Journal of Pragmatics 221 (2024) 123–136

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Communication: Inferring speaker intentions or perceiving the world? Insights from developmental research

Diana Mazzarella [*, 1], Edoardo Vaccargiu [1]

Cognitive Science Centre, University of Neuchâtel, Pierre-à-Mazel 7, 2000 Neuchâtel, Switzerland

ARTICLE INFO

Article history:
Available online 19 January 2024

Keywords:
Pragmatic inference
Direct perception
Trust
Epistemic vigilance
Testimony
Ontogeny

ABSTRACT

Communication is an effective and rich tool to transmit information to others. Different models of communication, though, disagree on how beliefs are acquired via testimony. While the Intentional-Inferential Model conceives communication as a matter of expressing, recognizing and evaluating intentions, the Direct Perception Model views communication as a means for direct belief transfer. What kind of experimental data can be brought to bear on this debate? We argue that developmental research can provide relevant insights for advancing the present debate. Drawing on recent experimental findings on the ontogeny of vigilance and trust, we question the idea that communication involves direct belief transfer and illustrate how children's reliance on communication is the result of smart processes of trust calibration.

1. Introduction

Communication serves the primary, although not unique, function of transmitting information. In pursuing this function, communicators typically have two distinct goals: to get their message across (to be understood) and to influence how the audience thinks or acts based on the message they have understood. For instance, when my partner utters “It is late” at a dinner party, they may want me to understand that it is late for the babysitter waiting for us at home and that we should leave the party soon. Crucially, though, they also want me to accept what they say, as well as what they implicitly communicate, as true and act accordingly. Indeed, it appears that in most circumstances “audiences cannot actually gain from communication unless they extend a degree of trust towards the communicator” (Heintz and Scott-Phillips, 2023: 8) that allows their beliefs to be updated and information to be gained.

However, belief formation and updating via testimony are open to the risk of misinformation: the speaker may accidently or intentionally communicate false information, thus preventing the audience from getting any benefit from the communicative exchange. As a result, the audience may have reasons to doubt the reliability of the message communicated. Does my partner want me to think that it is late for the babysitter just because they are bored and would like to go home? Is it really the case that we are breaking the agreement with the babysitter? (for a discussion of this example, see Sperber, 1994). From the audience's perspective, then, comprehension appears to be complemented by some epistemic assessment, which may lead to the rejection of the communicated message. I can understand what my partner wants to communicate but fail to believe it.

* Corresponding author.
E-mail addresses: diana.mazzarella@unine.ch (D. Mazzarella), edoardo.vaccargiu@unine.ch (E. Vaccargiu).
[1] Diana Mazzarella and Edoardo Vaccargiu share co-first authorship.

https://doi.org/10.1016/j.pragma.2023.12.014

Distinct models of communication offer different accounts of the relationship between comprehension and epistemic assessment. This paper focuses on two prominent models: the Intentional-inferential model, on the one hand, and the Direct Perception model, on the other hand. As it will become clear, these models hold radically different views on the place of speaker intentions in an account of ordinary communication. While the former conceives communication as involving the expression and recognition of speaker intentions, the latter assimilates communication to perceptual processes, with the audience directly "perceiving" the communicated content. This difference has important implications on the extent to which the acquisition of belief via testimony is taken to be mediated by epistemic assessment. Do we routinely evaluate speakers' intentions? Or do we believe by default everything we are told? Is epistemic assessment automatic, or is it called upon only under special circumstances? The Intentional-inferential model and the Direct Perception model depart on their answers to these questions. By grounding our discussion on experimental data, we aim to assess the debate between these two models in light of new findings in developmental psychology. We argue that the development of the epistemic attitudes towards communicated information during the first years of life can shed new light on the nature of communication and its contribution to belief formation and updating.

We proceed as follows. First, we introduce the Intentional-inferential model (Section 2) and the Direct Perception model (Section 3). Then, we provide an overview of the experimental literature that has directly addressed some of their empirical predictions, with mixed results (Section 4). Later, we illustrate the relevance of taking a developmental perspective to enrich this debate (Section 5). Finally, we take stock of the data and discuss their implications for some of the core questions in pragmatics on the nature of communication (Section 6).

2. The Intentional-inferential model

The Intentional-inferential model traces back to Paul Grice's work on speaker meaning and it was later developed in a more cognitive-oriented perspective by Sperber and Wilson's Relevance Theory. Following Neale (2022), we can think of this model as answering two distinct questions. The first is a constitutive question:

> What are the facts in virtue of which a speaker meant whatever he or she meant by uttering something on a given occasion? (Neale, 2022:126).

Grice (1957) suggested that speaker meaning is constitutively determined by the set of nested intentions the speaker has in producing her utterance. Thus, 'to mean something' reflects the speaker's intention to produce a certain response in the addressee by changing his mental representations via the recognition of the same intention. Specifically, the speaker intends the addressee.

a. to produce a particular response *r*;
b. to recognise that the speaker intends (a);
c. to fulfil (a) on the basis of the recognition of (b).

The second is an epistemic question, concerning the nature of the process through which the addressee reconstructs this intended meaning:

> What kinds of information, what principles, and what kinds of cognitive states and processes are (standardly) involved in the (typically spontaneous) arrival in the mind of an interpreter of a (typically resilient) conclusion about what a speaker means by uttering something on a given occasion? (Neale, 2022: 126).

Grice's (1975) suggestion was that hypotheses about the speaker meaning are derived via non-demonstrative inferences that are guided by the addressee's expectations that the utterance should be truthful, informative, relevant, and properly formulated. Utterance interpretation is thus conceived by Grice as an inferential process where candidate hypotheses are tested against specific pragmatic expectations raised in conversation, in line with the central assumption that interlocutors are expected to behave cooperatively.

Sperber and Wilson developed Grice's epistemological suggestion in a cognitively-oriented framework, by proposing a psychological account of utterance interpretation as guided by expectations of relevance. Importantly, while the Gricean project was mainly devoted to a philosophical analysis of speaker meaning, Relevance Theory was intended to apply to the full range of cases of ostensive (i.e., overt) communication, where the communicator overtly provides verbal and/or non-verbal evidence of their intention to convey information, and the addressee infers the content of this intention from the evidence provided (thus blurring the boundaries between 'meaning' and 'showing', Sperber and Wilson, 1995: 46–54). [1]

Sperber and Wilson claim that ostensive communication has a metarepresentational structure and build their version of the Intentional-inferential model around the hypothesis that speakers have both an *informative* and a *communicative*

[1] This more comprehensive approach is one of the major theoretical shifts from Grice to Relevance Theory. According to Sperber and Wilson, a theory of ostensive communication requires going beyond the notion of speaker meaning, by "demoting it from its central theoretical role to a loosely descriptive use that may nonetheless be adequate when dealing with fairly standard cases of linguistic communication" (Sperber and Wilson, 2015: 120).

intention (roughly corresponding to intentions a and b above). The informative intention is the speaker's intention to inform the audience of something, hence, to convey a propositional content *p* to the audience. The communicative intention is the speaker's intention to inform the audience of her informative intention, thus, to signal the addressee her will to communicate (Sperber and Wilson, 1995: 29).[2] According to Sperber and Wilson, communicative and informative intentions are generally expressed by speakers via the production of ostensive acts of communication (e.g., utterances) that are deliberately performed to make the listener attend to the speaker's intended meaning. Conversely, the recognition of an ostensive act of communication raises in the addressee pragmatic expectations about the relevance of communicated information and triggers an inferential comprehension mechanism in which candidate hypotheses about the speaker's intended meaning are tested in order of accessibility and selected once they meet the addressee's occasion-specific expectations of relevance (Sperber and Wilson, 2002). In short, Relevance Theory describes utterance interpretation as a non-demonstrative inferential process starting with a metarepresentation of an attributed utterance (i.e., the speaker has said *u*) and resulting in a metarepresentation of an attributed thought about the speaker's intended meaning (i.e., the speaker means that *p*), whose reconstruction is driven by the expectations of relevance raised by the utterance. Crucially, Sperber and Wilson argue that the output of comprehension is metarepresentational in nature as the inferred content is embedded within the metarepresentational structure of the speaker's communicative and informative intentions.

The metarepresentational structure of ostensive communication generates a distinction between comprehension (i.e., understanding that the speaker means that *p*) and epistemic acceptance (i.e., accepting *p* as true). In Sperber et al. (2010), this distinction is characterised as a function of the audience's attitude towards the speaker's informative intention. Comprehension is achieved when the speaker's informative intention is *recognized*, while the acceptance of communicated content follows from the audience's *fulfilment* of the same intention. Crucially, though, "[w]hether the informative intention itself is fulfilled depends on how much the audience trusts the communicator" (Wilson and Sperber, 2004: 611). This opens up the question of how trust is calibrated in conversation.

