., through eye contact), thus fulfilling the overtness requirement of Gricean communication. This functional separation considerably reduces the number of metarepresentations in the analysis. Adapting the above example, in addressing her gesture to Peter's attention through, say, an attention-getter:

Mary intends$_1$

Peter attends and responds to her gesture.

And in pointing at berries:

Mary intends$_1$

Peter looks at berries.

Therefore, two pairs of metarepresentation would be enough for comprehending such forms of communication. As the communicator does not need to represent her own intention, producing minimally Gricean acts does not involve metarepresentations at all. This formulation may permit the emergence of ostensive communication in preverbal infants and possibly non-human primates. If so, one could be justified in positing that sophisticated forms of metarepresentation evolved culturally from simpler forms by exploiting linguistic tools (R. Moore, 2021).

Another strategy in tackling the complexity of Gricean communication is to argue that there are intermediate cases of communication that share some features with full-blown ostensive communication but are simpler in structure (Armstrong, 2021; Green, 2019; Planer, 2017a, 2017b). If one could identify these intermediate cases, the argument goes, the task of explaining the transformation from largely coded animal signals to mentalistic communication in humans becomes much easier. Thus, while postulating non-Gricean forms of communication as intermediate links, proponents of incrementalism generally take issue more with the Gricean origin of human communication than with the Gricean analysis of meaning into metapsychological propositional attitudes. As will become clear below, I will take the opposite path; that is, while Grice's insights about the explanandum were correct, his intentional analysis does not provide a sufficient account of the emergence of ostensive communication.

To summarize this section, Grice accounted for the distinction between natural and nonnatural meaning by suggesting that human communication involves complex, overt intentions. Such intentions in turn necessitate, and build upon, multi-layered metarepresentations of mental states. The complexity of such metarepresentational inferences creates difficulties for a psychologically plausible account of the emergence

of ostensive communication in ontogeny and phylogeny. These difficulties have usually been addressed by qualifying or simplifying the purported representational demands of communication. In the next two sections, I will discuss the degree and kind of metarepresentations required for a parsimonious account of the concept of communication.

6.3. Higher-order Metarepresentations are not Necessary

The higher-order intentions in Gricean formulations serve mainly to distinguish communicative from non-communicative behavior (e.g., hidden authorship). The audience should, for instance, infer that the communicator wants him to believe that she has a message (i.e., an informative intention) for him. Therefore, most of the complexity of the proposed representations stems from such inferences to the communicative nature of the behavior. However, proponents of the theory of natural pedagogy have argued that human children are innately sensitive to a set of *ostensive signals* which, among other things, indicate to them the occurrence of communicative episodes (Csibra, 2010; Csibra & Gergely, 2006, 2009, 2011). These signals (e.g., eye contact, infant-directed speech, and contingent responsivity) allow caregivers to communicate and teach to infants adaptively-relevant knowledge by exploiting the infants' tendency to interpret such episodes as involving generic information. One of the implications of these signals is that, rather than metapsychologically *infer* the presence of communication, infants can simply *decode* the signals when they detect their occurrence (Csibra, 2010). They would subsequently need to infer only the message behind the communicative act. Although later in development they might also be able to postulate that the communicator has an intention to communicate, this is arguably not always needed. If so, then the requisite metarepresentational inferences in the Gricean analysis would be significantly reduced (see also Moore, 2014). But pedagogy is not the only context in which humans can directly recognize communication. Adult interactions, too, sometimes involve ostensive signals, such as eye contact, for rendering the behavior communicative (e.g., while demonstrating or pantomiming). A functional reading of Grice's second clause, which (contra Moore, 2016, 2018) takes into account its role in distinguishing communication from non-communicative behavior, leads us to define ostension as the marking of stimuli as communicative (see chapter 2). This marking$_{NN}$ permits the function of ostension to be fulfilled without necessarily drawing on metapsychological inferences. Decoding

ostensive signals, as discussed, is one mechanism for fulfilling this. However, the "communicative presumption" (Bach & Harnish, 1979) is at work whenever we use an established channel of communication. These channels, like the spoken and gestural modality, might be developmentally established through their co-occurrence with ostensive signals. In effect, in many, if not most, contexts, we can take the presence of communication for granted without needing to call on complex inferences. Thus, in development and evolution, humans could have utilized ostensive signals to mark$_{NN}$ their actions both for ad hoc communication and for establishing communicative channels—thereby bootstrapping the emergence of ostensive communication. The existence of specialized ostensive signals can ultimately simplify the Gricean analysis of communication into something like the following:

Mary (ostensively) intends$_2$

Peter believes$_1$

these berries are edible.

What is needed for this simpler formulation to work properly is that, although temporally and procedurally separate, the decoding of ostensive signals is conceptually linked to the interpretation of the content of communication. Additionally, a concept of ostension or communicative intention might be required (Csibra, 2010).

The separation of the functional counterparts of Grice's clause (a) and (b) might be not only a *possibility* but a *necessity* for the emergence of ostensive communication. This is because embedding the informative intention within the communicative intention, as the standard accounts do, makes a full grasp of the communicative intention impossible without also figuring out the informative intention. As young infants might not yet have the world knowledge or the cognitive wherewithal to infer the content of the latter intention, they would not be able to infer the former either. Consequently, neither could develop. One solution would be to hold that infants represent the informative intention embedded as a placeholder, specifying the content type, within the communicative intention. But if the complexity of the metarepresentational structure in more developed communication does create explanatory problems for a psychologically plausible account, entertaining complex metarepresentations about an empty first-order content should be seen as an even more formidable task for infants. (Consider: Mary intends Peter believes Mary intends Peter believes something.) If, instead, recognizing communication and inferring the content are carried out by two separate processes, the infant can do the former without

the latter. Thus, while the two must be conceptually linked, they may need to be (meta)representationally separate.

Although what adult interlocutors explicitly communicate often, if not always, underdetermines what they intend to convey, it does not follow that the audience must necessarily consider the communicator and her intention in the inferential processes (see also Recanati, 2002). There are cases of ostensive communication in which the specific mental state of the communicator is not relevant to the interpretation (Geurts, 2019b). For instance, when a priest utters "I hereby pronounce you husband and wife," (or in similar conventional speech acts) it does not usually matter what he intended by this. Likewise, when we interpret the meaning of a road sign that we have not seen before, the communicative intentions of the person(s) who installed it may not contribute much (Sterelny, 2017)—assuming that they are accessible at all. But these could be viewed as exceptions to an otherwise interpersonal form of communication in which the identity and the mental states of both the communicator and the audience are inferable and contribute to the interpretation[4]. While this is certainly true, it would be a mistake to conclude that the mechanisms involved in this final product are identical to the ones responsible for their emergence in development and evolutionary history. Again, some of the findings in the natural pedagogy framework could shed light on this.

In many studies on the development of communication, the identity of the communicator is not revealed (e.g., a recorded voice is played back). Children, nevertheless, seem to interpret the communicative acts appropriately, expanding our knowledge about the development of ostensive communication. Such "depersonalized" communication could, of course, be merely an artifact of experimental design. However, Egyed et al. (2013) have conducted a study that could address the issue more directly (see also Novack et al., 2014): Eighteen-month-old infants saw an experimenter show a positive emotional expression directed at one object and a negative emotional expression toward another. These were preceded either by ostensive signals or no communication. In the test phase, either the same or a new experimenter requested the infant to hand them one of the objects. The pattern of results suggested that in ostensive contexts infants interpreted the expression as

[4] Alternatively, one could argue that these are not exceptions but involve attribution of general intentions without an identified subject. My main point is not that these are definitive counterexamples, but rather that these *can* be given a different, more parsimonious interpretation without appealing to intentions.

communicating generalizable object-directed knowledge, encouraging them to hand the *new* experimenter the positively valenced object. Although one can still argue that the experimenters' communicative intentions are considered by the infants, it could alternatively imply that the identity, and thus the specific mental state, of the communicator is less relevant for infants in the interpretive process than the communicative act itself and what it conveys about the referent (see also Topál et al., 2009). Here communication seems to work more like the road sign example. If communicative acts are mostly interpreted this way early in infancy and if these pedagogical scenarios reflect the context in which humans evolved their communicative system (see chapter 4), then it might not be as dependent on the attribution of mental states as the standard account makes it out to be. This picture would create a dilemma for intentionalism, as it urges one to either claim that this is not a case of nonnatural meaning or to drop the mentalistic reduction altogether.

Beside representing other representations, another feature of metarepresentations is suggested to be that they decouple the metarepresented content from the rest of the cognitive system, adding to it information that specifies, among other things, the kind and source of the representation (Cosmides & Tooby, 2000; Leslie, 1987; see §6.6 for more discussion). Decoupling would allow the cognizer to make inferences about the content within its relevant scope without committing errors originating from confusing metarepresented and first-order representations. This is critical for attributing mental states, since otherwise one would take someone else's beliefs and intentions as their own and behave maladaptively (Cosmides & Tooby, 2000). This might also be crucial for adult communication in which the communicative nature of the representation and who produced it (and for whom) could potentially affect not only how one interprets it but also the credal value one attaches to it—as the representation could be either mistaken or deceptive. However, in parent-infant interactions, trust in the benevolence and competence of the communicator is built into the kin-based organization of the interaction. A pedagogical scenario for the origin of ostensive communication would, therefore, allow that one treats the communicated representation as equivalent, if not superior (Marno & Csibra, 2015; Topál et al., 2008), to information one obtains from perception and first-hand experience. Representing the self's belief as the target of communication would also be mostly unnecessary, for, besides its reliability, the communicative act is, due to the production of infant-directed

ostensive signals, naturally meant for the infant. If so, the representations required for the emergence of a concept of communication would be minimized considerably.

Another reason to doubt the necessity of higher-order metarepresentations is absence of convincing evidence for intermediate cases. While it would be difficult to find direct support for the proposal that ostensive communication requires several orders of metapsychological representations, the presence of populations that, because of lacking some of the metarepresentations, show only intermediate capacities would provide indirect support. One potential intermediate case could come from the developmental trajectory of communication. If ostensive communication does indeed rely on four orders of metarepresentations, and assuming that newborns do not yet possess such capacities, there may be a developmental phase at which children are capable of expressing and comprehending, say, informative intentions (i.e., intention to induce belief) but not communicative intentions (i.e., intention to reveal informative intentions). Contra Scott-Phillips (2015), findings from the domain of communication are not useful in this, since they would not enable teasing apart communicative from informative intentions. But cases that could clearly count as "absent authorship" (Grosse et al., 2013), where there is only an informative intention at play, seem rare. And their presence in early infancy (never mind before the emergence of communication) is questionable. A possible example of absent authorship is placing an object at a visible place for someone else to notice, without being concerned whether your intention is recognized. Hidden authorship is a perhaps more common case where, according to the standard account, the informative intention is concealed by the actor. An analogous example is placing an empty glass in a visible location so that your host fills it, but this time intending that she does not notice your intention. Hidden authorship is argued to have the same representational structure as ostensive communication (Scott-Phillips, 2015). However, intuitively at least, it seems much more cumbersome than communication and is unlikely to develop on a par with it. One implication of the existence of a comprehension module (Sperber & Wilson, 2002) might be that the complex metarepresentational structure of communication emerges together in a "package", making it unlikely to reveal a developmental trajectory from simpler structures. This may, however, render falsifying the underlying representations very difficult.

Yet another potential source of evidence is atypical development. If one can identify populations who, due to difficulty with higher-order metarepresentations, are

nevertheless able to entertain informative intentions but not communicative intentions, this could be used to support the standard picture. Hypothetically, these populations would intentionally convey information to change the mental state of their audience but be unable to make this intention overt. To my knowledge, there are no such populations. Moreover, although communicative problems associated with autism are sometimes taken as evidence for intentionalism (e.g., Happé, 1993), the presence of individuals with autism who, despite not passing the false-belief task, can communicate linguistically implies that mechanisms other than the formation and attribution of higher-order intentions may be involved in communication (Glüer & Pagin, 2003).

Finally, there does not seem to be any good examples of informative intention among non-human primates (Heintz & Scott-Phillips, 2022; Zuberbühler, 2018; but see Sperber, 2019). The existence of species which inform one another by inducing mental states without making that overt would exhibit that the Gricean reduction of communicative cognition to metapsychological processes is phylogenetically well-founded. Primate attention-getters may be one place to look for these, although it should be empirically tested whether they actually target belief-states rather than behavior. Of course, none of the above is direct counterevidence against intentionalism, but the absence of intermediate cases raises the question as to why only the end product of these processes exists. As I will argue in the next section, this is possibly because the very reduction is misdirected.

6.4. Metarepresentations (of Mental Representations) are not Sufficient

As mentioned before, Grice's formulation of the intentional structure of communication has been responded with a host of counterexamples targeting either the necessity or sufficiency of the analysis. For instance, it has been argued that torturing has the same structure as meaning$_{NN}$: the torturer intends the audience to produce a response *r*, to recognize this intention, and to produce *r* based on this recognition (Grice, 1969). However, a more widely-discussed type of counterexample was introduced by Strawson (1964): Mary wants Peter to mend her broken hairdryer. Thus, she pretends to mend the pieces together, hoping that Peter notices this and helps out. She intends Peter to realize that she intends him to help, but she does not intend him to realize that she intends him to realize that she intends him to help (example from Sperber & Wilson, 1995). Here, although all Grice's clauses seem to be fulfilled, advocates of the

general analysis do not wish to consider it as a case of meaning$_{NN}$, because, intuitively, this does not involve communication. As a result, the sufficiency of the analysis is threatened. Thus, Strawson proposed that the communicator "should not only intend A to recognize his intention to get A to think that *p*, but that he should also *intend A to recognize his intention to get A to recognize his intention* to get A to think that *p*" (p. 447). However, he noted correctly that even this condition is unlikely to rule out further counterexamples. No matter how many such intentions we add to the formulation, there will be counterexamples in which the actor has a further deceptive intention to hide the lower-order intention. This will result in an infinite regress of intentions (and metarepresentations)—undesirable for a cognitive account.

