
Millikan's Teleosemantics and Communicative Agency

Pierre Jacob

Millikan's teleosemantic approach constitutes a powerful framework for explaining the continued reproduction and proliferation of intentional conventional linguistic signs, and thereby the stability of human verbal intentional communication. While this approach needs to be complemented by particular proximate psychological mechanisms, Millikan rejects the mentalistic psychological mechanisms, which are part of the Gricean tradition in pragmatics. The goal of this paper is to assess the balance between Millikan's teleosemantic framework and the particular proximate psychological mechanisms that she favors.

Keywords

Acceptance (compliance)/understanding | Communicative/informative intention | Conventions | Coordination | Direct/derived proper function | Etiological theory of functions | Imitative learning | Mindreading | Natural signs | Perception theory of verbal understanding

Author

[Pierre Jacob](#)
jacob@ehess.fr
Institute Jean Nicod
Paris, France

Commentator

[Marius F. Jung](#)
mjung02@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Editors

[Thomas Metzinger](#)
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

In this paper, I wish to revisit a topic that I addressed many years ago (cf. [Jacob 1997](#)) from a novel perspective. Much philosophy of mind of the latter part of the twentieth century has been devoted to naturalizing intentionality or the contents of mental representations. One of the landmarks of naturalistic philosophy of mind of the past thirty years is unquestionably Ruth Millikan's teleosemantic framework. Teleosemantic theories are teleological theories that seek to explain content by appealing to the *functions* of representations. Like most teleosemantic approaches, [Millikan \(1984, 2004\)](#) embraces an *etiological* conception of function, ac-

ording to which functions are *selected effects* ([Millikan 1984, 1989b; Neander 1991, 1995, 2004; Wright 1973](#)): the function of a trait is the effect caused by the trait that explains the continued reproduction (survival or proliferation) of past tokens of this trait.

Millikan's teleosemantic approach is particularly impressive for two related reasons. First, it applies in a single stroke to the contents of intentional *mental* representations, whose function is to mediate between pairs of cognitive mechanisms located within single brains, and also to the meanings of intentional *conventional* linguistic signs, whose function is

to mediate between pairs of cognitive mechanisms located in the brains of distinct individuals. Second, her overall teleological (or teleo-functional) approach, based on the etiological theory of functions, is meant to offer an account of the proliferation or continued reproduction of both biological entities and non-biological cultural things, such as linguistic and non-linguistic conventions.

Following [Mayr \(1961\)](#), evolutionary biologists and philosophers have long argued that the distinction between so-called *ultimate* and *proximate* explanations of biological traits (e.g., behaviors) is central to evolutionary theorizing. Roughly speaking, ultimate explanations address *why*-questions: for example, why do birds sing? Why does singing confer a selectional advantage (or greater fitness) to birds? Proximate explanations address *how*-questions: for example, what are the particular external circumstances which trigger singing in birds? What are the internal brain mechanisms that allow birds to sing?

The distinction between ultimate and proximate biological explanations raises some deep scientific and philosophical questions. One such question is whether ultimate explanations should be construed as non-causal answers to why-questions. Some philosophers have argued that ultimate explanations are *selectional* explanations based on natural selection. Natural selection can account for the prevalence of some trait in a population of individuals, but it cannot track the causal process whereby the trait is generated in each individual in the first place ([Sober 1984](#), pp. 147–152; [Dretske 1988](#), pp. 92–93; [Dretske 1990](#), pp. 827–830). Other philosophers have replied that selectional explanations are causal explanations, on the grounds that no token of a trait whose type has been selected for fulfilling its (etiological) function could proliferate unless it was linked by a causal chain to the earlier production of the selected effect by ancestor tokens of the same type of trait ([Millikan 1990](#), p. 808).¹

In this paper, I will not address such perplexing issues. I will simply accept the validity

of the distinction and assume that (whether ultimate explanations are causal explanations or not) ultimate and proximate explanations are complementary, not competing, explanations. Given that why-questions are fundamentally different from how-questions, it is likely that ultimate explanations offer few (if any) constraints on proximate explanations, and vice versa. I will further assume that the distinction carries over from biological to cultural evolution and applies to the evolution of human communication (cf. [Scott-Philipps et al. 2011](#)). In particular, Millikan's basic teleosemantic account of the proliferation of intentional conventional linguistic signs can usefully be construed as a kind of ultimate explanation of human (verbal and non-verbal) communication. Its main task is to address questions such as: what is the evolutionary or cultural function of human communication? Why do humans engage in communication at all? As with other kinds of ultimate explanations, it needs to be supplemented by specific proximate explanations whose role is to disclose the particular human cognitive capacities and mental processes whereby humans produce and understand intentional conventional signs.