Communication is a rich source of information about the world, thus, extending trust towards speakers appears to be advantageous for acquiring new beliefs via testimony. However, this trust must be tentative and provisional to avoid the risk of being intentionally or accidently misinformed. According to Sperber et al. (2010), the trade-off between the benefits of communication and its risks is ensured by a suite of cognitive mechanisms for "epistemic vigilance", which evaluate incoming information and provide a cost-effective epistemic assessment of communicated contents. Specifically, epistemic vigilance is described as comprising two clusters of cognitive mechanisms. The first calibrates the addressee's trust as a function of the perceived reliability of the speaker (vigilance towards the source); the second evaluates the credibility of communicated contents relative to first-hand evidence and consistency with the hearer's background beliefs (vigilance towards the content). Importantly, Sperber and colleagues argue that epistemic vigilance is automatically triggered by the same ostensive acts that trigger comprehension, thus suggesting that "the abilities for overt intentional communication and epistemic vigilance must have evolved together, and must also develop together and be put to use together" (Sperber et al., 2010: 360). In this sense, the Intentional-inferential model conceives comprehension and epistemic assessment as two distinct processes working in parallel with each other; the first, underpinned by a relevance-guided inferential mechanism, the second, carried out by the capacity for epistemic vigilance.

3. The Direct Perception model

The Direct Perception model represents one of the most discussed alternatives to the Gricean account of communication. The model was first elaborated by the philosopher Ruth Millikan and later developed in more cognitively oriented terms by Mikhail Kissine. It involves a radical perspective on the nature of comprehension and belief acquisition via testimony, which are both assumed to be as 'direct' as perceptual processes. In what follows, we examine the Direct Perception model's analysis of these processes.

Let us begin with comprehension. Millikan conceives language understanding as a process of direct content transfer, that is, "a form of direct perception of whatever speech is *about*" (Millikan, 1984: 62). By analogy with sensory perception, she claims that "when communication proceeds 'normally' […] it is the *world*, not meanings, and not speaker intentions, that is immediately perceived when language is understood." (Millikan, 2005: 206). According to Millikan, content transfer is made possible by a reliable pairing of linguistic stimuli and cognitive responses, which allow the hearer to entertain the same content purposefully communicated by the speaker.[3] The Direct Perception model radically departs from the Gricean assumption that comprehension involves inferring some propositional content that is attributed to the speaker as her intended meaning. Indeed, according to the Direct Perception model, the output of comprehension is typically non-metarepresentational (i.e., *'p'*, rather than *'The speaker means that p'*). Millikan (2005: 203) suggests that considerations about the speaker's intentions (and beliefs) enter the picture only when the normal flow of conversation is disrupted, as they

[2] For this reason, utterance interpretation is conceived as "an exercise in mind-reading" (Sperber and Wilson, 2002: 3), since the recognition of the speaker's communicative and informative intention is part and parcel of the interpretation of ostensive acts of communication.

[3] For this reason, the Direct Perception model is often described as a code model of communication. For a critical discussion, see Origgi and Sperber (2000).

are too cumbersome to be routinely entertained in ordinary communication.[4] Millikan's argument thus hinges on the distinction between 'normal' and 'abnormal' communication. While Millikan resorts to the Gricean analysis for the latter, crucially, she provides a perception-like account of the former.

Kissine (2016) has later developed a weaker version of the dichotomy between normal and abnormal communication by spelling out a distinction between comprehension strategies involved in *typical* or *atypical* language understanding. According to Kissine, adults with a history of typical development can rely on a "Gricean, sophisticated interpretation" strategy that "involves complex inferences about speaker's communicative intentions" (Kissine, 2016: 5). However, he argues, language understanding can also be underpinned by more egocentric pragmatic processing, which characterises a simpler interpretative strategy:

> "*Egocentric relevance*: does not require any Theory of Mind and is entirely based on egocentric considerations of accessibility. The output is the representation of a certain content (viz. non-embedded within the representation of the speaker's informative intentions), and is limited to primary meanings, material implicatures and (some) indirect speech acts" (Kissine, 2016: 5)

This interpretative strategy would thus be available to non-neurotypical individuals, such as autistic individuals (for a critical discussion, see Mazzarella and Noveck, 2021).[5] In a nutshell, different versions of the Direct Perception model can vary in the extent to which they conceive communication as a direct content transfer (non-mediated by metarepresentations of the speaker's intentions). While Millikan applies this view to normal, ordinary communication, Kissine considers the obligatoriness of direct content transfer as the hallmark of atypical communication.[6]

All versions of the Direct Perception Model, though, have a similar take on the process of belief formation via testimony, to which we now turn. According to Millikan, communication is a device for direct belief transfer. Thus, understanding an utterance automatically results in believing what is communicated:

> […] coming to believe something by being told it is so, in the typical case, is the formation of a direct *perceptual* belief. Forming a belief about where Johnny is on the basis of being told where he is, is just as direct a process (and just as indirect) as forming a belief about where Johnny is on the basis of seeing him there (Millikan 2004: 120).

By blurring the distinction between hearsay and direct perception, Millikan's view echoes a long-standing approach to the epistemology of testimony that traces back to Reid (1764/2000). According to Reid, belief formation via testimony is justified by the principles of "veracity" and "credulity", in virtue of which humans are endowed by God with a disposition to speak the truth and to take what other people say as true. Setting aside religious principles, Millikan's view fits into a broad conceptual framework that explains language as the product of biological and cultural processes of evolution, whose stabilising function is the transmission of true beliefs. In a similar vein, Kissine and Klein (2013) argue that "acquiring beliefs from speech is as direct as perception" (p. 140).

Does this view imply that humans are inevitably gullible? Is this view incompatible with the very existence of epistemic vigilance? Obviously, not. Any plausible theory of human communication should account for the operations of cognitive mechanisms that can (and, as a matter of fact, often do) filter incoming information and avoid misinformation. Crucially, though, proponents of the Direct Perception model claim that epistemic vigilance is not an intrinsic feature of communication: "epistemic vigilance is not inherent in our capacity to understand others' statements" and "the cognitive processes that allow us to interpret utterances as conveying informative contents do not come with an inherent epistemic safeguard" (Kissine and Klein, 2013: 151; 146). Comprehension typically leads to the automatic acceptance of the communicated content unless specific conditions are present, which impose "boundary conditions for the operation of a filter" (Kissine and Klein, 2013: 144). Only under these special circumstances, a communicated content is epistemically evaluated and can be either disbelieved (if it had already been automatically integrated into one's beliefs) or prevented access to one's 'belief box'.

The Intentional-inferential model of communication and the Direct Perception model thus provide two different conceptualizations of the relationship between comprehension, on the one hand, and epistemic evaluation of the communicated content, on the other hand. The former considers the two processes as being closely intertwined and automatically triggered by the same set of inputs (i.e., ostensive acts). The latter thinks of epistemic assessment as an optional complement to comprehension, one which appears to be involved only under specific circumstances. This difference represents a fertile ground for empirical testing of the models. In the next section (Section 4), we illustrate the kind of experimental data that have been brought to bear on this debate so far. They concern adults' processing and evaluation of written or auditory information in various experimental settings.

[4] For a similar idea, see Recanati (2004: 38).

[5] According to Kissine (2016), while autistic individuals are limited to the *egocentric* interpretative strategy, young children can resort also to the *allocentric* interpretative strategy which allows ruling out interpretations that are incompatible with the speaker's perspective.

[6] Much in line with Millikan's view, Kissine (2016) considers *egocentric* and *allocentric relevance* as default interpretative strategies that are variably exploited by adult neurotypicals depending on the communicative situation. Instead, Gricean interpretative strategies relying on more sophisticated metarepresentations are called upon in the understanding of conversational implicatures and irony. For a similar account, see Jary (2010).