One type of measure to root out these "sneaky intentions" (Grice, 1969) has been to introduce a condition that bars deception. For instance, Neale (1992) adds to clauses (a) and (b) a further clause that U does not intend A to be deceived about U's intentions (a) and (b) (see also Grice, 1969; Moore, 2016). The issue with this measure is that, while it may be a good solution for a philosophical conceptual analysis, it does not provide a plausible cognitive account of communication. It appears unlikely that every time the audience is addressed in communication, he considers the absence of deceptive intentions. And it is even more unlikely that the less sophisticated individuals (i.e., earlier hominins and infants) developing ostensive communication could entertain such thoughts. Plausibly, taking into account this kind of deception is an exception to the default interpretation of communication as being honest with respect to its communicative nature.

Another measure, advocated by Schiffer (1972), is introducing a mutual-knowledge condition. U and A mutually know that *p*, if and only if U knows that *p*, A knows that *p*, U knows that A knows that *p*, A knows that U knows that *p*, U knows that A knows that U knows that *p*, A knows that U knows that A knows that *p*, and so on ad infinitum. Schiffer believes that it is the absence of mutual knowledge of this form that has led to the deceptive counterexamples. For communication to be properly overt, the intentions involved must be mutually known between the communicator and the audience. However, as a cognitive account, this would merely replace the "vertical" regress of the communicator's mental states with a "horizontal" regress of both interlocutors' mutual mental states. As a result, mutual knowledge, too, fails to provide a psychologically plausible explanation of communication (Sperber & Wilson, 1995). Sperber and Wilson (1995) suggest using, instead, the weaker concept of *mutual*

manifestness. According to them, "[a] proposition is manifest to an individual at a given time to the extent that he is likely to some positive degree to entertain it and accept it as true" (Sperber & Wilson, 2015, p. 134). In a mutual cognitive environment, propositions that identify the people who share that environment are *mutually* manifest as well. A more technical definition of a communicative intention is, then, an intention that makes it mutually manifest to audience and communicator that the communicator has a particular informative intention. However, the problem may lie more with the notion of mutuality than with knowledge per se (see also Garnham & Perner, 1990; Sperber & Wilson, 1990). That is, mutual manifestness may create the same kind of regress as mutual knowledge.

Geurts (2019, 2020) has recently suggested to substitute mutual knowledge with the arguably less demanding normative notions of "reasons to believe" (Lewis, 1969) or "mutual commitment"—both of which take the iterative structure of common ground to be a chain of implications rather than actual cognitive processes. Whichever view on common ground and the related notions we end up accepting, it will still be problematic to account for individual communicative cognition by appeal to such interpersonal concepts. Firstly, mutual mental states are superfluously complex concepts to attribute to prelinguistic creatures who are beginning to develop communication. As argued in the previous section, it is possible that, early in infancy, humans are endowed with an inferential communicative system which does not necessitate attributing complex mental states. Mastery of the interpersonal, mutual consequences of communication (as opposed to, say, hidden or absent authorship) and the norms that follow its practice arguably take place at a later stage in development. Secondly, measures such as an anti-deception condition or a mutual-knowledge condition have been proposed to rule out counterexamples to the conceptual analysis of the philosophical notion of meaning$_{NN}$—aimed ultimately at spelling out its necessary and sufficient conditions. Consequently, they are unlikely to yield good models of human cognition (see also Scarafone & Michael, 2022). Thirdly, and most importantly, the dominant approach of characterizing the interpreter's concept of communication largely as action produced with a particular mental set-up is problematic. As we saw in the torture example above, that an instrumental action is produced by an actor with an underlying mental state akin to communication is not enough for rendering it communicative. One must also spell out how communicative, as opposed to instrumental, action is distinguished and how (as I will try to show in the

next section) it relates to its informational content. Having a convincing account of this type of action, we can solidly ground our understanding of the unique, underlying cognitive mechanisms that generate it, too. However, as mentioned, the standard approach attempts to characterize ostensive communication by specifying the representations that are thought to produce it. These postulated representations are then weighed against intuitions as to what should or should not count as communication proper. I believe that this is the wrong explanatory direction. A sounder approach would be to directly address those intuitions by offering an account of communicative action (including how it is distinguished and how it is taken to convey information). Issues of the psychological plausibility of complex metarepresentations and mutual knowledge arise arguably because the behavioral is explained in terms of the psychological. And when this fails, ever more complex cognitive processes are added to compensate it. However, since interlocutors do not have direct access to each other's mental states, in spite of sophisticated metarepresentational capacities, they cannot achieve overt communication. If, instead, one takes account of the action, notions such as overtness and the issues associated with it might not emerge at all.

In Gricean formulations (much simplified in the previous section), communicative representations consist in a configuration of intentions and beliefs. This implies that by developing the two concepts of intention and belief and linking them humans come to possess a metarepresentational complex that enables communication. More specifically, the reduction involves casting the problem of explaining communication to another level, where, among the set of possible intentions, there is a type of belief-inducing intentions. Granting the validity of this reduction, we can make another prediction regarding the developmental trajectory of the concept of communication: (1) infants develop the ability to attribute intentions and beliefs; (2) they link these two propositional attitudes; (3) they make use of the latter link to express and attribute informative intentions; and (4) they develop a higher-order informative intention that allows them to express and attribute communicative intentions. This would be ostensive communication proper. The problem with this prediction is already visible in (1), as from early in infancy humans show a rich and flexible understanding of communication (Bohn & Frank, 2019; Csibra, 2010; Vouloumanos et al., 2014). Thus, communication emerges alongside, if not earlier than, mindreading capacities and its core features are unlikely to be dependent on such mentalistic concepts. The link in (2) is also problematic in explaining the development of communication. On the one

hand, this would delay the emergence of communication even further in development. On the other hand, prominent theories of conceptual change (Carey, 2009; Spelke, 2003; Xu, 2019) take language to be at the center of this process, either through the structures it generates in the mind or through its compositional semantics that permits combining information across concepts. However, since the hypothetical conceptual change that would link the concepts of intention and belief is itself postulated to explain communication, linguistic communication is clearly unavailable to the process. One must therefore specify how this process occurs (e.g., see Gopnik, 2011). Regarding stages (3) and (4), as it was argued in the previous section, it is question-begging, and as yet empirically unsupported, that the expression and attribution of informative intentions (e.g., in the form of absent authorship) precedes ostensive communication. Indeed, it appears more likely that absent and hidden authorship develop as offshoots of communication, that is, informative behaviors that suppress, or leave out, communicative cues. Therefore, on ontogenetic grounds, belief-inducing intentions fail to account for the relatively early emergence of the concept of communication.

That communication does not emerge from metapsychological representations becomes more evident when we apply the logic of the latter to the interpretation of communication. In attribution of ordinary, instrumental intentions, we predict the behavior of agents to follow from the content of their conative propositional attitudes. For example, when we see Mary walking in the direction of her house, we attribute to her the intention to go home, and we predict her behavior accordingly:

Mary intends [Mary goes home] $\rightarrow$ Mary goes home

This is possible because we have a good understanding of the means she can take to obtain her goal. For instance, we consult our knowledge that she lives nearby to infer that she will walk home. Clearly, the same procedure cannot be used to predict communicative action:

*Mary intends [Peter believes *p*] $\rightarrow$ Peter believes *p*

Thus, Mary's intention that Peter believe something (e.g., that the berries are edible) is not sufficient. (After all, she does not possess telepathic abilities.) This is possibly because the subject of the embedded proposition is another person. So, a more proper analogy might be the following:

Mary intends [Peter sits down] $\rightarrow$ Mary makes Peter sit down

Again, this inference is meaningful because we have a good grasp of the physical possibilities and constraints of human action. We subconsciously consult our knowledge to yield the conclusion that Mary pushes Peter down. But:

Mary intends [Peter believes *p*] → Mary makes Peter believe *p*

Although this time the inference is not wrong, it is trivial because the (communicative) action that leads to Peter's belief is the very thing we would like to predict. Since the range of communicative means is much broader (see also Sperber & Wilson, 2002), we cannot as easily predict them. And more importantly, the infant simply does not have prior access to them. The resources that are available for interpreting instrumental goals are unavailable in communication. These include, among other things, efficiency and simulation. The infant can predict behavior by assuming that agents choose the most efficient means given the environmental constraints (Gergely & Csibra, 2003). But efficiency is less relevant to communicative behavior. For instance, the English word "tree" is no more efficient or rational for denoting the concept TREE than the German word "Baum". Simulation too is unhelpful, because the infant cannot rely on her own limited repertoire to predict others' communicative behavior. Therefore, the above schema for action prediction is unlikely to be sufficiently useful for comprehending, and learning about, communicative behavior.

A similar difficulty emerges for action explanation. In explaining an action, we rely on the effect that the action caused to infer a corresponding intention. Thus:

Mary made Peter sit down → Mary intended [Peter sits down]

Yet this is not possible for explaining most communicative actions:

*Mary made Peter believe *p* → Mary intended [Peter believes *p*]

It is not possible in third-personal communication since, unlike instrumental actions, there is often no observable change in the audience, and we obviously do not have access to their beliefs. Considering the central role of declarative, as opposed to imperative, communication for humans (Tomasello, 2008), the problem becomes even more striking. In second-personal communication, the audience cannot start from the belief they form as a result of being addressed by an act, and then attribute that as the intention of the communicator. Recognizing this intention, on the standard accounts, is itself the very goal of the communicator (Sperber & Wilson, 2002). Faced with novel words or gestures, as infants are, there is no way of first forming a belief in order to interpret it as the intention of the communicator or as the meaning of the utterance. A

relevance-guided comprehension module may avoid some of the problems with intention attribution. However, attributing the inference of such a module as the communicator's intention would imply metacognitive access to the conclusion, complicating the developmental trajectory even more.

As a result, the standard treatment of utterances as just another case of instrumental action and the underlying processes as cases of mentalistic inferences is not a sufficient approach and is unlikely to explain the emergence of human communication in ontogeny and phylogeny. It might be more fruitful in accounting for language *use*, in which there is a more or less established channel in place and interlocutors can use the tools at their disposal to manipulate one another's mental states and actions. It could also prove helpful for explaining how non-human primates use their gestures and vocalizations which are relatively preestablished and limited but used flexibly to achieve various goals (Byrne et al., 2017). Human communication, however, is marked not only by flexible use but also flexible production of ad hoc communicative means (e.g., pantomiming and demonstration) or entirely learned devices (e.g., words and gestures). Then, an account is in order which explains the unique behavioral features of ostensive communication and the underpinning cognitive representations that make this system possible.

6.5. Communication as Representational Action

So far, I have claimed that we cannot unproblematically analyze communicative cognition into a configuration of intentions and beliefs. Ostensive communication may alternatively involve an irreducible, primitive concept that enables comprehending, learning about, and learning from communicative episodes. Thus, our example could be further simplified to only one metarepresentation:

Mary communicates$_1$

these berries are edible.

And sometimes, we saw, even the identity of the communicator may not be relevant. Now the question is: what is communication if not the expression and attribution of intentions? Perhaps the main phenomenon that a recourse to metapsychology makes us neglect is informative behavior. The presence of this class of behavior is, of course, not denied, but rather taken for granted. However, this should arguably be the central question for any account of the development and evolution of human-specific communication. From early in life, we, but apparently no other primate, treat a class

of behavior as having an informational function. This is unlikely to emerge from metarepresentations of mental states, for these are targeted at the relation between other minds and the external world. And besides, they serve to explain and predict actions that typically bring about a perceivable change in the world, not behavior that informs. A primate, armed with the most sophisticated metarepresentational capacities, would still struggle to understand why another agent moves its arms around in a strange, ineffective fashion. (Possibly it would think that the agent is desperately trying to drive away an insect, and so would go on with its business.) To us, however, it appears very trivial that a pantomiming action is informing about what it resembles—so much so that we might come to think that no extra cognitive mechanism is required for it. But no rationalization of otherwise instrumental behavior is likely to lead one to its representational nature if one does not already possess the relevant concept for interpreting informative behavior. Moreover, while, as shown above, it is possible to envisage a non-mentalistic form of inferential communication, there cannot be any realistic account of human communication that does not involve informative behavior—that is, manipulation of external stimuli for the purpose of conveying information. Therefore, an account of this class of behavior and its representation in the mind takes explanatory precedence over mental state attribution.

But what does it mean for an action to have an informational function? Everything in our environment is a potential source of information, including other people's behavior. Thus, ostensive communication is not special in transmitting information. To answer the question, we could go back to Grice. As I mentioned above, while Grice's analysis of meaning and communication might have been mistaken, he seems to have had the right insight about what should count as meaning$_{NN}$. We saw that, read functionally, clause (b) deals with the distinctive quality in human communication of marking$_{NN}$. Clause (c), on the other hand, requires the audience to produce *r* based on the fulfillment of the sub-intention in (b). Grice's insight was that if this is not the case, then what we have is an instance of showing—that is, "deliberately and openly letting someone know" (Grice, 1957). He believed that such cases (e.g., when you show a bandaged leg to convey that you have a bandaged leg) should not be considered meaning$_{NN}$. In these cases, as opposed to cases of "telling", the required inference can be made mostly based on the observable evidence and regardless of the purpose of communication. Of course, in keeping with his commitment to intentionalism, Grice's reasoning for this was that the inference be based on the

recognition of the communicator's intention. However, it is possible to take a non-intentional lesson from this.