The goal of this paper is to assess the balance between Millikan's broad teleosemantic approach to the cooperative function of human communication and the choice of particular proximate psychological mechanisms that she endorses. In particular, I will focus on her anti-mentalistic view, namely that verbal understanding of another's utterance is a kind of direct *perception* of whatever the utterance is about, and her correlative rejection of the basic Gricean pragmatic assumption that verbal understanding is an exercise in *mindreading*. One of the distinctive features of the human mindreading capacity is that it enables individuals to make sense of two kinds of agency: instrumental and communicative agency. In order to make sense of an agent's instrumental action, one must represent the contents of both her motivations and epistemic states. In order to make sense of an agent's communicative action, as Grice has basically argued, the addressee must infer what the agent is trying to convey, i.e., her communicative intention, whose very fulfilment

¹ For further discussion cf. [Jacob 1997](#), pp. 256–269.

requires that it is recognized by the addressee. What is distinctive of human intentional communication is that it enables the communicative agent to cause her addressee to acquire new psychological states, and thereby to manipulate his mind.

Thus, I shall examine the contrast between the particular proximate mechanisms favored by Millikan and the Gricean pragmatic tradition. In the first section, I shall spell out the basic Gricean mentalistic framework. In the second section, I will spell out Millikan's teleosemantic machinery. In the third section, I will examine Millikan's view that verbal understanding is an extended form of perception. In the fourth section, I will examine the extent to which Millikan's account of conventions can support her rejection of the Gricean assumption that verbal understanding is an exercise in mindreading. Finally, in the last section, I will show that recent developmental findings in the investigation of early human social cognition are relevant to the controversy between Millikan and the Gricean tradition over the choice of proximate mechanisms underlying human communication.

2 The Gricean mentalistic picture of communicative agency

The Gricean mentalistic tradition rests on three basic related assumptions.²

- The first is the assumption that the complete process whereby an addressee contributes to the full success of a speaker's communicative act should be decomposed into two separable psychological sub-processes: a process of *understanding* (or *comprehension*) of the speaker's utterance and a process of *acceptance*, which in turn can be construed as the addressee's acquiring either a new belief or a new desire for action (depending on the direction of fit of the speaker's utterance). I'll call this the *separability* thesis.

² Although the relevance-based approach advocated by Sperber & Wilson (1986) and Wilson & Sperber (2004) departs in some interesting respects from Grice's (1969, 1989) own approach, I will nonetheless call their approach "Gricean" because, in the context of the present paper, the continuities between the two frameworks are far more important than the discontinuities.

- The Gricean mentalistic tradition also rests on the assumption that verbal understanding is an exercise in mindreading, whereby the addressee recognizes the complex psychological state that underlies the speaker's communicative act. I'll call this the *mindreading* thesis. (Clearly, the mindreading thesis is presupposed or entailed by the separability thesis.)
- Third, the Gricean mentalistic tradition further rests on a fundamental hypothesized asymmetry between what is required for understanding *instrumental* non-communicative agency and *communicative* agency. An agent intends her instrumental action to satisfy her desire in light of her belief, and the desirable outcome of her instrumental action can be recognized even if the agent fails to fulfil her goal or intention. But the intended effect of a speaker's communicative action, which is the addressee's understanding of what she means, cannot be achieved unless the speaker's intention to achieve this effect is recognized (cf. Sperber 2000, p. 130). Unlike purely instrumental agency, communicative agency is *ostensive* in the following sense. A speaker's communicative act is ostensive because its desirable outcome cannot be identified unless the addressee recognizes what the speaker intends to make manifest to him, i.e., what Sperber & Wilson (1986) call the speaker's *informative* intention. Thus, the Gricean tradition rests on the thesis of the *ostensive* nature of communicative agency (Sperber & Wilson 1986).

2.1 The mindreading thesis

On the picture of pragmatics which is part of the Gricean tradition of the past forty years broadly conceived, a human agent could not achieve a verbal or non-verbal act of intentional communication unless she had a complex psychological state, which Grice (1957) called the "speaker's meaning" and which he construed as a set of three interrelated intentions.³ First of all, by producing an utterance (or any other piece of ostens-

³ For brevity, I'll use "speaker" instead of "communicative agent". But of course not all communicative actions are verbal.

ive communicative behavior), the speaker must have the basic intention (i) to act on her addressee's mind, i.e., to cause him to acquire a new belief or a new desire (or intention) to perform some action. Second, the speaker must intend (ii) her addressee to recognize the content of her basic intention. Third, she must further intend (iii) her addressee's recognition of her basic intention (in accordance with (ii)) to play a major role in his fulfilling her basic intention (i).

In the following, I will adopt (Sperber & Wilson's (1986) simplified two-tiered account, according to which a communicative agent who produces an utterance has *two* (not three) inter-related intentions: an *informative* and a *communicative* intention, the first of which is nested within the other. She has the informative intention to make some state of affairs manifest to her addressee and also the further communicative intention to make her informative intention manifest to her addressee. So in this framework, the speaker's communicative intention is fulfilled by the addressee as soon as the latter recognizes (or understands) which state of affairs it is the speaker's informative intention to make manifest. But more is required for the speaker's informative intention to be fulfilled: the addressee must further accept the speaker's epistemic or practical authority. Depending on the direction of fit of the speaker's utterance, the addressee must either believe the fact which it is the speaker's informative intention to make manifest to him, or he must acquire the desire to act so as to turn into a fact the possible state of affairs which it is the speaker's informative intention to make manifest to him.