4. Empirical tests: The state of the art

Much research in cognitive psychology has examined the extent to which the epistemic assessment of communicated information is an optional process that can be disrupted under cognitive load. Initial evidence was collected by psychologist Daniel Gilbert and his colleagues in the 1990s. In a series of studies, they showed that participants under some kind of cognitive load were more likely to misremember false statements as true than true statements as false. For instance, in the classic 'Hopi paradigm' devised by Gilbert et al. (1990, Experiment 1), participants were presented with a series of sentences concerning the meaning of Hopi words (e.g., "A Monishna is a star") accompanied by an explicit indication of their truth-value (true vs false). Crucially, for some of these items, participants were asked to perform a secondary task when truth-value information became available. Finally, in a subsequent recall task, participants' memory of the truth value of the statements was tested. Interestingly, the results showed that, when distracted by the secondary task, participants tended to misremember false statements as true, while the opposite pattern (i.e., misremembering true statements as false) did not occur. This was taken to suggest that cognitive load had an impact on the process of epistemic assessment: deprived of the cognitive resources to 'disbelieve' false statements, participants would fail to revise their automatic stance of acceptance. In line with this, in Gilbert et al. (1993, Experiment 1), participants were instructed to act as jury members and carefully read crime reports containing both true and false statements, depicted, respectively, in black and red font. Half of the participants were cognitively loaded during the reading of false statements (e.g., via a secondary digit detection task). The results showed that the cognitive load group, but not the other one, systematically misremembered false statements as true, and their judgments on perpetrators' guilt were significantly influenced by the content of these statements.

Taken together, this set of data has long been interpreted as supporting the idea that while acceptance is the spontaneous and automatic attitude towards incoming information, epistemic assessment is optional and cognitively effortful. More recently, Pantazi et al. (2018) have run an auditory adaptation of Gilbert et al. (1993)'s study, in which the statements' truth-value (true vs false) was associated with two distinct voices. The results did not show any effect of the cognitive-load manipulation: in both groups (cognitive load and no cognitive load), participants' guilt judgments were equally affected by false statements and these were likely to be misremembered as true. They concluded that the tendency to accept communicated information (or 'truth bias') is even more robust than previously assumed and holds independently of distracting factors.

The Direct Perception model appears to capture these findings aptly and to give rise to the right predictions concerning the alleged priority of acceptance over epistemic assessment. It is worth noting, though, that recent research sketches a more nuanced picture of the nature of epistemic assessment. First, later adaptations of Gilbert and colleagues' studies gave rise to interesting findings, with important theoretical implications. For instance, Hasson et al. (2005) showed that when the informativeness of the stimuli is increased, the effect of cognitive load disappears. By introducing statements whose falsity was informative (e.g., "George has a television", whose falsity contradicts the expectation that people typically own televisions), they found that distracted participants were equally likely to misremember false statements as true than true statements as false (see also Richter et al., 2009, Experiment 1 for similar results). As Sperber et al. (2010) suggest, this indicates that while individuals may be disposed to accept irrelevant information as true, their epistemic vigilance is well calibrated to the relevance of the incoming information: the more relevant (informative) the content, the higher the epistemic vigilance.

Second, experimental studies using implicit measures suggest that some form of epistemic assessment is spontaneous and occurs in parallel with comprehension. Relying on a Stroop-like paradigm, Richter et al. (2009, Experiments 3 and 4) asked participants to perform orthographical judgments about the spelling of individual words which appeared one by one on a screen. Sequences of words formed sentences that were either true or false. Crucially, the results showed a significant 'interference effect': participants took longer to approve the spelling of a word at the end of a sentence that contradicted their beliefs and, conversely, to disapprove of incorrect spelling in sentences whose content was perceived to be plausible. Background beliefs thus interfere with orthographical judgments, providing direct evidence for the operations of spontaneous epistemic assessment. Further evidence can be found in the ERP studies by Hagoort et al. (2004), which investigated the neurophysiological basis of the integration of world knowledge in linguistic comprehension. They recorded the brain activity of a group of Dutch participants while reading three different versions of a sentence like "The Dutch trains are yellow/white/sour". Assuming that Dutch people have strong background knowledge about the actual colour of Dutch trains, only the first version of the sentence (i.e., "The Dutch trains are yellow") was semantically coherent and consistent with their world knowledge, whereas the other two versions contrasted either with their background beliefs ("The Dutch trains are white") or with their semantic knowledge ("The Dutch train are sour"). Crucially, results showed a comparable N400 effect for both world knowledge and semantic violations, thus providing strong evidence that background beliefs are early and implicitly integrated during comprehension. Taken together, these studies support the view that epistemic assessment is a spontaneous process, which may occur in parallel with comprehension.[7] These results are therefore compatible with the Intentional-inferential model of communication, which conceives of comprehension and epistemic assessment as parallel processes that are automatically triggered by the same stimuli.

[7] Furthermore, when participants are confronted with communicated information that openly clashes with their prior beliefs they show a consistent bias towards their own opinion (see, e.g., Trouche et al., 2018).

It is worth stressing, though, that the debate is far from being settled. Indeed, some proponents of the Direct Perception model, such as Kissine and Klein (2013), insist that under certain circumstances (such as when the content of incoming information directly contradicts one's background beliefs) epistemic vigilance can prevent automatic acceptance. According to them, this should not overshadow the fact that acceptance is otherwise the default epistemic stance (as evidenced by data on the 'truth bias' previously discussed) and that epistemic assessment is often optional. We believe that to overcome this impasse, it is necessary to enrich this debate with a new perspective and new data. In the next section, we illustrate how data from developmental psychology can shed interesting light on the relationship between comprehension and epistemic assessment, and help us advance the debate between the two models of communication under discussion.

5. A developmental perspective

In this section, we illustrate how research in developmental psychology can be brought to bear on the debate at issue. The relevance of such a developmental approach has been emphasised by advocates of the Direct Perception Model:

> This line of thought receives support from data on the ontogenesis of epistemic vigilance. The capacity to assess the reliability of a communicator is quite a precocious one. From the age of four, children are capable to discriminate between a reliable and an unreliable puppet (Clément et al., 2004). Likewise, four-year-olds tend to distrust a puppet characterised as a liar (Mascaro and Sperber, 2009). Yet, this capacity is by no means part and parcel of the processing of communicative behaviour (as the IM would have it). To begin with, the same studies also revealed that at the age of three, children fail to adopt such selective trust. Furthermore, Vanderbilt et al. (2011) show that explicitly identifying an adult as an unreliable deceiver in three consecutive communicative exchanges does not prevent four-year-olds from trusting the information communicated by this same person right afterward (Kissine and Klein, 2013: 145).

According to Kissine and Klein (2013), then, the existence of developmental evidence showing that young children fail to be vigilant in these ways is suggestive of the empirical plausibility of the model. To make sense of their claim, we assume that the Direct Perception model is taken to predict that at least in the earliest stages of development, before the emergence of independent mechanisms for epistemic vigilance (or their robust application), infants and toddlers should accept communicated information even if the informant has repeatedly misinformed them in the past, either intentionally or accidently (in line with Vanderbilt et al., 2011 mentioned above). Furthermore, they should not be able to selectively trust a more reliable informant over a less reliable one when the two communicate conflicting pieces of information, and thus be at chance in endorsing one of the two testimonies (in line with the results of three-year-olds in Clément et al., 2004; Mascaro and Sperber, 2009). More specifically, based on the quotation discussed above, we assume that the ontogeny of epistemic vigilance is understood by proponents of the Direct Perception model as a development from *indiscriminate* to *sceptical trust* (Clément et al., 2004). Whereas 'indiscriminate trust' prevents children from distinguishing between reliable and unreliable informants (e.g., based on their past accuracy), 'sceptical trust' would allow them to selectively trust previously reliable informants over unreliable ones.[8]

As the Direct Perception model maintains that epistemic vigilance is not inherent in linguistic communication, supporters of this model may insist that the ontogeny of these two mechanisms should be independent, without committing to any specific developmental trajectory. However, in light of Kissine and Klein's discussion of the developmental data reported above, the existence of a developmental stage in which the relationship between comprehension and acceptance is not mediated by any "epistemic filter" (Kissine and Klein, 2013: 145) is seen as compatible with the Direct Perception model, but not with the Intentional-inferential model. From their perspective, the preschool years would mark a gradual transition from a developmental phase characterised by indiscriminate trust to one characterised by 'sceptical trust' (in line with Clément et al., 2004; Mascaro and Sperber, 2009; Vanderbilt et al., 2011, mentioned above) as resulting from the "acquisition of […] 'sceptic' strategies" (Kissine and Klein, 2013: 145) that allow children to discriminate between reliable and unreliable informants. The emergence of new, independent, epistemic filters should lead young children to avoid automatically integrating any piece of communicated information, thus revising their default trustful attitude downwards.