In paradigm cases of telling (e.g., in linguistic utterances and gestures), the relation between the communicative medium and the content is clearly one of representationality. This means that in order to arrive at the content of a linguistic utterance we cannot rely solely on the utterance and its physical features, but we need to infer a detached propositional content that the utterance is representing. That this is the case is most striking in depictions (Clark, 2016) because, despite perceptual similarities between the medium and the content, the depiction is meaningful only in relation to what it represents. Otherwise, a drawing of a cup is just a mark on a piece of paper without the affordances of an actual cup. Although perhaps less striking, this is at the heart of most of our communication. The word "cup" (or miming the affordance of a cup) is not an actual cup—nor should it resemble one. How it relates to a particular cup or a CUP concept is through representing. The dependence on detached contents becomes more discernible when we look at the acquisition history of the communicative behavior. That is, although the relation between "cup" and its meaning seems to be direct, hearing the word for the first time we must infer what it is a representation of. This is not the case in the bandaged leg example. There, drawing attention to the bandage does not represent anything—unless used to communicate something else. The bandaged leg is simply a bandaged leg.

At first blush, the theoretical reliance on representationality might appear to rule out a wide range of human communication. Is pointing to a cup not ostensive communication? Sometimes we point at a cup to evoke the kind associated with the proximal referent—for example, when we ask for any member of the same kind as the indicated cup. Here, the distal referent (i.e., the kind) is detached from, and represented by, the communicative medium. But even when we are pointing at a cup to convey something about that very cup, we are seldom merely drawing attention to it. The same holds when we show a bandaged leg. Often, we point at something to lead the audience to inferences that they would not otherwise make. In most such cases, even if the referent is not detached from our communication, that is, it is part of the proximal medium we utilize to inform, the predicate is nevertheless detached from it. By pointing at the cup, we request our audience, say, to FILL it. What is distinctive about human pointing is not only that it can refer to things. Even some species of fish might be able to do so (R. Moore, 2018; Vail et al., 2013). Human pointing is special

for its capacity to invite detached inferences (see also Tomasello, 2008). Hence, reference in pointing should not be considered in isolation from predication. Instead, pointing should be seen as an "utterance" that involves a full propositional content. Or take cases of demonstration. Sometimes, as in most pedagogical demonstrations, both the referent and the predicate fall out of the scope of the communicative medium. For example, we denote an action kind on an object kind, none of which are by definition present here and now. And even when we demonstrate an action on a particular object (e.g., a specific machine), the predicate (e.g., the action kind or the observer's future action) is represented by the communicative act.

Representationality can, nonetheless, rule out the familiar counterexamples. Torture is not a representational action, even if its intentional structure turns out to be akin to communication. It is, alas, an instrumental action taken to bring about a change (albeit peculiar) in another person. Examples of hidden authorship, too, involve (at least for the audience) non-representational action. Regardless of the mental states of the actor, the audience treats planted evidence as he would do, were it not slyly arranged by someone else. Yet when the audience takes some arrangement to be communicative, he interprets it as representing a detached content. A key on the table would then not only be a key, but also represent a content that is to be inferred—say, that the audience should lock the door using that key. As I said, Grice wished also to rule out showing in his account of meaning$_{NN}$ for its apparent natural feature—although he included cases such as nonspontaneous frowning (Grice, 1982). This move has been criticized as unnecessarily excluding important forms of communication (Neale, 1992; Sperber & Wilson, 2015). One way of approaching the question would be to hold that admitting showing in our account achieves inclusivity at the expense of explanatory power. If we allow examples like the bandaged leg, we might miss the crucial feature in most human communication (e.g., in linguistic utterances) of informing about something only indirectly. However, as I mentioned above, we rarely communicate like the limiting case of the bandaged leg. This is also reflected in other approaches to communication which require utterances to convey relevant information (Grice, 1975; Sperber & Wilson, 1995). The relevance expectation often leads the audience to seek information that is not readily available. Besides, the function for which a cognitive system has evolved need not entirely overlap with how it can actually be used (see also Sperber, 1994a; Sperber & Hirschfeld, 2004). While, as I claim, our communicative concept may be geared to representational action, it can be exploited

also to direct attention to stimuli. Then, even linguistic utterances, as paradigm cases of representational communication, could be used to manipulate attention: when we say “I am here!” to a friend who is looking for us in a crowd, regardless of the semantic features of the constituent words, our voice is likely to attract her attention and lead her to the same conclusion (Recanati, 1986). The present account would also incorporate examples like nonspontaneous frowning: how it transmits the information is not necessarily or only the communicative intention behind it, but also its representational relation to its content (i.e., actual frowns and their implication). Lastly, soliloquy (i.e., communication without an audience) can create explanatory problems for intentionalist accounts, as there is no audience in whom you can intend to induce a response or belief. However, you can perform a representational action whether or not it is addressed to anyone.

The point is not that communicative acts are the only domain in which we make inferences about things detached from the perceivable stimuli (see also Gärdenfors, 1995; Planer, 2021). This happens almost all the time—for instance, when we infer the presence of things that we cannot currently see. Nor does it mean simply that ostensive communication evokes representations in our minds. Some alarm calls in birds and chimpanzees appear to evoke the representation corresponding to their (functional) referent (Sato et al., 2022; Suzuki, 2018). What is unique about ostensive communication is that it creates in our mind the *expectation* that the stimulus is representational. It relies, thus, on the existence of a class of stimuli in communication which have an aboutness. Despite the referential quality of some animal signals and enculturated apes’ acquisition of language-like signs (reviewed in Gillespie-Lynch et al., 2014), it is doubtful that other species possess any matching disposition to attribute representationality to unfamiliar stimuli in novel channels (see also Novack & Waxman, 2020; Warren & Call, 2022). Humans, on the other hand, interpret novel communicative behavior in various modalities as representational from very early on (Ferguson & Waxman, 2016; Novack et al., 2014; Tauzin & Gergely, 2018).

The representational understanding of communicative acts provides an extraordinary possibility for humans to link concepts to other concepts or entities—with potential consequences for conceptual development. By drawing on the representational nature of communication, you can spontaneously specify a referent, perform an action that calls to mind your concept of choice, and connect the concept to the referent. In this way, you can communicate that you want your phone to be

brought to you or even suggest (in pretense) that a banana is your phone (Leslie, 1987). Thus, due to their representational nature, our communicative acts permit us to establish arbitrary (or nonnatural) informational links. We can call this informing$_{NN}$. Inferences in other domains arguably draw on the existence of preestablished informational links, based, for example, on statistical or natural relations. Smoke means fire because there is a causal and statistical relation between the two (see also Piccinini & Scarantino, 2011; Scarantino & Piccinini, 2010). An alarm call means snake, due to genetically and/or statistically encoded associations. However, ostensive communication *is* a way of establishing informational relations. As such, its functioning does not require (although it can use) the preexistence of informational links between the communicative action and the referent. We can, thus, account for the difference between what Grice called natural and nonnatural meaning without appealing to the intentions behind them.

When we communicate, we use one entity E_1 (typically an action) to inform about another entity E_2 (i.e., the referent). This way we set up an asymmetrical informational link between E_1 and E_2, such that E_1 can be used to draw inferences about E_2—but not vice versa. Consequently, the task of the audience is to identify in the cognitive environment the scope of the representational medium, on the one hand, and the scope of the representational content, on the other. Specifying these representational scopes is not a trivial matter, of course. It is likely that early in development these are largely prespecified. For instance, as I suggested in the previous chapter, the action may be taken to designate the predicate (most evidently in demonstrations) and the object may be construed as exemplifying its kind. By building on this link between the action and the object kind, and the spatiotemporal features of the action, the infant can both acquire generic knowledge and also bootstrap the development of the communicative system. Later in ontogeny, depending on the context (and, of course, the attributed mental state of the communicator), the scope can vary dramatically. Sometimes the object is part of the communicative medium, sometimes it is not. Sometimes the action informs us about an entity, sometimes (as in the example with the key) we do not observe any action whatsoever.

The communicative system, then, typically functions along these lines in comprehension: (1) communicative episodes are detected through ostensive stimuli; (2) the referent is identified (mostly through following deictic gestures); (3) an informational link is established between the communicative act, a conceptual

predicate, and the referent; (4) inferences are drawn about the referent. Let us see how this works. In action demonstrations, after detecting ostension, the infant identifies the object (or object kind) as the referent, by following both the adult's gaze and her manual manipulation of the object (Pomiechowska & Csibra, 2022). A predicate placeholder is generated, filled by conceptual information drawn from the action (e.g., an action kind or a physical property), and linked to the referent object. Finally, the infant makes inferences about the referent—say, that this kind of object is to be manipulated in the demonstrated fashion. Sometimes, however, the predicate is not iconically represented in the action. In cases of pointing, for example, infants must fill the predicate placeholder based on the context (or even on the referent). Detecting communication, they use the adult's pointing gesture to identify the referent, they generate a predicate placeholder and fill it with conceptual information from the context (e.g., GIVE in a play context), and use this to yield inferences about the referent (e.g., that they must give the object to their mother). And still sometimes the predicate might be codified in the action (as in linguistic communication). Note that such inferences might not have been made outside the domain of communication. Observing someone perform a purely instrumental action on an object, infants may or may not draw those conceptual inferences. (They may, for instance, encode it as a transient relation between that specific action and the object.) Or seeing their mother search for an object, they might eventually realize that she wishes the object to be handed to her. However, by exploiting the above-mentioned simple procedure, communication can constrain and secure the necessary inferences. While leaving the key on the table could remind your flatmates to lock the door, if you do it conspicuously to tap into their communicative concept, you secure your intended inference. They will now take it to represent a detached content by, say, ascribing a predicate (e.g., LOCK) to an implied referent (e.g., the key or the door). This, of course, is an atypical example, where there is only a trace of the communicative action. As a perhaps more typical example, the caregiver demonstrates an action to secure in the infant's mind an informational link between the action and the indexed referent, leading the infant to otherwise opaque inferences about the object kind (Csibra & Gergely, 2011). And in linguistic communication, one can even exploit the representational nature of communicative action to create, and inform about, fictional entities. Furthermore, whereas information use is abundant across cognitive domains, "referential information" (Scarantino &

Clay, 2015), in which one entity is directly stipulated to inform about another entity, seems unique to communication.

Thus, the use of representational means to convey detached propositional contents or simply informing$_{NN}$ enables an inferential communicative system that does not necessitate attributing mental states. This is certainly not to deny the importance of mentalizing in communication. Often, we rely on the beliefs or intentions of the communicator to identify the referent or predicate. And sometimes the content of the utterance is itself a mental state. However, the attribution of mental states would be unhelpful in communication if the representational structure were not in place. With it, one can comprehend communication both when the mental state is irrelevant or unavailable and when it is necessary to arrive at the right interpretation. The schema of intention attribution failed to predict and explain communication (§6.4) because communicative acts involve a different type of goal. The goal in instrumental actions is typically a two-place relation between an action *a* and a change of state *b*: *I(a, b)*. However, in communication the goal is a three-place relation between an action *x*, what it represents *y*, and (sometimes) the change of state (e.g., belief or action) *z*: *C(x, y, z)*. Thus, with respect to the goal, communication involves components that are distinct from instrumental, goal-directed action.

To sum up, human communication is characterized by marking$_{NN}$ and informing$_{NN}$. Marking$_{NN}$ allows us to open-endedly mark actions as communicative through ostensive stimuli. Informing$_{NN}$ is about how the message in the marked$_{NN}$ action is communicated. The informational relation between the communicative action and its message is one of representationality, which involves a detached propositional content (i.e., one with a predicate-argument structure). By establishing concept-mediated informational links between the communicative action and a referent, humans can open-endedly convey information to one another. Although it can utilize postulated mental states to home in on the content, this process generally involves distinct components. We thus have an account of communication that, although certainly sketchy at this point, takes account both of the action (i.e., communicatively marked, representational action) and its cognitive underpinning.

6.6. The Evolution of Metarepresentation

Evolutionary theories of the origin of metarepresentational capacities can be classified into three groups: first, theories that propose metarepresentation evolved to solve

mostly individual, rather than social, problems such as metacognition (Couchman et al., 2009) and decoupling (Cosmides & Tooby, 2000); second, theories that suggest metarepresentation originated in social cognition (Baron-Cohen, 1999; Byrne & Whiten, 1997; Sperber, 2000); and third, theories that claim metarepresentation evolved culturally in language (Geurts, 2021; R. Moore, 2021).

Prominent among the first strand of theories is Cosmides and Tooby's (2000) suggestion that metarepresentation is an adaptation to the "cognitive niche". This is an adaptive mode that involves increasing use of contingent and local information for the regulation of improvised behavior (see also Pinker, 2010). Through decoupling (Leslie, 1987), metarepresented propositional content is separated from the rest of the cognitive system to allow inferences that are valid within the relevant scope but harmful if applied outside of it. These representations are stored with source tags which indicate how they have been obtained (e.g., self vs other). Subsequent information about the source (e.g., its reliability) can affect the truth-status of the representation and move it upwards or downwards in the cognitive architecture in terms of credibility. Such metarepresentational capacities are useful not only for solving socio-cognitive problems (e.g., belief attribution and communication) but also, among other things, for planning, episodic memory, counterfactuals, and suppositions. If this metarepresentational system is compromised, it can result in maladaptive cognitive and behavioral conditions such as those symptomatic of schizophrenia.

A more popular approach to the evolution of metarepresentation views it as an adaptation to a "socio-cognitive niche" (Whiten & Erdal, 2012). According to this view, the evolution of distinctive cognitive abilities in primates (also called "Machiavellian intelligence") is largely determined by living in large, semi-permanent groups of long-lived individuals and the problems it poses (Byrne, 1996; Byrne & Whiten, 1997). This environment favors, on the one hand, the use of deceptive, social manipulation to achieve individual benefits at the expense of other group members and, on the other hand, cooperation and coalition building. This causes an arms race between the social skills of those seeking higher ranks in the group and those collaborating to counter the alpha's dominance—a positive feedback loop that leads to the evolution of ever more complex socio-cognitive specializations. Mentalistic metarepresentation is among these adaptations, enabling individuals to interpret and predict the behavior of conspecifics not just as bodily movement but as action guided by beliefs and desires (Sperber, 2000). Such metapsychological reasoning helps individuals to protect

themselves from others, to exploit them, and to cooperate more successfully with them. Hence, metarepresentation could have evolved independently of communication in response to social selection pressures (Scott-Phillips, 2014; Sperber, 2000). As a result, communication could emerge, in ontogeny and phylogeny alike, on the back of psychological metarepresentation (Baron-Cohen, 1999). Evidence for preverbal infants' sensitivity to false beliefs in non-linguistic tasks provides some support for this (Kovács et al., 2010; Onishi & Baillargeon, 2005; Surian et al., 2007). Perhaps the main challenge for this approach is to explain how and why humans transitioned from a "perception-goal psychology" (Call & Tomasello, 2008), characteristic of non-human primates, to a belief-desire psychology, with its unique recursive structure, or, in other words, how higher-order representations emerged from primary representations.

Advocates of the third approach attempt to address the latter challenge. According to them, natural languages provide humans with representational tools that also enable expressing and entertaining propositional attitudes. These tools include recursion (e.g., embedding a sentence within a sentence), representing falsities, and evidential marking (De Villiers, 2013; R. Moore, 2021). Besides, language reveals the logical structure of propositional attitudes which allows contrasting and combining them with the content of other propositional attitudes to yield further inferences (Bermúdez, 2017; R. Moore, 2021). Since we represent others' mental states in order to use them in our own conscious practical decision-making, it is argued, they must be consciously accessible representations of an external language rather than sentences in a language of thought (Bermúdez, 2017). If true, this dependence on natural languages could imply that full-blown metarepresentations are the outcome of relatively recent cultural, rather than biological, evolution. Geurts (2021) notes that in many languages, quotative verbs are also used for attributing mental states, including beliefs and intentions. Thus, one can imagine an evolutionary trajectory from quotation to the public practice of attributing mental states (corresponding to the quoted expression) and eventually to implicit mental state attribution. Similarly, Moore (2021) suggests that human-specific forms of metapsychology are linguistically constructed folk models of the human mind which have been invented and modified by humans for various purposes. According to him, beside quotation, a major source for propositional attitude concepts has been perception verbs such as "see" and "hear" which have culturally evolved to attain cognitive senses. Proponents of this approach should

explain how an inferential communicative system that went beyond the simple coded signaling of non-human animals could get off the ground absent a sophisticated metapsychology (Sperber, 2000). More crucially, linguistic expressions are themselves external, representational devices that necessitate metarepresentation. Therefore, their existence cannot be taken for granted in accounting for the evolution of metarepresentation.

In the absence of direct paleoanthropological evidence as to the emergence of metarepresentation, any hypothesis about its evolutionary root in humans will have to be largely speculative—unless, of course, we obtain more evidence in support of metarepresentational mindreading in non-human primates (see also Krupenye et al., 2016). However, the plausible alternatives would still be worth considering. One of the main reasons for the popularity of the idea that psychological metarepresentation precedes communication (e.g., Baron-Cohen, 1999) is the assumption that the latter requires the former. However, if my account of communication is correct and communicative cognition emerges independently of mentalizing, we can also envisage the opposite direction: metarepresentation evolved to enable external, communicative representations and only later was it exapted for postulating mental states to interpret instrumental, goal-directed behavior.

This proposal is explanatorily more powerful than the language-first proposal, as it does not take linguistic representations for granted. Rather, it proposes a representational system that, on the one hand, permitted both linguistic and non-linguistic external representations, and, on the other, provided a platform for metapsychological representations. Similarly to the language-first proposal, however, it suggests an ecology in which there are perceivable objects in the world that can promote incremental evolution from organisms lacking metarepresentation to ones with increasingly sophisticated metarepresentational capacities, which they can exploit for various communicative, metacognitive, and metapsychological functions.

Consider two types of representing the representational medium: in the first (admittedly very shallow) type, the representational medium is represented and used in learning, but only as a non-representational entity; in the second (full-fledged) type, the representational medium is represented as a representation proper, that is, with a representational content. The first type is clearly simpler and can potentially support the evolution of the more sophisticated type. However, this is only possible with public representations, for mental representations are not available to perception; and even

if they were somehow inferable, their representation as non-representational objects would be futile. As a result, taking metapsychology as the original function demands an evolutionary leap from organisms capable of only primary representations to ones with the ability to postulate abstract, higher-order constructs (i.e., mental states) directed at similarly abstract contents (e.g., a false belief-content or a future state of affairs).

The hypothesis that ostensive communication evolved mainly for teaching generalizable knowledge to the offspring, plausible in its own right, offers the intriguing possibility of a scenario in which pedagogical communicators can *use* external stimuli representationally to transmit information about kinds, in the absence of specialized cognitive mechanisms. The teachers could, say, perform adaptively-relevant actions (e.g., knapping flints) in the presence of their children. This already provides a suitable environment for learning. They could, additionally, monitor their children's gaze or emit sounds to ensure that their attention is oriented towards the action. Because these cues (i.e., eye contact and vocalizations) are associated with adaptive information, children would benefit from evolving a preference for them—eventually promoting them to child-directed, ostensive signals. Moreover, as these proto-demonstrations often entail generalizable knowledge, it would be advantageous for the learner to develop a cognitive shortcut from the ostensively marked action-object pair to the respective kinds, that is, interpret the demonstration as *representing* a generic predicate on an object kind. This would, thus, be an ecology in which actions have an "aboutness"—a feature which can lead to the emergence of specialized cognitive mechanisms capable of utilizing it efficiently. (The attraction of this scenario notwithstanding, one can imagine further alternative scenarios where actors manipulate public stimuli to convey information, and observers subsequently evolve a conceptual framework in which to make better sense of the stimuli.)

In addition, metarepresentation has some features that are instrumental for mentalizing but not for representing external representations. Decoupling, for instance, is an integral feature of mental state attribution, for otherwise someone else's belief or desire would be detrimentally taken as one's own. In a kin-selected communicative system (Fitch, 2004, 2007), however, you would be safe to encode and store the transmitted knowledge without needing to quarantine it with a source tag. Likewise, if the parent is both competent and benevolent in communication, you would not need to worry about whether the representation misrepresents the content—

suggested to be necessary for a full-blown understanding of representations (Perner, 1991).

Thus, at least in theory, metarepresenting representations of a public nature can have a relatively simpler cognitive structure. This, in turn, enables an incremental evolutionary trajectory from a non-representational understanding of representational entities to comprehending recursive representations (i.e., representations of representations). Once recursive representation was in place for communication between kin, it could further be used to transmit generic and, eventually, episodic information to non-kin. This extension in use would, however, necessitate some form of decoupling. Decoupled, recursive representations would not only enable communication but could be exploited also to attribute the mental states corresponding to the utterances. This representational format could ultimately be applied to domains in which the medium is abstract and can only be contextually inferred. In this way, instrumental action interpretation would be significantly enhanced, for now the causes for behavior can be expanded beyond what is immediately observable. This application could be through an intermediary factor such as language, as suggested by the language-first approach, but it could occur more directly through repurposing the same representational structure. Moreover, communication itself creates strong selection pressure for the evolution of increasingly sophisticated mindreading, augmenting the resources for successful communication.

One implication of considering external, communicative representations as the original domain of metarepresentation could be that what set apart human communication from the limited communication of other great apes was not cognitive constraints linked to the capacity for complex recursive mindreading, but rather an environment that favored ever more flexible communication. Since other animals face only a limited range of prespecified signals in their lifetime, adequate for the phylogenetically recurring problems which drove their evolution, they do not need an encompassing “naïve signaling theory”, that is, a metarepresentational concept of communication. However, an environment that involves ever-changing and cognitively opaque knowledge and technology, like the one our hominin ancestors inhabited (Boyd & Silk, 2014; Csibra & Gergely, 2011; Roche et al., 2009; Toth & Schick, 2015), fosters an open-ended communicative system for the faithful transfer of generic information. Such an environment creates a “symbolic niche” in which it is beneficial to interpret (communicative) action as representationally conveying information that is

applicable beyond the locally perceived behavior to displaced conceptual entities. As emphasized, this scenario is inevitably speculative. But if plausible, it can potentially change our perspective on the evolution of uniquely human forms of communication and perhaps social cognition.

6.7. Final Remark

Human communication is commonly understood in terms of the intentional structure that is at play in its production and comprehension alike. This account has several shortcomings in explaining communicative cognition which were discussed above. Particularly, complex metarepresentations of intentions and beliefs are neither necessary nor sufficient in accounting for the design features of ostensive communication as a behavior and as a cognitive concept. Chiefly, the standard account may lead to overlooking informative (i.e., representational) action—common across our diverse uses of communication. As an evolutionary account, intentionalism has arguably hindered progress in comparative research. Firstly, emphasizing the intentions behind ostensive communication obscures what is truly unique about it and leads one to seek its origin in the intentionality of primate communication (Zuberbühler, 2018)—rather than, for instance, behavior that is potentially homologous with representational communication. Secondly, as the purported mental states are inaccessible, devising paradigms in which to test similar traits in non-human animals will prove difficult, if not impossible. However, if my account is plausible, one may conduct studies which can empirically test whether other primates are capable of flexibly marking their actions as communicative, and, more pertinent to the present chapter, whether they interpret unfamiliar stimuli as representing a detached content. Such studies will likely move research on the evolutionary origin of ostensive communication forward and shed light on what genuinely separates (or unifies) our interactions and those of our primate cousins.

Chapter 7. General Conclusion

In this thesis, I have tried to approach the question as to why we communicate ostensively from various angles. In chapter 2, I suggested that an explicitly functional account of ostensive communication may facilitate the identification of the communicative system and the adaptive problem that it solves. I surveyed some of the notions evoked in the literature to characterize ostension including higher-order intentionality, attention manipulation, and audience specification. These notions are used in mostly mechanistic terms and so do not clarify the function that ostension serves or whether this function can be implemented in different mechanisms. Crucially, these notions do not account for the ability in humans to generate novel communicative means. I proposed, instead, that the notion of marking$_{NN}$ can explain this ability, while also providing a functional account that is applicable across species and age ranges—relatively independently of the proximate mechanisms. I defined marking$_{NN}$ as flexibly marking actions (and entities targeted by action) as communicative. This function can be realized both in specialized ostensive signals (e.g., eye contact) or inference-directed stimuli (e.g., doing something in a stylized manner suggestive of communication). Moreover, this function corresponds to Grice's higher-order intention, whereas the function of informing$_{NN}$, discussed in chapter 6, corresponds to the first-order intention to induce a belief or response.

In chapter 3, I investigated the diverse selective scenarios proposed for explaining the evolution of human communication. I concluded that scenarios focused on the basic pragmatic competence for ostensive communication are more compatible with the criteria suggested for assessing evolutionary accounts of human communication—that is, uniqueness, immediate utility, generality, honesty, and cooperativeness. Particularly, these scenarios can explain how ostensive communication could be adaptive from the outset without any culturally developed code: a non-verbal system using "natural" means of communication can emerge and provide a platform for language. According to a widely held idea, this system evolved in a collaborative and cooperative context to permit various joint activities.

In chapter 4, I defended the idea that ostensive communication evolved, instead, for a specific kind of cooperation, namely teaching. Much of what we know about the life of early hominins is due to ancient stone tools. While this could be simply because

of a bias for preserving non-perishable artifacts, it provides strong evidence for hominin behavior. As suggested by theoretical and empirical studies, complex tools may have been difficult to make without acquiring knowledge and skills from others. Pedagogical communication is an important means of knowledge acquisition. Thus, tools involving opaque skills may have necessitated open-ended communication. Open-ended communication is arguably useful in all adaptive domains, but teaching open-ended technology is possible only through such a system as opposed to typical animal signals. Therefore, teaching meets the uniqueness criterion. Demonstrations constituting a non-verbal teaching channel can additionally use actions and objects to communicate generic knowledge—thus accounting for immediate utility. These demonstrations enable the generality of human communication because they can communicate open-endedly, they are about displaced entities (i.e., kinds of objects and their predicates), and they use actions and objects representationally. Teaching in these contexts is likely to be honest and cooperative due to inclusive fitness. Once such a system was in place, it would also be available to other selection pressures—leading eventually to linguistic communication.

Chapter 5 asked whether action demonstration already shares features of language. The predicate-argument structure which is a distinctive property of language is the primary focus. This structure may be seen in the complementary contribution of actions and objects to demonstrations. While objects can be understood as functioning like symbols standing for their kind as the arguments, actions performed on them reveal their properties and relations like linguistic predicates. This creates an open-ended compositional system and may even contain syntactic qualities.

Chapter 6 dealt with the broader question of the cognitive concept of communication and the role of metarepresentation. According to the dominant account, namely intentionalism, this concept comprises representations of mental states. I presented extensive arguments against this purported priority of mentalistic metarepresentations. I suggest that many of these metarepresentational levels are not necessary, if we consider the notion of marking$_{NN}$ and the role of specialized ostensive signals: these can be decoded rather than metarepresentationally inferred. More importantly, metarepresentation of mental states is insufficient for explaining the concept of communication. This is because the existing developmental evidence does not reveal a progression from mental state attribution to communication. In addition, intention attribution is better suited to instrumental action interpretation in that this type

of action often causes a perceivable effect that can be taken as the goal of the agent following from her intention. Communication often does not produce such effects. Instead, I propose that our concept of communication involves representational action. This kind of action entails a detached content that needs to be inferred. Representing contents in action allows establishing informational links between the represented content and the identified referent. I then discussed the evolutionary implication of the conceptual priority of such representational actions: if communication can emerge independently of mental state attribution, then one can also envisage that metarepresentation evolved originally to represent external, communicative representations and was then coopted both for language and metapsychology.