Thanks to the growing literature on the ontogeny of vigilance and trust, we are now in a better position to establish whether this developmental scenario is empirically plausible. In Section 5.1 we review novel key findings on the ontogeny of vigilance, while in Section 5.2 we present recent experimental data on the ontogeny of trust in the first years of life. Taken together, these findings highlight the limits of the Direct perception model in providing a coherent explanation of the available developmental data. Finally, in Section 5.3, we address one of the uncontroversial commitments of the Direct Perception model, that is, that "any epistemic filtering of hearsay information is independent from the interpretation process"

[8] As remarked by Kissine and Klein (2013: 145) "[t]he point is not that young children are blindly gullible. They are not. For instance, children below five years tend to privilege first-hand, perceptual information over verbal claims made by an adult (Robinson et al., 1995)". This passage indicates that Kissine and Klein are not committed to the existence of a developmental stage characterised by *gullible trust*, where "children suspend or inhibit their normally routine, empirical checking […] they might accept testimony from that informant even if it contradicts what they have observed from themselves" (Clément, et al., 2004: 362–363). However, while they do not consider young children as blindly gullible, they suggest that they are indiscriminately trustful. Indiscriminately trustful receivers are able to reject testimony if this contradicts their first-hand perceptual experience, but they are not able to selectively trust reliable informants over unreliable ones. For a discussion of the notion of 'indiscriminate trust' and its relation to the Direct perception model, see Clément et al. (2004: 361).

(Kissine and Klein, 2013: 142), and we discuss experimental data on reference assignment and disambiguation which point to a role of epistemic vigilance in the very process of pragmatic comprehension.

5.1. The ontogeny of vigilance

The existence of an early developmental phase of unlimited reliance on others' testimony – which has long been taken for granted – has become an empirical question in the last few years. Indeed, only recently researchers in developmental psychology have started to systematically investigate the emergence of epistemic vigilance in infancy and toddlerhood (for reviews, see Harris and Lane, 2014; Poulin-Dubois and Brosseau-Liard, 2016). The results of this investigation suggest that even infants are sensitive to several cues of (un)trustworthiness and orient their learning behaviours accordingly. In what follows we highlight some key findings in this literature.

First, recent evidence shows that infants do not treat any potential informant as if they were equally trustworthy. For instance, Tummeltshammer et al. (2014) show that 8-month-olds are sensitive to the reliability of potential informants whose gaze direction and face orientation have provided cues to correct (reliable adult) or incorrect (unreliable adult) location of animated animals. Their findings reveal that infants were more likely to search longer in boxes cued by the reliable adult than in those cued by the unreliable one. This tendency to rely on the cue provided by the reliable adult generalised to circumstances in which the cue was directed at locations where no animated object had previously appeared. Furthermore, Bazhydai et al. (2020) show that 12-month-olds can reliably identify knowledgeable adults and selectively seek information from them in situations of epistemic uncertainty. When confronted with an impossible choice (i.e., choosing which is the *[pseudoword]* among two novel and unfamiliar objects), they systematically looked at the adult who had reliably labelled objects in a previous phase over the adult who interacted with them but was ignorant with respect to the objects' labels. This selective social referencing shows that infants are disposed to rely more on information coming from knowledgeable adults, and actively seek it even before being able to formulate verbal questions.

Second, infants calibrate their trust as a function of the perceived reliability of the informant. For instance, in a study by Varró-Horváth et al. (2017), 12- to 15-month-olds appear to lose trust in the testimony of an adult who repeatedly pointed to the wrong container when the child had to guess in which of two containers the toy was hidden. This trust loss occurs both when the adult openly expresses her ignorance, as well as when the adult tricks the child (and reveals her intention by laughing and stating "I have deceived you!"). Across both conditions, trust decreased as early as the second trial and further dropped until the fourth test trial (e.g., resulting in an acceptance rate that decreased from 87 % to 33 % in the ignorant condition, and from 90 % to 38 % in the deceptive condition). Furthermore, from very early on, infants display some selective social learning. For instance, Crivello et al. (2021) demonstrate that, when confronted with two informants whose verbal competence has been manipulated in a labelling task, 18-month-olds were less likely to learn new labels from the unreliable informant than the reliable one (see also Crivello et al., 2017; Luchkina et al., 2018).

Third, infants calibrate their trust as a function of the first-hand evidence at their disposal. Mascaro and Kovács (2022) show that both 15- and 24-month-olds rely more on their direct perception than on the testimony of others when the two provide conflicting information (see Study 2). Infants are confronted with a situation in which they have to search for a toy hidden in one of two buckets, which are transparent from their perspective, and the experimenter points towards the empty bucket. Results show that infants and toddlers do not follow the pointing, and rely on the visual information at their disposal to find the toy, thus suggesting that "infants' and toddlers' trust is not unconditional, and that it is sensitive to the quality of evidence that is pitted against pointing" (Mascaro and Kovács, 2022: 6). Furthermore, when communication contradicts their perception-based beliefs, toddlers are sensitive to the quality of the arguments presented to support the counterintuitive claim. Castelain et al. (2018) asked 24-month-olds to name hybrid animals (e.g., a bird–fish, with 75 % of the features of a bird and 25 % of the features of a fish) and later provided them with a conflicting label coming from a third-party informant (e.g., "It's a fish" after the child named the hybrid "bird"). Crucially, they manipulated the strength of the argument offered by the informant, who produced either a strong argument (e.g., "It's a fish because I saw it swimming in the water"), a circular argument ("It's a fish because I saw it's a fish") or no argument. In a second phase, participants were presented with new hybrid and conflicting labels from the same informant, but without any supporting argument. Their results reveal that toddlers kept track of the quality of the argument provided in the first phase and, as a result, they were more likely to endorse the informant's labels (even if no argumentation was involved) and to retain them if the informant had previously offered a strong argument than if they had offered a circular argument or no argument. These precocious argument evaluation skills allow toddlers to evaluate the content of the arguments and grant more trust to informants whose arguments have proven to be solid enough.

Bringing these findings together, a clear picture emerges: far from being unconditionally trusting, infants and toddlers appear to be already equipped with a suite of cognitive mechanisms that allow them to navigate their social world with epistemic vigilance. There is thus no evidence that infants and toddlers trust any informant indiscriminately or that early epistemic acceptance is not mediated by epistemic vigilance. As suggested above, such a developmental scenario would have been compatible with the Direct Perception model, but not with the Intentional-inferential model, thus supporting the former over the latter. However, in contrast with the idea that epistemic filters may complement the process of belief-formation only at a later developmental stage characterised by 'sceptical trust', even infants appear to be sensitive to the quality of the information that is provided (and its coherence with their background assumptions), and to the reliability of the individuals who dispense it. Furthermore, their vigilance does not seem to kick in only under specific circumstances, e.g.,

disruptions of the normal flow of conversation.[9] Indeed, some early vigilance appears to play a crucial role in buttressing and orienting social learning throughout development. As shown by Bazhydai et al. (2020), it is not only the acceptance of a piece of communicated information but also the very search for information from communicators to be mediated by epistemic vigilance. This sheds further doubts on the assumption that epistemic vigilance acts as a "filter" which is activated under some "boundary conditions" (Kissine and Klein, 2013: 144). Rather, a certain degree of vigilance appears to be ordinarily in place since infancy.

In sum, the purported developmental change from indiscriminate to sceptical trust that, in line with Kissine and Klein (2013), would provide support to the Direct perception model, falls short of accounting for recent evidence on the early ontogeny of epistemic vigilance.

5.2. The ontogeny of trust

The early emergence of various epistemic vigilance mechanisms may seem incompatible with the naiveté that children display until around the age of four. Much research in developmental psychology shows that infants are surprisingly oblivious to the risk of being misinformed until the age of four to five. This *prima facie* gullibility is particularly evident in experimental settings involving what Mascaro and Morin (2014) call "false communication tasks", that is, tasks where a single unreliable informant communicates a piece of false information to the child, who in turn has to infer that the opposite is true (and act accordingly). In Couillard and Woodward (1999), for instance, 3- and 4-year-olds had to find a sticker hidden in one of two opaque buckets by relying on the pointing gesture of an informant, who repeatedly deceived the child and was thus described as "tricky". To succeed, children had to come to mistrust the misleading informant and infer that the sticker was hidden in the bucket that the informant had not pointed at. Interestingly, the results showed that children have a robust tendency to trust the informant and endorse their testimony, irrespective of how misleading or mistaken they have proved to be (see, *inter alia,* Call and Tomasello, 1999; Heyman et al., 2013; Mascaro and Sperber, 2009; Vanderbilt et al., 2011).