Overall, my account of ostensive communication diverges from dominant accounts in several respects. As mentioned, it can be contrasted most clearly with intentionalism. Although intentions certainly contribute to the expressive power of ostensive communication, I have tried to show that they may not be central to it. Both the recognition of communicative acts and the interpretation of its informational content can be dealt with minimally by the complementary roles of ostensive markers and representational actions. The cognitive underpinnings for producing and understanding such communication can then utilize mindreading to reason about the intentions behind acts, say, to make sure that actions are produced for a communicative purpose or to zero in on the intended content. Thus, although my account attempts to attribute minimal requirements to communicators, it is not incompatible with more sophisticated mentalistic communication. Future theoretical and empirical research can shed light on the actual role of mental state attribution in interactions. I hope that the specification of problems in my thesis will guide this research and enable a balanced account of the contribution of various mechanisms to our unique communicative system.

References

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(2), 105–124. https://doi.org/10.1017/S0140525X05000038

Armstrong, J. (2021). Communication before communicative intentions. *Noûs*.

Bach, K. (1987). On Communicative Intentions: A Reply to Recanati. *Mind & Language*, *2*(2), 141–154. https://doi.org/10.1111/j.1468-0017.1987.tb00112.x

Bach, K. (2012). Meaning and Communication. In G. Russell & D. G. Fara (Eds.), *Routledge Companion to the Philosophy of Language* (pp. 79–90). London: Routledge.

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Bar-On, D. (2013). Origins of Meaning: Must We 'Go Gricean'? *Mind & Language*, *28*(3), 342–375. https://doi.org/10.1111/mila.12021

Bar-On, D. (2017). Communicative intentions, expressive communication, and origins of meaning. In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 301-312). Routledge.

Baron-Cohen, S. (1999). The evolution of a theory of mind. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 261–277). Oxford University Press.

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology & Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492–499. https://doi.org/10.1111/j.1467-7687.2005.00440.x

Bermúdez, J. L. (2017). Can nonlinguistic animals think about thinking? In K. Andrews & J. Beck (Eds.), *The Routledge handbook of philosophy of animal minds* (pp. 119–130). Routledge.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press.

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, *117*(3), 510–526. https://doi.org/10.1016/j.lingua.2005.02.006

Bickerton, D. (2009). *Adam's tongue: How humans made language, how language made humans*. Macmillan.

Bickerton, D. (2011). The origins of syntactic language. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 456–468). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0049

Bickerton, D., & Szathmáry, E. (2011). Confrontational scavenging as a possible source for language and cooperation. *BMC Evolutionary Biology*, *11*(1), 261. https://doi.org/10.1186/1471-2148-11-261

Birch, J. (2021). Toolmaking and the evolution of normative cognition. *Biology & Philosophy*, *36*(1), 4. https://doi.org/10.1007/s10539-020-09777-9

Biro, D., Haslam, M., & Rutz, C. (2013). Tool use as adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1630), 20120408. https://doi.org/10.1098/rstb.2012.0408

Bloom, P. (2000). *How children learn the meanings of words*. The MIT Press.

Boeckx, C. A., & Fujita, K. (2014). Syntax, action, comparative cognitive science, and Darwinian thinking. *Frontiers in Psychology*, *5*, 627.

Boesch, C. (1991). Teaching among wild chimpanzees. *Animal Behaviour*, *41*, 530–532. https://doi.org/10.1016/S0003-3472(05)80857-7

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*, 223–249.

Bohn, M., Liebal, K., Oña, L., & Tessler, M. H. (2022). Great ape communication as contextual social inference: A computational modelling perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1859), 20210096. https://doi.org/10.1098/rstb.2021.0096

Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How Could Language Have Evolved? *PLOS Biology*, *12*(8), e1001934. https://doi.org/10.1371/journal.pbio.1001934

Booth, A. E., Schuler, K., & Zajicek, R. (2010). Specifying the role of function in infant categorization. *Infant Behavior and Development*, *33*(4), 672–684. https://doi.org/10.1016/j.infbeh.2010.09.003

Boyd, B. (2018). The evolution of stories: From mimesis to language, from fact to fiction. *WIREs Cognitive Science*, *9*(1), e1444. https://doi.org/10.1002/wcs.1444

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918-10925.

Boyd, R., & Silk, J. B. (2014). *How Humans Evolved: Seventh Edition*. W. W. Norton & Company.

Boyette, A. H., & Hewlett, B. S. (2018). Teaching in Hunter-Gatherers. *Review of Philosophy and Psychology*, *9*(4), 771–797. https://doi.org/10.1007/s13164-017-0347-2

Breheny, R. (2006). Communication and Folk Psychology. *Mind & Language*, *21*(1), 74–107. https://doi.org/10.1111/j.1468-0017.2006.00307.x

Bretherton, I., & Bates, E. (1979). The emergence of intentional communication. *New Directions for Child and Adolescent Development*, *1979*(4), 81–100. https://doi.org/10.1002/cd.23219790407

Brinck, I., & Gärdenfors, P. (2003). Co–operation and Communication in Apes and Humans. *Mind & Language*, *18*(5), 484–501. https://doi.org/10.1111/1468-0017.00239

Burdett, E. R. R., Dean, L. G., & Ronfard, S. (2018). A Diverse and Flexible Teaching Toolkit Facilitates the Human Capacity for Cumulative Culture. *Review of Philosophy and Psychology*, *9*(4), 807–818. https://doi.org/10.1007/s13164-017-0345-4

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(5), 175–186. https://doi.org/10.1002/evan.20222

Butler, L. P., & Markman, E. M. (2012). Preschoolers Use Intentional and Pedagogical Cues to Guide Inductive Inferences and Exploration. *Child Development*, *83*(4), 1416–1428. https://doi.org/10.1111/j.1467-8624.2012.01775.x

Butler, L. P., & Markman, E. M. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, *130*(1), 116–127. https://doi.org/10.1016/j.cognition.2013.10.002

Byrne, R. W. (1996). Machiavellian intelligence. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *5*(5), 172-180.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: Intentional communication with a rich set of innate signals. *Animal Cognition*, *20*(4), 755–769. https://doi.org/10.1007/s10071-017-1096-4

Byrne, R. W., & Rapaport, L. G. (2011). What are we learning from teaching? *Animal Behaviour*, *5*(82), 1207–1211.

Byrne, R. W., & Whiten, A. (1997). Machiavellian intelligence. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian Intelligence II: Extensions and Evaluations* (pp. 1–23). Cambridge University Press. https://doi.org/10.1017/CBO9780511525636.002

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carlson, G. N. (1977). *Reference to kinds in English.* University of Massachusetts Amherst.

Caro, T. M., & Hauser, M. D. (1992). Is There Teaching in Nonhuman Animals? *The Quarterly Review of Biology*, *67*(2), 151–174. https://doi.org/10.1086/417553

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Castro, L., & Toro, M. A. (2014). Cumulative cultural evolution: The role of teaching. *Journal of Theoretical Biology*, *347*, 74–83. https://doi.org/10.1016/j.jtbi.2014.01.006

Chomsky, N. (2006). *Language and mind*. Cambridge University Press.

Christiansen, M. H., Kirby, S. (2003). Language Evolution: The Hardest Problem in Science? In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 1–15). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199244843.003.0001

Cimpian, A. (2016). The Privileged Status of Category Representations in Early Development. *Child Development Perspectives*, *10*(2), 99–104. https://doi.org/10.1111/cdep.12166

Clark, B. (2011). Scavenging, the stag hunt, and the evolution of language. *Journal of Linguistics*, *47*(2), 447–480.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. https://doi.org/10.1037/rev0000026

Comrie, B. (2008). Argument structure. In *Syntax: An International Handbook of Contemporary Research.* Berlin. New York: De Gruyter Mouton.

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 53–115). Oxford University Press.

Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2009). Metacognition is prior. *Behavioral and Brain Sciences*, *32*(2), 142–142. https://doi.org/10.1017/S0140525X09000594

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology*, *22*(2), 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Csibra, G. (2007a). Action mirroring and action understanding: An alternative account. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorymotor Foundations of Higher Cognition. Attention and Performance XXII*, (pp. 435–459). Oxford University Press, USA.

Csibra, G. (2007b). Teachers in the wild. *Trends in Cognitive Sciences*, *11*(3), 95–96. https://doi.org/10.1016/j.tics.2006.12.001

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

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.)*, Processes of Change in Brain and Cognitive Development. Attention and Performance XXI*, (pp. 249–274). OUP Oxford.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157. https://doi.org/10.1098/rstb.2010.0319

Csibra, G., & Shamsudheen, R. (2015). Nonverbal Generics: Human Infants Interpret Objects as Symbols of Object Kinds. *Annual Review of Psychology*, *66*, 689–710. https://doi.org/10.1146/annurev-psych-010814-015232

Davidson, D. (2009). *Truth and Predication*. Harvard University Press.

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology*. John Wiley & Sons.

de Boer, B. (2011). Infant-directed speech and language evolution. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 322–327). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0033

De Villiers, J. (2013). Language and reasoning about beliefs. In M. Banaji & S. Gelman (Eds.), *Navigating the Social World* (pp. 96–100). Oxford University Press.

Deacon, T. W. (1998). *The Symbolic Species: The Co-evolution of Language and the Brain*. W. W. Norton & Company.

Dean, L. G., Vale, G. L., Laland, K. N., Flynn, E., & Kendal, R. L. (2014). Human cumulative culture: A comparative perspective. *Biological Reviews*, *89*(2), 284–301. https://doi.org/10.1111/brv.12053

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution*, *1*(5), Article 5. https://doi.org/10.1038/s41559-017-0112

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*.

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology*, *47*, 1499–1503. https://doi.org/10.1037/a0025659

DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences*, *8*(2), 66–70. https://doi.org/10.1016/j.tics.2003.12.004

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences*, *6*(3), 343–355.

Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.

Dessalles, J.-L. (2007). *Why We Talk: The Evolutionary Origins of Language*. OUP Oxford.

Dessalles, J.-L. (2008). From metonymy to syntax in the communication of events. *Interaction Studies*, *9*(1), 51–65. https://doi.org/10.1075/is.9.1.05des

Dessalles, J.-L. (2014). Why talk? In D. Dor, C. Knight, & J. Lewis (Eds.), *The Social Origins of Language* (pp. 284–296). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665327.003.0020

Dewar, K., & Xu, F. (2009). Do Early Nouns Refer to Kinds or Distinct Shapes?: Evidence From 10-Month-Old Infants. *Psychological Science*, *20*(2), 252–257. https://doi.org/10.1111/j.1467-9280.2009.02278.x

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.

Dunbar, R. I. M. (1997). Groups, Gossip, and the Evolution of Language. In A. Schmitt, K. Atzwanger, K. Grammer, & K. Schäfer (Eds.), *New Aspects of Human Ethology* (pp. 77–89). Springer US. https://doi.org/10.1007/978-0-585-34289-4_5

Dunbar, R. I. M. (1998a). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.

Dunbar, R. I. M. (1998b). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190.

Dunbar, R. I. M. (2017). Group size, vocal grooming and the origins of language. *Psychonomic Bulletin & Review*, *24*(1), 209–212. https://doi.org/10.3758/s13423-016-1122-6

Egyed, K., Király, I., & Gergely, G. (2013). Communicating Shared Knowledge in Infancy. *Psychological Science*, *24*(7), 1348–1353. https://doi.org/10.1177/0956797612471952

Engh, A. L., Hoffmeier, R. R., Cheney, D. L., & Seyfarth, R. M. (2006). Who, me? Can baboons infer the target of vocalizations? *Animal Behaviour*, *71*(2), 381–387. https://doi.org/10.1016/j.anbehav.2005.05.009

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(4), 491–503. https://doi.org/10.1017/S0140525X04000111

Ferguson, B., & Waxman, S. R. (2016). What the [beep]? Six-month-olds link novel communicative signals to meaning. *Cognition*, *146*, 185–189. https://doi.org/10.1016/j.cognition.2015.09.020

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-Month-Old Infants: An Advantage of Words Over Tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Fitch, W. T. (2004). Kin Selection and "Mother Tongues": A Neglected Component in Language Evolution. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach*, (pp. 275–296). MIT Press.

Fitch, W. T. (2007). Evolving Meaning: The Roles of Kin Selection, Allomothering and Paternal Care in Language Evolution. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence of Communication and Language* (pp. 29–51). Springer. https://doi.org/10.1007/978-1-84628-779-4_2

Fitch, W. T. (2015). Evolving pragmatics. *Current Biology*, *25*(23), R1110–R1112. https://doi.org/10.1016/j.cub.2015.10.013

Fitch, W. T. (2017a). On externalization and cognitive continuity in language evolution. *Mind & Language*, *32*(5), 597–606. https://doi.org/10.1111/mila.12162

Fitch, W. T. (2017b). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 3–33. https://doi.org/10.3758/s13423-017-1236-5

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

Fogarty, L., Strimling, P., & Laland, K. N. (2011). The Evolution of Teaching. *Evolution*, *65*(10), 2760–2770. https://doi.org/10.1111/j.1558-5646.2011.01370.x

Franco, F., Perucchini, P., & March, B. (2009). Is Infant Initiation of Joint Attention by Pointing Affected by Type of Interaction? *Social Development*, *18*(1), 51–76. https://doi.org/10.1111/j.1467-9507.2008.00464.x

Franks, N. R., & Richardson, T. (2006). Teaching in tandem-running ants. *Nature*, *439*(7073), 153-153.