At first glance, these data would speak in favour of the assumption that young children are *indiscriminately* trustful. They would lack the mechanisms for epistemic vigilance — or at least the cognitive resources to exercise it — and thus fail to reject misleading information. In other terms, in line with the predictions of the Direct Perception model, young children would fall short of revising the default acceptance that comes with comprehension. It is worth noting, though, that this interpretation of the literature on false communication tasks is hardly compatible with the growing body of evidence on the emergence of epistemic vigilance in infancy (discussed in the previous section). How could preschoolers be so oblivious to misinformation while infants show fairly sophisticated epistemic vigilance? If vigilance emerges precociously, why are preschoolers not capable of rejecting the misleading testimony of an unreliable informant? To address this puzzle, we now focus on the ontogeny of trust and its developmental changes from infancy to childhood. In what follows, we outline a solution to this apparent puzzle which relies on a closer inspection of the nature of children's trust in communicated information and its dynamic development.

To begin with, let us consider the relationship between epistemic vigilance, on the one hand, and trust, on the other hand. The Direct Perception model conceives epistemic vigilance as a filter that disrupts the automatic process of default acceptance of communicated information. More specifically, Kissine and Klein (2013) argue for the optional nature of vigilance filters, as means of epistemic safeguards that are activated only under specific circumstances (i.e., when testimony clashes with background beliefs or first-hand perception). Furthermore, they conceive the development of epistemic vigilance in terms of the "acquisition of […] 'sceptic' strategies" (Kissine and Klein, 2013: 145), thus framing it in light of the previously discussed distinction between 'indiscriminate' and 'sceptical trust'. Departing from this view, we explore the assumption that "[v]igilance (unlike distrust) is not the opposite of trust; it is the opposite of blind *[or indiscriminate]* trust" (Sperber et al., 2010: 363, *our addition in brackets*). Indeed, according to Sperber and colleagues, being vigilant does not amount to being *a priori* sceptical and mistrustful towards informants. Epistemic vigilance is first and foremost a suite of mechanisms for trust calibration whose job is to regulate individuals' expectations about the reliability of communicative acts. According to this view, trust itself is buttressed by epistemic vigilance: it is by exercising some mutual vigilance that individuals come to trust each other in most ordinary — and benevolent — interactions, and thus rely on each other to gain information about the world. As a result, a modicum of epistemic vigilance is needed to set one's priors about the reliability of communicated information by modulating the expectations of trustworthiness in communication; thus, understanding how these priors evolve across development can be fruitful in explaining the seemingly contradictory data presented so far.

Mascaro and Morin (2011, 2014) argue that children before the age of 4 have particularly strong and optimistic expectations of trustworthiness towards communicated information. According to them, these expectations are (at least in part) the result of the trust-inducing social environment that characterises the first years of life. In this period, children are typically

[9] It must be noted that proponents of the Direct Perception model provide different characterisations of disruptions of the normal flow of conversation. According to Millikan (2005), normal conversation is determined by the direct proper function of informative acts (i.e., transferring beliefs among interlocutors), which purports to explain the ultimate survival of linguistic communication throughout biological and cultural evolution. Instead, Kissine and Klein (2013) loosely describe such interruptions in terms of conflict between the communicator's testimony and the addressee's background beliefs or first-hand experience. Crucially, though, as discussed by Sperber et al. (2010), it is arguably the case that some degree of vigilance is required *a priori* to detect such disruptions in the first place. Kissine and Klein's (2013) description of epistemic vigilance as optional filtering mechanisms falls short of accounting for this role.

exposed to daily interactions with a limited number of benevolent caregivers, whose interests tend to converge with those of the child. In such a social environment, a trustful attitude supports social and cultural learning processes from beneficial interactions with caregivers (Csibra and Gergely, 2009; Harris et al., 2018; Tomasello, 2016). Frequent opportunities to learn from (few) benevolent informants constitute a significant part of children's social environment during preschool years, thus fostering strong expectations of trustworthiness towards communicators and incoming information. Furthermore, as discussed by Mascaro et al. (2017), these optimistic expectations also play an important role in the process of acquiring and practising linguistic and pragmatic abilities, which require seizing the opportunities to communicate and treating incoming information as worthwhile, at the cost of being eventually oblivious to deception and misinformation.

Crucially, in experimental settings involving false communication tasks, young children are confronted with situations in which their strong expectations of trustworthiness, otherwise well-justified in their daily interactions, interfere with their ability to reject the information communicated. It is for this reason, according to Mascaro and Morin (2011), that children's later success in this kind of task goes hand in hand with important changes in their social environment. Indeed, typically around the age of four, peer-to-peer interactions progressively acquire a more prominent role in the social life of the child. Children can thus experience, in a more systematic way, the fact that interlocutors are neither always competent nor always benevolent, and can be more inclined than caregivers to deceive them. Furthermore, they find themselves, for the first time, in a partner-choice ecology, in which detecting malevolence is useful to avoid uncooperative relationships (see Mascaro and Morin, 2011: 33–34). The emergence of this "sense of deceit", and the revision of the optimistic expectations of trustworthiness that it brings about, can thus explain why children start passing false communication tasks at around the age of four and systematically succeed at around the age of five. The revision of such expectations would make children more aware of the potential risks of communication and more sensitive to the possibility that informants can be unreliable because of a lack of knowledge or misleading communicative intentions.[10]

By looking at the dynamics of trust across development, it is thus possible to reconcile the early emergence of some mechanisms dedicated to epistemic vigilance with the robustness of young children's tendency to trust informants and their obliviousness to the risk of misinformation. According to this developmental-social hypothesis, pre-schoolers' trustful attitude towards communicators is the product of early mechanisms for epistemic vigilance which calibrate children's trust to the experience provided by their social environment. Later on, as children become more acquainted with deceitful affordances within peer-to-peer interactions, epistemic vigilance gets modulated to this newfound social ecology, thus bringing about a downgrade revision of their baseline positive assumptions about the reliability of communicators and communication. In this sense, rather than being blind or indiscriminate, children's trust is well-calibrated to the ever-changing experience provided by their social environment, and epistemic vigilance is the set of (early) cognitive mechanisms responsible for this modulation.[11]

To further probe this smart conception of trust, Mascaro and Kovács (2022) have recently investigated the very emergence of the disposition to trust communicated information in infancy and toddlerhood. By comparing the performance of 15- and 24-month-olds in a simple false communication task, their objective was to establish whether trust would increase during the second year of life. Indeed, this developmental trajectory can tell us much about the very nature of the disposition to trust communicated information. Specifically, if the disposition to trust is the result of an appropriate calibration to the regular positive experience with reliable communication, it should be strengthened during the second year of life. In Study 1, 15- and 24-month-olds were asked to find a reward hidden under one of two containers. They first saw in which container the reward was hidden and they were later exposed to the testimony of the adult informant, who always pointed towards the incorrect container. The testimony thus contradicted the participant's perception-based belief concerning the actual location of the reward. Crucially, while both 15- and 24-month-olds showed the tendency to follow the pointing (over their own belief), this tendency was significantly stronger for the 24- than 15-month-olds, thus revealing a developmental change in young children's reliance on communicated information that is compatible with the smart conception of trust. Furthermore, in another series of studies, the authors investigated the extent to which children's trust may be calibrated as a function of the informant's signalling of her intention to communicate. Specifically, they measured whether toddlers' reliance on testimony was influenced by the presence of specific "ostensive cues", such as direct eye gaze and infant-directed speech, that make salient the informant's communicative intentions (Csibra, 2010). Firstly, 24-month-olds were familiarised with a novel cue (a plastic marker) which was repeatedly placed above the correct container to induce participants to interpret it as a reliable indicator of the reward's location. Then, in the ostensive condition (Studies 3 and 4a), children were involved in a false communication task almost identical to the one previously described, except that instead of cueing by pointing, the informant addressed ostensively the child (by saying "*Look*" in infant-directed speech while establishing eye contact) before placing the marker on the top of the empty container. By contrast, in the non-ostensive condition (Study 4) the marker was glued on the

[10] For recent evidence of how training with strategic deception can enhance preschoolers' performance in resisting information from misleading informants, see Ding et al. (2022).

[11] It must be noted that the developmental-social hypothesis advanced by Mascaro and Morin (2011, 2014) relies on an intuitive comparison between the social ecologies characterising preschool years and school age, thus assuming that adult caregivers are benevolent and accurate communicators in the early stages of children's development. As one of the reviewers pointed out, this can be partially disputed by the contrasting intuition whereby adults' child-directed communication begins in a rather inconsistent and imprecise manner and progressively becomes more accurate as children mature and become more inquisitive. Though further empirical research is needed to adjudicate between these clashing intuitions, we find it uncontroversial to assume that adult caregivers are essentially benevolent communicators.

top of the wrong container, and the experimenter did not produce any cue indicating her intention to communicate by means of it. Interestingly, results revealed that 24-month-olds were significantly more inclined to trust the marker when it was accompanied by ostensive cues while they relied more on their own perception-based belief in the non-ostensive condition. Thus, as the authors conclude, "experience with the accuracy of cues alone cannot explain toddlers' trust in them. Instead, our data suggest that the communicative use of cues plays a central role in triggering toddlers' trust" (Mascaro and Kovács, 2022: 10).