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative Function Demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, *117*(1), 1–8. https://doi.org/10.1016/j.cognition.2010.06.003

Gallistel, C. R. (2011). Prelinguistic Thought. *Language Learning and Development*, *7*(4), 253–262. https://doi.org/10.1080/15475441.2011.578548

Gärdenfors, P. (1995). Cued and detached representations in animal cognition. *Behavioural Processes*, *35*(1), 263–273. https://doi.org/10.1016/0376-6357(95)00043-7

Gärdenfors, P., & Högberg, A. (2017). The Archaeology of Teaching and the Evolution of Homo docens. *Current Anthropology*, *58*(2), 188–208. https://doi.org/10.1086/691178

Garnham, A., & Perner, J. (1990). Does manifestness solve problems of mutuality? *Behavioral and Brain Sciences*, *13*(1), 178–179. https://doi.org/10.1017/S0140525X00078225

Gelman, S. A. (2004). Learning Words for Kinds: Generic Noun Phrases in Acquisition. In *Weaving a lexicon* (pp. 445–484). MIT Press.

Gentner, D. (1982). Why are nouns learned before verbs: Linguistic relativity versus natural partitioning. *Language Development: Language, Thought and Culture*, *2*, 301–334.

Genty, E., & Zuberbühler, K. (2014). Spatial Reference in a Bonobo Gesture. *Current Biology*, *24*(14), 1601–1605. https://doi.org/10.1016/j.cub.2014.05.065

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, *10*(1), 139–146. https://doi.org/10.1111/j.1467-7687.2007.00576.x

Geurts, B. (2019a). What's wrong with Gricean pragmatics. *ExLing 2019: Proceedings of the 10th International Conference of Experimental Linguistics*, 1–9.

Geurts, B. (2019b). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*(1–2), 1–30. https://doi.org/10.1515/tl-2019-0001

Geurts, B. (2020). *Common ground as a normative condition*. [Unpublished manuscript].

Geurts, B. (2021). First saying, then believing: The pragmatic roots of folk psychology. *Mind & Language*, *36*(4), 515–532. https://doi.org/10.1111/mila.12345

Geurts, B. (2022). Evolutionary pragmatics: From chimp-style communication to human discourse. *Journal of Pragmatics*, *200*, 24–34. https://doi.org/10.1016/j.pragma.2022.07.004

Gil, D. (2012). Where does predication come from? *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *57*(2), 303–333. https://doi.org/10.1017/S0008413100004795

Gillespie-Lynch, K., Greenfield, P. M., Lyn, H., & Savage-Rumbaugh, S. (2014). Gestural and symbolic development among apes and humans: Support for a multimodal theory of language evolution. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01228

Glüer, K., & Pagin, P. (2003). Meaning Theory and Autistic Speakers. *Mind & Language*, *18*(1), 23–51. https://doi.org/10.1111/1468-0017.00213

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*(6664), Article 6664. https://doi.org/10.1038/34646

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163–9168.

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning While Babbling: Prelinguistic Object-Directed Vocalizations Indicate a Readiness to Learn. *Infancy*, *15*(4), 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x

Gómez, J. C. (1994). Mutual awareness in primate communication: A Gricean approach. In S. T. Parker, R. M. Mitchell, & M. L. Boccia (Eds.), *Self-awareness in animals and humans: Developmental perspectives* (pp. 61–80). Cambridge University Press. https://doi.org/10.1017/CBO9780511565526.007

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

Gómez, J. C. (2010). The emergence of eye contact as an intersubjective signal in an infant gorilla: Implications for models of early social cognition. *Acción Psicológica*, *7*(2), 35–43. https://doi.org/10.5944/ap.7.2.213

Gopnik, A. (2011). A unified account of abstract structure and conceptual change: Probabilistic models and early learning mechanisms. *Behavioral and Brain Sciences*, *34*(3), 129.

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze Following Is Not Dependent on Ostensive Cues: A Critical Test of Natural Pedagogy. *Child Development*, *89*(6), 2091–2098. https://doi.org/10.1111/cdev.13026

Green, M. (2019). Organic Meaning: An Approach to Communication with Minimal Appeal to Minds. In A. Capone, M. Carapezza, & F. Lo Piparo (Eds.), *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications* (pp. 211–228). Springer International Publishing. https://doi.org/10.1007/978-3-030-00973-1_12

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Grice, H. P. (1969). Utterer's Meaning and Intentions. *The Philosophical Review*, *78*(2), 147–177.

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

Grice, P. (1982). Meaning revisited. *Mutual Knowledge*, 222–243.

Griffin, A. S., & West, S. A. (2002). Kin selection: Fact and fiction. *Trends in Ecology & Evolution*, *17*(1), 15–21. https://doi.org/10.1016/S0169-5347(01)02355-2

Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology*, *46*(6), 1710–1722. https://doi.org/10.1037/a0020727

Grosse, G., Scott-Phillips, T. C., & Tomasello, M. (2013). Three-year-olds hide their communicative intentions in appropriate contexts. *Developmental Psychology*, *49*(11), 2095–2101. https://doi.org/10.1037/a0032017

Gustison, M. L., Tinsley Johnson, E., Beehner, J. C., & Bergman, T. J. (2019). The social functions of complex vocal sequences in wild geladas. *Behavioral Ecology and Sociobiology*, *73*(1), 1–12.

Hamilton, W. (1964). The genetical theory of social behaviour, I and II. *Journal of Theoretical Biology*, 7, 1-52.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*(3), 571–581. https://doi.org/10.1016/j.anbehav.2003.11.011

Harnad, S. (2011). From sensorimotor categories and pantomime to grounded symbols and propositions. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 387–392). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0042

Harris, D. W. (2019). Intention recognition as the mechanism of human communication. In A. Sullivan (Ed.), *Sensations, Thoughts, Language* (pp. 11-37). Routledge.

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

Havy, M., & Waxman, S. R. (2016). Naming influences 9-month-olds' identification of discrete categories along a perceptual continuum. *Cognition*, *156*, 41–51. https://doi.org/10.1016/j.cognition.2016.07.011

Heintz, C., & Scott-Phillips, T. (2022). Expression unleashed: The evolutionary & cognitive foundations of human communication: VERSION December 2021. *Behavioral and Brain Sciences*, 1–46. https://doi.org/10.1017/S0140525X22000012

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *12*(3), 123–135. https://doi.org/10.1002/evan.10110

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *22*(4), e12779. https://doi.org/10.1111/desc.12779

Hewlett, B. S., & Roulette, C. J. (2016). Teaching in hunter–gatherer infancy. *Royal Society Open Science*, *3*(1), 150403.

Heyes, C. (2016). Born Pupils? Natural Pedagogy and Cultural Pedagogy. *Perspectives on Psychological Science*, *11*(2), 280–295. https://doi.org/10.1177/1745691615621276

Hirai, M., & Kanakogi, Y. (2019). Communicative hand-waving gestures facilitate object learning in preverbal infants. *Developmental Science*, *22*(4), e12787. https://doi.org/10.1111/desc.12787

Hiscock, P. (2014). Learning in Lithic Landscapes: A Reconsideration of the Hominid "Toolmaking" Niche. *Biological Theory*, *9*(1), 27–41. https://doi.org/10.1007/s13752-013-0158-3

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

Hogan, J. A. (1994). The concept of cause in the study of behavior. In J. A. Hogan & J. J. Bolhuis (Eds.) *Causal mechanisms of behavioural development* (pp. 3–15). Cambridge University Press. https://doi.org/10.1017/CBO9780511565120.003

Hoppitt, W. J. E., Brown, G. R., Kendal, R., Rendell, L., Thornton, A., Webster, M. M., & Laland, K. N. (2008). Lessons from animal teaching. *Trends in Ecology & Evolution*, *23*(9), 486–493. https://doi.org/10.1016/j.tree.2008.05.008

Hurford, J. R. (1998). The evolution of language and languages. In R. Dunbar, C. Knight, & C. Power (Eds.), *The Evolution of Culture*. Edinburgh University Press.

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, *26*(3), 261–283. https://doi.org/10.1017/S0140525X03000074

Hurford, J. R. (2006). Proto-propositions. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.), *The evolution of language. The Proceedings of the 6th International Conference (Evolang6)* (pp. 131–138). World Scientific.

Hurford, J. R. (2007). *The Origins of Meaning: Language in the Light of Evolution*. OUP Oxford.

Hurford, J. R. (2012). *The Origins of Grammar: Language in the Light of Evolution II*. OUP Oxford.

Ittelson, W. H. (1996). Visual perception of markings. *Psychonomic Bulletin & Review*, *3*(2), 171–187. https://doi.org/10.3758/BF03212416

Jackendoff, R., & Wittenberg, E. (2017). Linear grammar as a possible stepping-stone in the evolution of language. *Psychonomic Bulletin & Review*, *24*(1), 219–224.

Kachel, G., Moore, R., & Tomasello, M. (2018). Two-year-olds use adults' but not peers' points. *Developmental Science*, *21*(5), e12660. https://doi.org/10.1111/desc.12660

Kaminski, J. (2009). Dogs (Canis familiaris) are Adapted to Receive Human Communication. In A. Berthoz & Y. Christen (Eds.), *Neurobiology of "Umwelt": How Living Beings Perceive the World* (pp. 103–107). Springer. https://doi.org/10.1007/978-3-540-85897-3_9

Kaminski, J., & Piotti, P. (2016). Current Trends in Dog-Human Communication: Do Dogs Inform? *Current Directions in Psychological Science*, *25*(5), 322–326. https://doi.org/10.1177/0963721416661318

Kaminski, J., Schulz, L., & Tomasello, M. (2012). How dogs know when communication is intended for them. *Developmental Science*, *15*(2), 222–232. https://doi.org/10.1111/j.1467-7687.2011.01120.x

Kano, F., Moore, R., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2018). Human ostensive signals do not enhance gaze following in chimpanzees, but do enhance object-oriented attention. *Animal Cognition*, *21*(5), 715–728. https://doi.org/10.1007/s10071-018-1205-z

Keenan, E. L., & Ochs, E. (1979). Becoming a competent speaker of Malagasy. *Languages and Their Speakers*, 113–158.

Keenan, E. O. (1976). The universality of conversational postulates. *Language in Society*, *5*(1), 67–80.

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology*, *116*(2), 471–486. https://doi.org/10.1016/j.jecp.2012.12.003

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, *38*, e31. https://doi.org/10.1017/S0140525X14000090

Knight, C. (1998). Ritual/speech coevolution: A solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight, (Eds.), *Approaches to the Evolution of Language*, (pp. 68–91). Cambridge University Press.

Kolodny, O., & Edelman, S. (2018). The evolution of the capacity for language: The ecological context and adaptive value of a process of cognitive hijacking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170052. https://doi.org/10.1098/rstb.2017.0052

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy*, *19*(6), 543–557. https://doi.org/10.1111/infa.12060

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kovács, Á. M., Téglás, E., Gergely, G., & Csibra, G. (2017). Seeing behind the surface: Communicative demonstration boosts category disambiguation in 12-month-olds. *Developmental Science*, *20*(6), e12485. https://doi.org/10.1111/desc.12485

Krifka, M. (2008). Functional similarities between bimanual coordination and topic/comment structure. In R. Eckardt, G. Jäger, & T. Veenstra (Eds.), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, (Vol. 197, pp. 307–336). Walter de Gruyter.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Lakusta, L., & Carey, S. (2015). Twelve-Month-Old Infants' Encoding of Goal and Source Paths in Agentive and Non-Agentive Motion Events. *Language Learning and Development*, *11*(2), 152–175. https://doi.org/10.1080/15475441.2014.896168

Lakusta, L., & DiFabrizio, S. (2017). And the Winner Is…A Visual Preference for Endpoints over Starting Points in Infants' Motion Event Representations. *Infancy*, *22*(3), 323–343. https://doi.org/10.1111/infa.12153

Laland, K. (2017a). The origins of language in teaching. *Psychonomic Bulletin & Review*, *24*(1), 225–231. https://doi.org/10.3758/s13423-016-1077-7

Laland, K. N. (2017b). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton University Press.

Langus, A., & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*(4), 291–318. https://doi.org/10.1016/j.cogpsych.2010.01.004

Leadbeater, E., Raine, N. E., & Chittka, L. (2006). Social Learning: Ants and the Meaning of Teaching. *Current Biology*, *16*(9), R323–R325. https://doi.org/10.1016/j.cub.2006.03.078

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

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review*, *94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind.' *Trends in Cognitive Sciences*, *8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001

Leslie, S.-J. (2008). Generics: Cognition and Acquisition. *The Philosophical Review*, *117*(1), 1–47. https://doi.org/10.1215/00318108-2007-023

Leslie, S.-J. (2015). Generics oversimplified. *Noûs*, *49*(1), 28–54.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell.

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *American Journal of Primatology*, *64*(4), 377–396. https://doi.org/10.1002/ajp.20087

Liebesman, D. (2011). Simple generics. *Noûs*, *45*(3), 409–442.

Liebesman, D. (2015). Predication as Ascription. *Mind*, *124*(494), 517–569. https://doi.org/10.1093/mind/fzu182

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259–280. https://doi.org/10.1017/S0140525X0600906X

Lombao, D., Guardiola, M., & Mosquera, M. (2017). Teaching to make stone tools: New experimental evidence supporting a technological hypothesis for the origins of language. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-14322-y

Lombard, M. (2015). Hunting and Hunting Technologies as Proxy for Teaching and Learning During the Stone Age of Southern Africa. *Cambridge Archaeological Journal*, *25*(4), 877–887. https://doi.org/10.1017/S0959774315000219

Lonsdorf, E. V. (2006). What is the role of mothers in the acquisition of termite-fishing behaviors in wild chimpanzees (*Pan troglodytes schweinfurthii*)? *Animal Cognition*, *9*(1), 36–46. https://doi.org/10.1007/s10071-005-0002-7

Lucas, A. J., Kings, M., Whittle, D., Davey, E., Happé, F., Caldwell, C. A., & Thornton, A. (2020). The value of teaching increases with tool complexity in cumulative cultural evolution. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1939), 20201885. https://doi.org/10.1098/rspb.2020.1885

Luuk, E. (2009). The noun/verb and predicate/argument structures. *Lingua*, *119*(11), 1707–1727.