Contrary to the idea that indiscriminate trust is the earliest epistemic attitude, this body of evidence shows that young children's trust is smart and calibrated during early ontogeny, thus undermining the hypothesis that they would first trust unconditionally and revise their strong disposition thanks to the later acquisition of epistemic vigilance mechanisms. This disposition to trust appears to develop during toddlerhood and is particularly robust until the age of four when children become more attuned to the opportunities to deceive and be deceived.

Bringing together data on the early ontogeny of vigilance (Section 5.1) with data on the ontogeny of trust (Section 5.2), we can draw the following conclusions. First, there is no evidence for a developmental stage in which the relationship between comprehension and acceptance is not mediated by epistemic vigilance. Second, recent findings suggest that young children's trust is smart since very early in ontogeny, and epistemic vigilance buttresses both trustful and cautious epistemic attitudes by calibrating the expectations raised by communicative acts. This picture indicates that, even during infancy, indiscriminate trust or default acceptance are not the basic epistemic attitudes, thus highlighting some important implications for the debate between the Direct Perception model and the Intentional-inferential model.

First, it rules out the only developmental scenario that is straightforwardly compatible with the Direct perception model, but not with the Intentional-inferential model, i.e., that there is an early developmental stage characterised by default acceptance and lack of vigilance. Second, against the most radical versions of the Direct Perception model (à la Gilbert, see Sect. 4), it disconfirms the view that epistemic vigilance is optional and effortful, and thus beyond the reach of infant cognition. Finally, against all versions of the Direct Perception model, it emphasises the limits of conceiving epistemic vigilance as an additional, independent "filter" that operates on the process of belief formation via testimony.

5.3. Vigilant interpreters

As already discussed (Section 3), the nature of the relationship between pragmatic comprehension and epistemic vigilance capacities marks a critical distinction between the Direct Perception model and the Intentional-inferential model. While the Intentional-inferential model conceives the two as deeply intertwined capacities that work in parallel and are automatically triggered by ostensive acts of communication, the Direct Perception model denies that epistemic vigilance is part and parcel of the processing of communicative behaviours: "any epistemic filtering of hearsay information is independent from the interpretation process" (Kissine and Klein, 2013: 142). The issue of the interplay between comprehension and epistemic vigilance has been discussed in theoretical pragmatics (see, for instance, Mazzarella, 2013, 2016; Padilla-Cruz, 2012), but alternative positions have not been tested experimentally. To what extent can developmental data help shed light on these alternative conceptualisations?

We suggest that there are at least two sets of developmental data that are promising concerning the question of the interplay between pragmatic comprehension and epistemic vigilance. The first type of data comes from the empirical investigation of preschoolers' interpretation of novel communicative acts. As discussed in Section 5.2, smart trust can lead preschoolers to display some *prima facie* naiveté in assigning meaning to novel communicative behaviours. For instance, Mascaro and Sperber (2019) showed that unfamiliar ostensive acts, such as placing a marker or an arrow to (deceptively) indicate the box containing a reward, raise expectations of trustworthiness strong enough to bring 4-year-olds, after a certain amount of trials, to reinterpret the meaning of the ostensive act so as to indicate the box where the reward was *not* located. These findings indicate that the optimistic expectations modulated by epistemic vigilance can influence and redirect the upshot of pragmatic comprehension, thus allowing for the possibility that the later fine-tuning of vigilance mechanisms could play a prominent role in children's development of pragmatic abilities.[12]

The second type of data comes from research on reference assignment in contexts of false belief or ignorance, i.e. where the informant's epistemic incompetence and unreliability are made particularly salient in the experimental setting. More precisely, these studies assess young children's ability to take into account the epistemic competence of the speaker and interpret their utterances accordingly. As an example, Southgate et al. (2010) conducted a false-belief referential task which required infants to interpret new labels and pronouns based on the speaker's false beliefs. The experiment involved 17-month-olds who were introduced to two unfamiliar objects placed in separate boxes by the experimenter. After the experimenter left the room, an accomplice changed the positions of the objects. When the experimenter returned, they pointed to one of the boxes while uttering the following statements: "Do you know what's in here? There's a *Sefo* in the box! Can you get *it* for me?". To assign the correct referent to the new label "Sefo" and the pronoun "it", thus succeeding in the task,

[12] Crucially, according to the Intentional-inferential model, revising one's expectations towards communication would contribute to the scaffolding of more sophisticated strategies of utterance interpretation that are sensitive to the dishonesty of the speaker. Thus, a refinement of interpreters' vigilance may also buttress the development of more complex pragmatic abilities. For a proposal along these lines, focused on the developmental trajectory of irony understanding, see Mazzarella and Pouscoulous, 2021.

participants must take into account that the speaker held a false belief about the location of the objects; hence, they need to understand that the experimenter was referring to the object that was in the box she had *not* pointed to and respond accordingly. Most of the 17-month-olds successfully passed the test, thus demonstrating that their pragmatic interpretation of the speaker's utterances depended on their assessment of the speaker's epistemic (in)competence in the context at issue (for a thorough discussion of this and other relevant studies, see Mazzarella and Pouscoulous, 2021).[13]

More recently, the 'Sefo task' devised by Southgate and colleagues (2010) has been replicated with both positive (Király et al., 2018) and negative results (Dörrenberg et al., 2018; Wenzel et al., 2020). Interestingly, by re-analysing data from these studies, Mascaro and Kovács (2022) found that participants' success on the Sefo task decreases with age. That is, while 17-month-olds prove to be more likely to assign the correct referent to the label "Sefo" by disregarding the pointing of a false believer, 24- and 30-month-old participants tend to follow the pointing without taking their belief into account. Taken together, such attempted replications indicate that young children's tendency to follow the experimenter pointing to fulfil their ambiguous request increases from 17 to 30 months of age, thus further corroborating the developmental trajectory observed in recent studies on the ontogeny of smart trust during toddlerhood (see Section 5.2).

While more empirical research is needed to aptly figure out the nature of the relationship between pragmatic comprehension and epistemic vigilance, and their interplay with children's smart trust in communication, these findings appear to be consistent with a model of communication, like the Intentional-inferential model, that conceives of comprehension and epistemic vigilance as deeply intertwined in their co-development.

6. General discussion

In the previous sections, we have illustrated how a variety of findings from the developmental literature can be brought to bear on the debate between alternative models of human communication. While these findings are far from being conclusive, they allow us to take a new enriching perspective on this debate.

The Direct Perception model and the Intentional-inferential model are ultimately grounded on two contrasting conceptualisations of the capacity for epistemic vigilance: the Intentional-inferential model construes vigilance as an early-developing capacity for epistemic trust calibration; instead, the Direct Perception model describes it as an optional (and possibly effortful) filtering mechanism that occasionally prevents automatic acceptance of communicated contents. Throughout this paper, we have argued that these two conceptualisations go together with two different ways of understanding the ontogeny of epistemic vigilance.

Let us start by reviewing the empirical data that both models can account for. As discussed before, both models can arguably explain the finding that infants' trust is not gullible but well calibrated to the first-hand evidence at their disposal, as evidenced by Mascaro and Kovács (2022, Study 2; see Section 5.1). On the one hand, the Direct Perception model relies on the claim that epistemic vigilance mechanisms of coherence checking are activated by the detection of a conflict between communicated information and first-hand evidence, a special circumstance (or "boundary condition") that would trigger the operation of epistemic vigilance.[14] On the other hand, the Intentional-inferential model maintains that the "same background information which is used in the pursuit of relevance can also yield an imperfect but cost-effective epistemic assessment" (Sperber et al., 2010, p. 374), thus emphasising the link between comprehension and epistemic assessment. Although equally compatible with the available data, these two proposals differ with respect to their psychological plausibility. Contrary to the Direct Perception model, the Intentional-Inferential model, does not require any ad-hoc activation of vigilance mechanisms and specifies the scope of such a process of coherence checking in a principled way, thus providing a parsimonious analysis of the cognitive cost involved that is arguably better suited to describe the type of cognitive resources mobilised in infant cognition.

Let us now turn to the discussion of the developmental trajectory that characterises the ontogeny of vigilance. Based on Kissine and Klein (2013), we showed that the Direct Perception model describes this trajectory as resulting from the acquisition of sceptic strategies that bring about a developmental change from indiscriminate to sceptical trust occurring in the preschool years (see Section 5). While this view found some support from early data on children's selective trust abilities, it falls short of explaining recent findings on the precocious emergence of vigilance in infancy (see Section 5.1). Furthermore, the attested role played by epistemic vigilance in orienting infants' learning behaviours (see Bazhydai et al., 2020) highlights the limits of conceiving it as an optional mechanism (an "epistemic filter") activated only under special circumstances, such as when a conflict between testimony and first-hand experience emerges.

[13] Despite Southgate et al.'s (2010) paradigm being based on a false belief context, we are not committed to the view that epistemic vigilance necessarily requires full-blown metarepresentational abilities. By definition, epistemic vigilance comprises a suite of mechanisms that can vary on the extent to which they rely (or not) on metarepresentational resources. Data on the ontogeny of vigilance suggest that infants' selective trust is guided by early detection of cues of reliability that presumably do not recruit metarepresentational processes (e.g., Tummeltshammer et al., 2014; Bazhydai et al., 2020; see Sect. 5.1). In line with this, Ostashchenko et al. (2020) show that 7-year-old children with Autism Spectrum Disorder are able to selectively trust an accurate speaker over an inaccurate one on the basis of such surface cues, while they perform worse than neurotypical peers in tasks that require tailoring epistemic trust to more metarepresentational assessment. This suggests that, even if higher metarepresentational proficiency can positively impact the sophistication of epistemic vigilance, the calibration of epistemic trust does not necessarily rely on metarepresentational abilities.

[14] It is worth noting that if such a conflict is a necessary condition to activate vigilance, this account does not specify the nature of the mechanisms which would identify the presence of a conflict in the first place.

An early capacity for epistemic vigilance, however, seems to be *prima facie* incompatible with preschoolers' robust credulity in false communication tasks (see Section 5.2). We have argued that these seemingly contrasting data can be coherently explained by combining the early ontogeny of epistemic vigilance with a smart conception of trust, as suggested by proponents of the Intentional-inferential model. According to this view, epistemic trust calibration would be the result of a process of vigilant monitoring of the environment (Sperber et al., 2010), and would thus be positively affected by children's positive experience with benevolent caregivers in the first years of life. From this developmental-social perspective, young children's trust in communication should increase over the first years of life, to be then revised downward as they become more sensitive to the possible risks of communication within peer-to-peer interactions (Mascaro and Morin, 2014). Interestingly, this view finds preliminary support from the developmental trajectory observed in Mascaro and Kovács' (2022) recent comparison between 15- and 24-month-olds' performances on false communication tasks, which shows an increasing reliance on communicated information from infancy to toddlerhood that is compatible with a smart conception of trust.

The Intentional-inferential model maintains that ostensive acts of communication raise in the audience positive expectations (cooperativeness, informativeness, relevance, and so on) that are not raised by ordinary actions (see Wilson, 2005, p. 1138). Importantly, ostensive forms of communication permeate daily caregiving relationships, and the use of ostensive cues in learning contexts has long been taken to play a facilitating role in the transfer and the acquisition of generic knowledge throughout the early stages of development (Csibra and Gergely, 2009). For this reason, following Mascaro and Kovács (2022), it is reasonable to assume that children's strong expectations of trustworthiness can be enhanced when information is provided ostensively, and that their vigilance and trust calibration may be properly attuned to these child-directed forms of ostensive communication (for a similar interpretation of converging results, see Jaswal, 2004; Heyman et al., 2013).

Finally, the two models at issue diverge in the way through which they construe the relationship between comprehension and epistemic vigilance. The experiments discussed in Section 5.3, specifically the 'Sefo task' by Southgate et al. (2010), show that infants' pragmatic comprehension can be influenced by epistemic assessment, thus empirically supporting the Intentional-inferential view.

Building on this, we suggest that a conception of epistemic vigilance as a composite suite of trust calibration mechanisms which buttress both trustful and cautious epistemic attitudes provides a more explanatory and comprehensive framework for interpreting data from recent developmental research, and should thus be preferred over a conception of vigilance as an optional filter that occasionally allows children to adopt a sceptical stance towards informants. Furthermore, combining the early ontogeny of epistemic vigilance with a smart conception of trust allows us to overcome the rigid distinction between indiscriminate and sceptical trust implicitly endorsed by proponents of the Direct Perception model. This is heuristically valuable to direct further systematic research on young children's epistemic attitudes toward testimony. For these reasons, we argue, the Intentional-inferential model should be preferred over the Direct Perception model on both empirical and metatheoretical grounds.

7. Conclusions

The Intentional-inferential model and the Direct Perception model offer two alternative accounts of the nature of communication and the process of belief formation via testimony. In this paper, we focused on how they conceive the relationship between comprehension, on the one hand, and epistemic assessment, on the other hand. By drawing on the most recent findings in developmental research, we put these models to the test. We argued that the Intentional-inferential model is better positioned than the Direct Perception model to account for the following three observations. First, children display early vigilance towards communicated information and do not indiscriminately rely on others to acquire beliefs about the world. Second, children's trust is smart: far from being blindly or indiscriminately trustful, children calibrate their trust during the first years of life. Third, children's pragmatic interpretation is directly affected by their vigilance and trust calibration. This opens up a new testing ground for models of communication and illustrates the relevance of turning to infant research in the study of pragmatics.

Funding

This work was supported by the University of Neuchâtel (*Fonds des donations*) and by the Swiss National Science Foundation Eccellenza Grant 186931 (awarded to DM).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Bazhydai, Marina, Westermann, Gert, Parise, Eugenio, 2020. “I don't know but I know who to ask”: 12-month-olds actively seek information from knowledgeable adults. Dev. Sci. 23 (5), e12938.

Call, Josep, Tomasello, Michael, 1999. A nonverbal false belief task: the performance of children and great apes. Child Dev. 70 (2), 381–395.

Castelain, Thomas, Bernard, Stéphane, Mercier, Hugo, 2018. Evidence that two-year-old children are sensitive to information presented in arguments. Infancy 23 (1), 124–135.

Clément, Fabrice, Koenig, Melissa, Harris, Paul L., 2004. The ontogenesis of trust. Mind Lang. 19 (4), 360–379.

Couillard, Nicole L., Woodward, Amanda L., 1999. Children's comprehension of deceptive points. Br. J. Dev. Psychol. 17 (4), 515–521.

Crivello, Cristina, Grossman, Shawna, Poulin-Dubois, Diane, 2021. Specifying links between infants' theory of mind, associative learning, and selective trust. Infancy 26 (5), 664–685.

Crivello, Cristina, Phillips, Sara, Poulin-Dubois, Diane, 2017. Selective social learning in infancy: looking for mechanisms. Dev. Sci. 21 (3), e12592.

Csibra, Gergely, Gergely, György, 2009. Natural pedagogy. Trends Cognitive Sci. 13 (4), 148–153.

Csibra, Gergely, 2010. Recognizing communicative intentions in infancy. Mind Lang. 25 (2), 141–168.

Ding, Xiao P., Lim, Hui Y., Heyman, Gail D., 2022. Training young children in strategic deception promotes epistemic vigilance. Dev. Psychol. 58 (6), 1128–1138.

Dörrenberg, Sebastian, Rakoczy, Hannes, Liszkowski, Ulf, 2018. How (not) to measure infant Theory of Mind: testing the replicability and validity of four non-verbal measures. Cognit. Dev. 46, 12–30.

Gilbert, Daniel T., Krull, Douglas S., Malone, Patrick S., 1990. Unbelieving the unbelievable: some problems in the rejection of false information. J. Personal. Social Psychol. 59 (4), 601–613.

Gilbert, Daniel T., Tafarodi, Romin W., Malone, Patrick S., 1993. You can't not believe everything you read. J. Personality Social Psychol. 65 (2), 221–233.

Grice, Paul, 1957. Meaning. Phil. Rev. 66, 377–396.

Grice, Paul, 1975. Logic and conversation. In: Cole, P., Morgan, J. (Eds.), Syntax and Semantics, Vol. 3: Speech Acts. Academic Press, New York, pp. 41–58.

Hagoort, Peter, Hald, Lea, Bastiaansen, Marcel, Petersson, Karlus M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304, 438–441.

Harris, Paul L., Lane, Jonathan D., 2014. Infants understand how testimony works. Topoi 33, 443–458.

Harris, Paul L., Koenig, Melissa A., Corriveau, Kathleen H., Jaswal, Vikram K., 2018. Cognitive foundations of learning from testimony. Annu. Rev. Psychol. 69, 251–273.