Luuk, E. (2012). The origins of linguistic predicate/argument structure. In T. C. Scott-Phillips, M. Tamariz, & E. A. Cartmill (Eds.), *The Evolution of Language* (pp. 204–211). World Scientific. https://doi.org/10.1142/9789814401500_0027

Maffongelli, L., D'Ausilio, A., Fadiga, L., & Daum, M. M. (2019). The Ontogenesis of Action Syntax. *Collabra: Psychology*, *5*(1), 21. https://doi.org/10.1525/collabra.215

Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences*, *41*, e1. https://doi.org/10.1017/S0140525X17000012

Malle, B. F. (2002). The relation between language and theory of mind in development and evolution. In B. F. Malle, & T. Givón (Eds.), *The Evolution of Language out of Pre-Language*, *18*, (pp. 265–284). John Benjamins Publishing.

Marno, H., & Csibra, G. (2015). Toddlers Favor Communicatively Presented Information over Statistical Reliability in Learning about Artifacts. *PLOS ONE*, *10*(3), e0122129. https://doi.org/10.1371/journal.pone.0122129

Matthews, D. (2020). Learning how to communicate in infancy. In C. F. Rowland, A. L. Theakston, B. Ambridge, & K. E. Twomey (Eds.), *Current Perspectives on Child Language Acquisition: How Children Use Their Environment to Learn* (pp. 11–38). John Benjamins Publishing Company.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, *201*(4924), 1145–1147.

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. OUP Oxford.

Mayr, E. (1961). Cause and Effect in Biology. *Science*, *134*(3489), 1501–1506.

Miller, G. (2000). *The mating mind: How sexual choice shaped the evolution of human nature*. Anchor.

Mishkin, M., Lewis, M. E., & Ungerleider, L. G. (1982). Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. *Behavioural Brain Research*, *6*(1), 41–55. https://doi.org/10.1016/0166-4328(82)90080-8

Mitchell, R. W. (1991). Bateson's concept of "metacommunication" in play. *New Ideas in Psychology*, *9*(1), 73–87. https://doi.org/10.1016/0732-118X(91)90042-K

Moore, M. (2010). "Grammars of action" and stone flaking design space. In A. Nowell & I. Davidson (Eds.), *Stone tools and the evolution of human cognition* (pp. 13–43). University Press of Colorado.

Moore, R. (2014). Ontogenetic Constraints on Grice's Theory of Communication. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 87–104). John Benjamins Publishing Company.

Moore, R. (2016a). Meaning and ostension in great ape gestural communication. *Animal Cognition*, *19*(1), 223–231. https://doi.org/10.1007/s10071-015-0905-x

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

Moore, R. (2017). Social cognition, Stag Hunts, and the evolution of language. *Biology & Philosophy*, *32*(6), 797–818. https://doi.org/10.1007/s10539-017-9598-7

Moore, R. (2018). Gricean communication, language development, and animal minds. *Philosophy Compass*, *13*(12), e12550. https://doi.org/10.1111/phc3.12550

Moore, R. (2021). The cultural evolution of mind-modelling. *Synthese*, *199*(1), 1751–1776. https://doi.org/10.1007/s11229-020-02853-3

Moore, R., Mueller, B., Kaminski, J., & Tomasello, M. (2015). Two-year-old children but not domestic dogs understand communicative intentions without language, gestures, or gaze. *Developmental Science*, *18*(2), 232–242. https://doi.org/10.1111/desc.12206

Morgan, T. J., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., ... & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature communications*, *6*(1), 6029. https://doi.org/10.1038/ncomms7029

Moro, A. (2014a). On the similarity between syntax and actions. *Trends in Cognitive Sciences*, *18*(3), 109–110. https://doi.org/10.1016/j.tics.2013.11.006

Moro, A. (2014b). Response to Pulvermüller: The syntax of actions and other metaphors. *Trends in Cognitive Sciences*, *18*(5), 221. https://doi.org/10.1016/j.tics.2014.01.012

Morrison, D. M. (2020). Disambiguated Indexical Pointing as a Tipping Point for the Explosive Emergence of Language Among Human Ancestors. *Biological Theory*, *15*(4), 196–211. https://doi.org/10.1007/s13752-020-00355-6

Motes-Rodrigo, A., McPherron, S. P., Archer, W., Hernandez-Aguilar, R. A., & Tennie, C. (2022). Experimental investigation of orangutans' lithic percussive and sharp stone tool behaviours. *PLOS ONE*, *17*(2), e0263343. https://doi.org/10.1371/journal.pone.0263343

Musgrave, S., Morgan, D., Lonsdorf, E., Mundry, R., & Sanz, C. (2016). Tool transfers are a form of teaching among chimpanzees. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep34783

Mussavifard, N., & Csibra, G. (in press). The co-evolution of cooperation and communication: alternative accounts. *Behavioral and Brain Sciences.*

Napoli, D. J. (1989). *Predication Theory: A Case Study for Indexing Theory*. Cambridge University Press.

Neale, S. (1992). Paul Grice and the Philosophy of Language. *Linguistics and Philosophy*, *15*(5), 509–559.

Nehaniv, C. L. (2005). Open Problems in the Emergence and Evolution of Linguistic Communication: A Road-Map for Research. In *Symposium on the Emergence and Evolution of Linguistic Communication (EELC'05).*

Nesse, R. (2013). Tinbergen's four questions, organized: A response to Bateson and Laland. *Trends in Ecology & Evolution*, *28*. https://doi.org/10.1016/j.tree.2013.10.008

Novack, M. A., Brentari, D., Goldin-Meadow, S., & Waxman, S. (2021). Sign language, like spoken language, promotes object categorization in young hearing infants. *Cognition*, *215*, 104845. https://doi.org/10.1016/j.cognition.2021.104845

Novack, M. A., Henderson, A. M. E., & Woodward, A. L. (2014). Twelve-Month-Old Infants Generalize Novel Signed Labels, but Not Preferences Across Individuals. *Journal of Cognition and Development*, *15*(4), 539–550. https://doi.org/10.1080/15248372.2013.782460

Novack, M. A., & Waxman, S. (2020). Becoming human: Human infants link language and cognition, but what about the other great apes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180408. https://doi.org/10.1098/rstb.2018.0408

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. *Cognition*, *195*, 104082. https://doi.org/10.1016/j.cognition.2019.104082

Okumura, Y., Kobayashi, T., & Itakura, S. (2016). Eye Contact Affects Object Representation in 9-Month-Old Infants. *PLOS ONE*, *11*(10), e0165145. https://doi.org/10.1371/journal.pone.0165145

O'Madagain, C., Kachel, G., & Strickland, B. (2019). The origin of pointing: Evidence for the touch hypothesis. *Science Advances*, *5*(7).

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

Origgi, G., & Sperber, D. (2000). Evolution, Communication and the Proper Function of Language. In P. Carruthers & A. Chamberlain (Eds.), *Evolution and the Human Mind: Language, Modularity and Social Cognition* (pp. 140–169). Cambridge: Cambridge University Press.

Otte, D. (1974). Effects and Functions in the Evolution of Signaling Systems. *Annual Review of Ecology and Systematics*, *5*, 385–417.

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–B113. https://doi.org/10.1016/j.cognition.2005.01.001

Pain, R. (2022). Stone tools, predictive processing and the evolution of language. *Mind & Language*.

Palagi, E., Burghardt, G. M., Smuts, B., Cordoni, G., Dall'Olio, S., Fouts, H. N., Řeháková-Petrů, M., Siviy, S. M., & Pellis, S. M. (2016). Rough-and-tumble play as a window on animal communication. *Biological Reviews*, *91*(2), 311–327. https://doi.org/10.1111/brv.12172

Palmer, C. T. (1991). Kin-selection, reciprocal altruism, and information sharing among Maine lobstermen. *Ethology and Sociobiology*, *12*(3), 221–235. https://doi.org/10.1016/0162-3095(91)90005-B

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 103–117.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, *37*(1), 1–38. https://doi.org/10.1007/s10867-010-9195-3

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(supplement_2), 8993–8999. https://doi.org/10.1073/pnas.0914630107

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–727. https://doi.org/10.1017/S0140525X00081061

Planer, R. J. (2017a). Protolanguage Might Have Evolved Before Ostensive Communication. *Biological Theory*, *12*(2), 72–84. https://doi.org/10.1007/s13752-017-0262-x

Planer, R. J. (2017b). Talking About Tools: Did Early Pleistocene Hominins Have a Protolanguage? *Biological Theory*, *12*(4), 211–221. https://doi.org/10.1007/s13752-017-0279-1

Planer, R. J. (2021). What is Symbolic Cognition? *Topoi*, *40*(1), 233–244. https://doi.org/10.1007/s11245-019-09670-5

Pomiechowska, B., & Csibra, G. (2022). Nonverbal Action Interpretation Guides Novel Word Disambiguation in 12-Month-Olds. *Open Mind*, *6*, 51–76. https://doi.org/10.1162/opmi_a_00055

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–189. https://doi.org/10.2307/1166159

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

Proust, J. (2016). The Evolution of Primate Communication and Metacommunication. *Mind & Language*, *31*(2), 177–203. https://doi.org/10.1111/mila.12100

Pulvermüller, F. (2014). The syntax of action. *Trends in Cognitive Sciences*, *18*(5), 219–220.

Putt, S. S., Woods, A. D., & Franciscus, R. G. (2014). The Role of Verbal Interaction During Experimental Bifacial Stone Tool Manufacture. *Lithic Technology*, *39*(2), 96–112. https://doi.org/10.1179/0197726114Z.00000000036

Pylyshyn, Z. W. (1978). When is attribution of beliefs justified? [P&W]. *Behavioral and Brain Sciences*, *1*(4), 592–593. https://doi.org/10.1017/S0140525X00076895

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2022). The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences. *Behavioral and Brain Sciences*, 1–55. https://doi.org/10.1017/S0140525X22002849

Rakoczy, H., & Behne, T. (2019). Commitment sharing as crucial step toward a developmentally plausible speech act theory? *Theoretical Linguistics*, *45*(1–2), 93–97. https://doi.org/10.1515/tl-2019-0007

Reboul, A. (2001). Foundations of reference and predication. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals. An international handbook* (Vol. 1, pp. 509-522). Walter de Gruyter.

Reboul, A. (2015). Why language really is not a communication system: A cognitive view of language evolution. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01434

Recanati, F. (1986). On Defining Communicative Intentions. *Mind & Language*, *1*(3), 213–241. https://doi.org/10.1111/j.1468-0017.1986.tb00102.x

Recanati, F. (2002). Does Linguistic Communication Rest on Inference? *Mind & Language*, *17*(1–2), 105–126. https://doi.org/10.1111/1468-0017.00191

Revencu, B., & Csibra, G. (2021). For 19-Month-Olds, What Happens On-Screen Stays On-Screen. *Open Mind*, *5*, 71–90. https://doi.org/10.1162/opmi_a_00043

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

Rocci, A., & Luciani, M. (2016). Semantics and verbal communication. In A. Rocci & L. de Saussure (Eds.), *Verbal Communication* (pp. 57–76). Walter de Gruyter GmbH & Co KG.

Roche, H., Blumenschine, R.J., Shea, J.J. (2009). Origins and Adaptations of Early *Homo*: What Archeology Tells Us. In F. E. Grine, J. G. Fleagle, R. E. Leakey (Eds.), *The First Humans – Origin and Early Evolution of the Genus Homo. Vertebrate Paleobiology and Paleoanthropology* (pp. 135–147). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9980-9_12

Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science*, *15*(3), 426–435. https://doi.org/10.1111/j.1467-7687.2012.01139.x

Roy, A. C., Curie, A., Nazir, T., Paulignan, Y., Des Portes, V., Fourneret, P., & Deprez, V. (2013). Syntax at hand: Common syntactic structures for actions and language. *PloS One*, *8*(8), e72677.

Royka, A., Chen, A., Aboody, R., Huanca, T., & Jara-Ettinger, J. (2022). People infer communicative action through an expectation for efficient communication. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-31716-3

Sato, Y., Kano, F., Morimura, N., Tomonaga, M., & Hirata, S. (2022). Chimpanzees (*Pan troglodytes*) exhibit gaze bias for snakes upon hearing alarm calls. *Journal of Comparative Psychology*, *136*(1), 44–53. https://doi.org/10.1037/com0000305

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Scalise Sugiyama, M. (2021). Co-occurrence of Ostensive Communication and Generalizable Knowledge in Forager Storytelling. *Human Nature*, *32*(1), 279–300. https://doi.org/10.1007/s12110-021-09385-w

Scarafone, A., & Michael, J. (2022). Getting Ready to Share Commitments. *philosophical topics*, *50*(1), 135-160.

Scarantino, A. (2013). Animal communication as information-mediated influence. In U. Stegmann (Ed.), *Animal Communication Theory: Information and Influence*, (pp. 63–88). Cambridge University Press.

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, *100*(100), e1–e8. https://doi.org/10.1016/j.anbehav.2014.08.017

Scarantino, A., & Piccinini, G. (2010). Information Without Truth. *Metaphilosophy*, *41*(3), 313–330. https://doi.org/10.1111/j.1467-9973.2010.01632.x

Schick, J., Fryns, C., Wegdell, F., Laporte, M., Zuberbühler, K., Schaik, C. P. van, Townsend, S. W., & Stoll, S. (2022). The function and evolution of child-directed communication. *PLOS Biology*, *20*(5), e3001630. https://doi.org/10.1371/journal.pbio.3001630

Schiffer, S. R. (1972). *Meaning*. New York, NY, USA: Oxford, Clarendon Press.

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

Scott-Phillips, T. C. (2011). Evolutionarily Stable Communication and Pragmatics. In A. Benz, C. Ebert, G. Jäger, & R. van Rooij (Eds.), *Language, Games, and Evolution: Trends in Current Research on Language and Game Theory* (pp. 117–133). Springer. https://doi.org/10.1007/978-3-642-18006-4_6

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

Scott-Phillips, T. C. (2015a). Nonhuman Primate Communication, Pragmatics, and the Origins of Language. *Current Anthropology*, *56*(1), 56–80. https://doi.org/10.1086/679674

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

Scott-Phillips, T. C. (2016). Meaning in great ape communication: Summarising the debate. *Animal Cognition*, *19*(1), 233–238. https://doi.org/10.1007/s10071-015-0936-3

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary Theory and the Ultimate–Proximate Distinction in the Human Behavioral Sciences. *Perspectives on Psychological Science*, *6*(1), 38–47. https://doi.org/10.1177/1745691610393528

Scott-Phillips, T. C., Gurney, J., Ivens, A., Diggle, S. P., & Popat, R. (2014). Combinatorial Communication in Bacteria: Implications for the Origins of Linguistic Generativity. *PLOS ONE*, *9*(4), e95929. https://doi.org/10.1371/journal.pone.0095929

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*(2), 226–233. https://doi.org/10.1016/j.cognition.2009.08.009

Scott-Phillips, T., & Heintz, C. (2023). Animal Communication in Linguistic and Cognitive Perspective. *Annual Review of Linguistics*, *9*(1), null. https://doi.org/10.1146/annurev-linguistics-030421-061233

Searcy, W. A., & Nowicki, S. (2010). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton University Press. https://doi.org/10.1515/9781400835720

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Seed, A., & Byrne, R. (2010). Animal Tool-Use. *Current Biology*, *23*(20), R1032–R1039.

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Seyfarth, R., & Cheney, D. (2018). Pragmatic flexibility in primate vocal production. *Current Opinion in Behavioral Sciences*, *21*, 56–61. https://doi.org/10.1016/j.cobeha.2018.02.005

Seyfarth, R. M., & Cheney, D. L. (2011). Primate social cognition as a precursor to language. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 59–70). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0004

Seyfarth, R. M., & Cheney, D. L. (2018). *The social origins of language.* Princeton University Press.

Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, *80*(1), 3–8. https://doi.org/10.1016/j.anbehav.2010.04.012

Skerry, A. E., Lambert, E., Powell, L. J., & McAuliffe, K. (2013). The Origins of Pedagogy: Developmental and Evolutionary Perspectives. *Evolutionary Psychology*, *11*(3), 147470491301100320. https://doi.org/10.1177/147470491301100306

Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant Pointing: Communication to Cooperate or Communication to Learn? *Child Development*, *78*(3), 735–740. https://doi.org/10.1111/j.1467-8624.2007.01028.x

Spelke, E. S. (2003). What Makes Us Smart? Core Knowledge and Natural Language. In D. Getner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 277–311). MIT Press.

Sperber, D. (1994a). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 39–67). Cambridge University Press.

Sperber, D. (1994b). Understanding Verbal Understanding. In J. Khalfa (Ed.), *What is Intelligence?* (pp. 179-198). Cambridge University Press.

Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117-137). Oxford University Press.

Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, *29*(1/2), 401–413.

Sperber, D. (2018). Rethinking ostension: (1) A terminological issue. *International Cognition and Culture Institute*. https://cognitionandculture.net/blogs/dan-sperber/rethinking-ostension-1/

Sperber, D. (2019). Personal Notes on a Shared Trajectory. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, Pragmatics and Interpretation* (pp. 13–20). Cambridge University Press. https://doi.org/10.1017/9781108290593.002

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, *8*(1), 40–46. https://doi.org/10.1016/j.tics.2003.11.002

Sperber, D., & Wilson, D. (1990). Spontaneous deduction and mutual knowledge. *Behavioral and Brain Sciences*, *13*(1), 179–184.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell Publishing.

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

Sperber, D., & Wilson, D. (2015). Beyond Speaker's Meaning. *Croatian Journal of Philosophy*, *15*(44), 117–149.

Stalmaszczyk, P. (2014). The legacy of Frege and the linguistic theory of predication. In P. Stalmaszczyk (Ed.), *Philosophy of Language and Linguistics* (pp. 225-254). De Gruyter.

Stegmann, U. E. (2013). Introduction: A primer on information and influence in animal communication. In U. E. Stegmann (Ed.), *Animal Communication Theory* (pp. 1–40). Cambridge University Press. https://doi.org/10.1017/CBO9781139003551.002

Sterelny, K. (2017). From code to speaker meaning. *Biology & Philosophy*, *32*(6), 819–838. https://doi.org/10.1007/s10539-017-9597-8

Stout, D., & Chaminade, T. (2009). Making Tools and Making Sense: Complex, Intentional Behaviour in Human Evolution. *Cambridge Archaeological Journal*, *19*(1), 85–96. https://doi.org/10.1017/S0959774309000055

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 75–87. https://doi.org/10.1098/rstb.2011.0099

Strauss, S., Calero, C. I., & Sigman, M. (2014). Teaching, naturally. *Trends in Neuroscience and Education*, *3*(2), 38–43. https://doi.org/10.1016/j.tine.2014.05.001

Strauss, S., & Ziv, M. (2012). Teaching Is a Natural Cognitive Ability for Humans. *Mind, Brain, and Education*, *6*(4), 186–196. https://doi.org/10.1111/j.1751-228X.2012.01156.x

Strawson, P. F. (1964). Intention and Convention in Speech Acts. *The Philosophical Review*, *73*(4), 439–460.

Suddendorf, T. (1999). The rise of the metamind. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 218–260). Oxford University Press.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science*, *18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

Suzuki, T. (2018). Alarm calls evoke a visual search image of a predator in birds. *Proceedings of the National Academy of Sciences*, *115*, 201718884. https://doi.org/10.1073/pnas.1718884115

Szabó, Z. G. (2008). Compositionality. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Számadó, S. (2010). Pre-Hunt Communication Provides Context for the Evolution of Early Human Language. *Biological Theory*, *5*(4), 366–382. https://doi.org/10.1162/BIOT_a_00064

Számadó, S., & Szathmáry, E. (2006). Selective scenarios for the emergence of natural language. *Trends in Ecology & Evolution*, *21*(10), 555–561. https://doi.org/10.1016/j.tree.2006.06.021

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention?. *Scientific Reports*, *4*(1), 1-4.

Tallerman, M. (2011a). Protolanguage. In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 479–491). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0051

Tallerman, M. (2011b). What is syntax? In K. R. Gibson & M. Tallerman (Eds.), *The Oxford Handbook of Language Evolution* (pp. 442–455). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541119.013.0048

Tallerman, M. (2013). Kin selection, pedagogy, and linguistic complexity: Whence protolanguage? In R. Botha & M. Everaert (Eds.), *The Evolutionary Emergence of Language: Evidence and Inference* (pp. 77–96). Oxford University Press.

Tallerman, M., & Gibson, K. R. (2011). *The Oxford handbook of language evolution*. Oxford University Press.

Tauzin, T., & Gergely, G. (2018). Communicative mind-reading in preverbal infants. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-27804-4

Téglás, E., Gergely, A., Kupán, K., Miklósi, Á., & Topál, J. (2012). Dogs' Gaze Following Is Tuned to Human Communicative Signals. *Current Biology*, *22*(3), 209–212. https://doi.org/10.1016/j.cub.2011.12.018

Tehrani, J. J., & Riede, F. (2008). Towards an archaeology of pedagogy: Learning, teaching and the generation of material culture traditions. *World Archaeology*, *40*(3), 316–331. https://doi.org/10.1080/00438240802261267

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415. https://doi.org/10.1098/rstb.2009.0052

Tennie, C., Premo, L. S., Braun, D. R., & McPherron, S. P. (2017). Early Stone Tools and Cultural Transmission: Resetting the Null Hypothesis. *Current Anthropology*, *58*(5), 652–672. https://doi.org/10.1086/693846

Thiele, M., Hepach, R., Michel, C., & Haun, D. B. M. (2021). Observing others' joint attention increases 9-month-old infants' object encoding. *Developmental Psychology*, *57*, 837–850. https://doi.org/10.1037/dev0001189

Thompson, J. R. (2014). Meaning and mindreading. *Mind & Language*, *29*(2), 167–200. https://doi.org/10.1111/mila.12046

Thorgrimsson, G. B., Fawcett, C., & Liszkowski, U. (2014). Infants' expectations about gestures and actions in third-party interactions. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00321

Thornton, A., & McAuliffe, K. (2006). Teaching in Wild Meerkats. *Science*, *313*(5784), 227–229. https://doi.org/10.1126/science.1128727

Thornton, A., & Raihani, N. J. (2008). The evolution of teaching. *Animal Behaviour*, *75*(6), 1823–1836. https://doi.org/10.1016/j.anbehav.2007.12.014

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*. MIT Press.

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Harvard University Press. https://doi.org/10.4159/9780674988651

Tomasello, M., & Call, J. (2019). Thirty years of great ape gestures. *Animal Cognition*, *22*(4), 461–469. https://doi.org/10.1007/s10071-018-1167-1

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, *53*(6), 673–692. https://doi.org/10.1086/668207

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, *17*(4), 563–584. https://doi.org/10.1348/026151099165483

Topál, J., Gergely, G., Erdőhegyi, Á., Csibra, G., & Miklósi, Á. (2009). Differential Sensitivity to Human Communication in Dogs, Wolves, and Human Infants. *Science*, *325*(5945), 1269–1272. https://doi.org/10.1126/science.1176960

Topál, J., Gergely, G., Miklósi, Á., Erdőhegyi, Á., & Csibra, G. (2008). Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation. *Science*, *321*(5897), 1831–1834. https://doi.org/10.1126/science.1161437

Toth, N., & Schick, K. (2015). Overview of Paleolithic Archaeology. In W. Henke & I. Tattersall (Eds.), *Handbook of Paleoanthropology* (pp. 2441–2464). Springer. https://doi.org/10.1007/978-3-642-39979-4_64

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., Braga Goncalves, I., Burkart, J. M., Flower, T., Gaunet, F., Glock, H. J., Gruber, T., Jansen, D. A. W. A. M., Liebal, K., Linke, A., Miklósi, Á., Moore, R., van Schaik, C. P., Stoll, S., … Manser, M. B. (2017). Exorcising Grice's ghost: An empirical approach to studying intentional communication in animals. *Biological Reviews*, *92*(3), 1427–1433. https://doi.org/10.1111/brv.12289

Vail, A. L., Manica, A., & Bshary, R. (2013). Referential gestures in fish collaborative hunting. *Nature Communications*, *4*(1), Article 1. https://doi.org/10.1038/ncomms2781

van Schaik, C. P., Pradhan, G. R., & Tennie, C. (2019). Teaching and curiosity: Sequential drivers of cumulative cultural evolution in the hominin lineage. *Behavioral Ecology and Sociobiology*, *73*(1), 2. https://doi.org/10.1007/s00265-018-2610-7

Vesper, C., Morisseau, T., Knoblich, G., & Sperber, D. (2021). When is ostensive communication used for joint action? *Cognitive Semiotics*, *14*(2), 101–129. https://doi.org/10.1515/cogsem-2021-2040

Vicari, G., & Adenzato, M. (2014). Is recursion language-specific? Evidence of recursive mechanisms in the structure of intentional action. *Consciousness and Cognition*, *26*, 169–188.

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, *17*(6), 872–879. https://doi.org/10.1111/desc.12170

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Warren, E., & Call, J. (2022). Inferential communication: bridging the gap between intentional and ostensive communication in non-human primates. *Frontiers in Psychology*, *12*, 6366.

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, *7*(3), 155–159.

Waxman, S. R., & Markow, D. B. (1995). Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*(2), 415–432. https://doi.org/10.1111/j.1420-9101.2006.01258.x

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

Whiten, A., & Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2119–2129. https://doi.org/10.1098/rstb.2012.0114

Wich, S. A., & Vries, H. de. (2005). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2005.3320

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, *308*(5955), 181–184.

Wilson, D., & Sperber, D. (2006). Relevance Theory. In L. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 606–632). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470756959.ch27

Wilson, D., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge University Press.

Wimmer, H., & Perner, J. (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

Woensdregt, M., & Smith, K. (2017). Pragmatics and Language Evolution. *Oxford Research Encyclopedia of Linguistics.* https://doi.org/10.1093/acrefore/9780199384655.013.321

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Wrangham, R. W., Jones, J. H., Laden, G., Pilbeam, D., & Conklin-Brittain, N. (1999). The Raw and the Stolen: Cooking and the Ecology of Human Origins. *Current Anthropology*, *40*(5), 567–594. https://doi.org/10.1086/300083

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841–864. https://doi.org/10.1037/rev0000153

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, *105*(36), 13690–13695.

Zahavi, A., & Zahavi, A. (1999). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford University Press.

Zohar, I., Alperson-Afil, N., Goren-Inbar, N., Prévost, M., Tütken, T., Sisma-Ventura, G., Hershkovitz, I., & Najorka, J. (2022). Evidence for the cooking of fish 780,000 years ago at Gesher Benot Ya'aqov, Israel. *Nature Ecology & Evolution*, *6*(12), Article 12. https://doi.org/10.1038/s41559-022-01910-z

Zuberbühler, K. (2018). Intentional communication in primates. *Revue Tranel (Travaux neuchâtelois de linguistique)*, *68*, 69–75.

Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20190062.

Zuberbühler, K., & Gómez, J.-C. (2018). Primate Intentional Communication. In H. Callan & S. Coleman (Eds.), *The International Encyclopedia of Anthropology* (pp. 1–10). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118924396.wbiea2211