Hasson, Uri, Simmons, Joseph P., Todorov, Alexander, 2005. Believe it or not: on the possibility of suspending belief. Psychol. Sci. 16, 566–571.

Heintz, Cristophe, Scott-Phillips, Thom, 2023. Expression Unleashed: the evolutionary and cognitive foundations of human communication. Behav. Brain Sci. 46, e1.

Heyman, Gail D., Sritanyaratana, Lalida, Vanderbilt, Kimberly E., 2013. Young children's trust in overtly misleading advice. Cognit. Sci. 37, 646–667.

Jary, Mark, 2010. Assertion. Palgrave Macmillan, London.

Jaswal, Vikram K., 2004. Don't believe everything you hear: preschoolers' sensitivity to speaker intent in category induction. Child Dev. 75 (6), 1871–1885.

Király, Ildikó, Oláh, Katalin, Csibra, Gergely, Kovács, Ágnes Melinda, 2018. Retrospective attribution of false beliefs in 3-year-old children. Proc. Natl. Acad. Sci. USA 115 (45), 11477–11482.

Kissine, Mikhail, Klein, Olivier, 2013. Models of communication, epistemic trust and epistemic vigilance. In: Forgacs, J.P., Vincze, O., Lázló, J. (Eds.), Social Cognition and Communication. Psychology Press New York, NY, pp. 139–154.

Kissine, Mikhail, 2016. Pragmatics as metacognitive control. Front. Psychol. 6, 2057.

Luchkina, Elena, Sobel, David M., Morgan, James L., 2018. Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. Dev. Sci. 21 (6), e12663.

Mascaro, Olivier, Kovács, Ágnes Melinda, 2022. The origins of trust: humans' reliance on communicative cues supersedes firsthand experience during the second year of life. Develop. Sci., e13223

Mascaro, Olivier, Morin, Olivier, 2011. L'éveil du mensogne. Terrain 57, 20–35.

Mascaro, Olivier, Morin, Olivier, 2014. Gullible's travel: how honest and trustful children become vigilant communicators. In: Robinson, E.J., Einav, S. (Eds.), Trust and Skepticism: Children's Selective Learning from Testimony. Psychology Press, pp. 69–82.

Mascaro, Olivier, Sperber, Dan, 2019. The pragmatic role of trust in young children's interpretation of unfamiliar signals. PLoS ONE 14 (10), e0224648.

Mascaro, Olivier, Morin, Olivier, Dan, Sperber, 2017. Optimistic expectations about communication explain children's difficulties in hiding, lying, and mistrusting liars. J. Child Lang. 44 (5), 1041–1064.

Mazzarella, Diana, 2013. Optimal relevance as a pragmatic criterion: the role of epistemic vigilance. UCL Working Pap. Linguist. 25, 20–45.

Mazzarella, Diana, 2016. Pragmatics, modularity and epistemic vigilance. Argumenta 1 (2), 181–193.

Mazzarella, Diana, Noveck, Ira, 2021. Pragmatics and mindreading: the puzzle of autism (Response to Kissine). Language 97 (3), e198–e210.

Mazzarella, Diana, Pouscoulous, Nausicaa, 2021. Pragmatics and epistemic vigilance: a developmental perspective. Mind Lang. 36 (3), 355–376.

Millikan, Ruth G., 1984. Language, Thought and Other Biological Categories. MIT Press, Cambridge.

Millikan, Ruth G., 2004. Varieties of Meaning. MIT Press, Cambridge.

Millikan, Ruth G., 2005. Language: a Biological Model. Oxford University Press, Oxford.

Neale, Stephen, 2022. Determinations of meaning. In: LePore, E., Sosa, D. (Eds.), Oxford Studies in the Philosophy of Language, vol. 2. Oxford University Press, pp. 104–143.

Origgi, Gloria, Sperber, Dan, 2000. Evolution, communication and the proper function of language. In: Carruthers, P., Chamberlain, A. (Eds.), Evolution and the Human Mind. Cambridge University Press, Cambridge, pp. 140–169.

Ostashchenko, Ekaterina, Deliens, Gaétane, Durrleman, Sthephanie, Kissine, Mikhail, 2020. An eye-tracking study of selective trust development in children with and without autism spectrum disorder. J. Exp. Child Psychol. 189, 104697.

Padilla Cruz, Manuel, 2012. Epistemic vigilance, cautious optimism and sophisticated understanding. Res. Lang. 10 (4), 365–386.

Pantazi, Myrto, Kissine, Mikhail, Klein, Olivier, 2018. The Power of Truth Bias: false information affects memory and Judgment even in the absence of distraction. Soc. Cognit. 36 (2), 167–198.

Poulin-Dubois, Diane, Brosseau-Liard, Patricia, 2016. The developmental origins of selective social learning. Curr. Dir. Psychol. Sci. 25 (1), 60–64.

Recanati, François, 2004. Literal Meaning. Cambridge University Press, Cambridge.

Reid, Thomas, 2000. An Inquiry into the Human Mind on the Principles of Common Sense: A Critical Edition. Penn State Press, University Park, PA.

Richter, Tobias, Schroeder, Sascha, Wöhrmann, Britta, 2009. You don't have to believe everything you read: background knowledge permits fast and efficient validation of information. J. Personality Social Psychol. 96, 538–558.

Robinson, Elizabeth J., Mitchell, Peter, Nye, Rebecca, 1995. Young children's treating of utterances as unreliable sources of knowledge. J. Child Lang. 22 (3), 663–685.

Southgate, Victoria, Chevallier, Coralie, Csibra, Gergely, 2010. Seventeen-month-olds appeal to false belief to interpret others' referential communication. Dev. Sci. 13 (6), 907–912.

Sperber, Dan, 1994. Understanding verbal understanding. In: Khalfa, J. (Ed.), What Is Intelligence? Cambridge University Press, Cambridge, pp. 179–198.

Sperber, Dan, Clément, Fabrice, Heintz, Cristophe, Mascaro, Olivier, Mercier, Hugo, Origgi, Gloria, Wilson, Deirdre, 2010. Epistemic vigilance. Mind Lang. 25 (4), 359–393.

Sperber, Dan, Wilson, Deirdre, 1995. Relevance: Communication and Cognition. Blackwell, Oxford.

Sperber, Dan, Wilson, Deirdre, 2002. Pragmatics, modularity and mind-reading. Mind Lang. 17, 3–23.

Sperber, Dan, Wilson, Deirdre, 2015. Beyond speaker's meaning. Croat. J. Philos. XV (44), 117–149.
Tomasello, Michael, 2016. The ontogeny of cultural learning. Curr. Opinion Psychol. 8, 1–4.
Trouche, Emmanuel, Johansson, Peter, Hall, Lars, Mercier, Hugo, 2018. Vigilant conservatism in evaluating communicated information. PLoS One. 13 (1), e0188825.
Tummeltshammer, Kristen S., Wu, Rachel, Sobel, David M., Kirkham, Natasha Z., 2014. Infants track the reliability of potential informants. Psychol. Sci. 25, 1730–1738.
Vanderbilt, Kimberly E., Liu, David, Heyman, Gail D., 2011. The development of distrust. Child Dev. 82 (5), 1372–1380.
Varró-Horváth, Diána Á., Dorn, Krisztina, Lábadi, Beatrix, 2017. Understanding deceptive intentions behind pointing gestures in 12-15-month-old infants. Infant Behav. Dev. 47, 121–124.
Wenzel, Lisa, Dörrenberg, Sebastian, Proft, Marina, Liszkowski, Ulf, Rakoczy, Hannes, 2020. Actions do not speak louder than words in an interactive false belief task. R. Soc. Open Sci. 7, 191998.
Wilson, Deirdre, Sperber, Dan, 2004. Relevance theory. In: Horn, L.R., Ward, G.L. (Eds.), Handbook of Pragmatics. Blackwell Publishing, Oxford, pp. 607–632.
Wilson, Deirdre, 2005. New directions for research on pragmatics and modularity. Lingua 115 (8), 1129–1146.

Diana Mazzarella is a Professor in Communication and Cognitive Science at the University of Neuchâtel, Switzerland. She has a PhD in Linguistics from University College London. Her main research interests are in theoretical and experimental pragmatics, particularly the interface between linguistic communication and other cognitive systems, such as Theory of Mind and epistemic vigilance.

Edoardo Vaccargiu is a Postdoctoral researcher at the Cognitive Science Centre of the University of Neuchâtel. He has a PhD in Philosophy of Mind and Language from the University of Genoa. He works on foundational issues in cognitive pragmatics with a focus on cognitive development.