In a nutshell, much of (Sperber & Wilson's (1986) relevance-based framework rests on their insightful recognition that, on the broad Gricean picture of the speaker's meaning, the task of the addressee can be usefully divided into two basic psychological processes: one is the process whereby the addressee *understands* (or recognizes) the speaker's informative intention and the other is the process whereby he *fulfils* the speaker's informative intention. The first process involves the addressee's recognition of the speaker's informative intention, whereby the addressee fulfils the speaker's communicat-

ive intention that he recognize the speaker's informative intention. By recognizing the speaker's informative intention, the addressee comes automatically to both fulfil the speaker's communicative intention and to understand (or comprehend) the speaker's utterance. But for the addressee to *recognize* the speaker's informative intention is not *ipso facto* to *fulfil* it. So the second process needed for the success of the speaker's communicative act involves the addressee's fulfilment of the speaker's informative intention, whereby the addressee either accepts a new belief (in accordance with the content of the speaker's assertion) or forms a new desire to act (in accordance with the content of the speaker's request; cf. Jacob 2011).

2.2 The separability thesis

While the relevance-based account of communication clearly presupposes the mindreading thesis, Sperber (2001) has offered further support in favor of the separability thesis. Following Krebs & Dawkins (1984), Sperber (2001) has argued that for cooperative communication to stabilize in human evolution, it must be advantageous to both senders and receivers. Since the interests of speakers and hearers are not identical, the cooperation required for the stabilization of communication is vulnerable to deception. When her utterance is descriptive, the speaker can speak either *truthfully* or *untruthfully*. The addressee can either *trust* the speaker or not. The speaker is better off if her addressee trusts her and worse off if he distrusts her, whether or not the speaker is truthful. If the addressee trusts the speaker, then he is better off if the speaker is truthful and worse off if the speaker is not truthful, while the addressee remains unaffected if he distrusts the speaker.

Clearly, not every speaker is (or should be) granted equal epistemic or practical authority on any topic by every addressee. As Sperber et al. (2010) have further argued, given the risks of deception, it is likely that human cooperative communication would not have stabilized in human evolution unless humans had evolved mechanisms of *epistemic vigilance*, whereby they filter the reliability of descriptive utter-

ances. Focusing on a speaker's assertions at the expense of her requests, a speaker's epistemic authority depends to a large extent on the addressee's evaluation of her reliability (or trustworthiness), which in turn depends jointly on the addressee's evaluation of the speaker's competence on the topic at hand and on the addressee's representation of how benevolent are the speaker's intentions towards him. According to Sperber et al. (2010), an addressee's epistemic vigilance can apply to either or both the *source* of the information being communicated and its *content*.

3 Millikan's teleosemantic machinery

3.1 Teleosemantics and informational semantics

One of the first attempts at a naturalistic account of content (or intentionality) in the philosophy of mind was Dretske's (1981) informational semantics, according to which a sign or signal s carries information about property F iff there is a nomic (or lawful) covariation between instances of F and tokenings of s . As Millikan (1984, 2004) emphasized shortly after, informational semantics faces the puzzle of accounting for the possibility of *misrepresentation*. If the conditional probability that F is instantiated given s is 1, then how could s ever misrepresent instances of F ? This puzzle is neatly solved by teleological approaches: if a representation has a *function*, then it can fail to fulfil its function and thereby misrepresent what it is designed to represent (Millikan 2004, Ch. 5). According to Dretske's (1988, 1995) own later attempt at preserving informational semantics as part of teleosemantics, a sign or signal s could not represent some property F unless s had the function of carrying information about (or indicating) instances of F .

Millikan's (1984, 2004) teleosemantic approach sharply departs from Dretske's (1988, 1995) information-based framework in at least two fundamental respects. First of all, in Millikan's earliest (1984) teleosemantic framework, there was no room for information-theoretic notions at all. In her later (1989a, 2004) work, she

argued that carrying information could not be a teleological function of a sign on the following grounds. Whether a sign carries information about some property depends on how the sign was *caused* or *produced*. But according to the etiological theory of functions, the function of a sign is one of its own *effects*, i.e., the selected effect that explains the continued reproduction of tokens of signs of this kind. How a sign was caused cannot be one of its effects, let alone its selected effect. If and when a sign happens to carry information about something, carrying information cannot be its selected effect, i.e., its etiological function.⁴

Second, Dretske's (1981) informational semantics could only be suitably naturalistic in the required sense if information is construed as the converse of nomological covariation (or necessity), i.e., as an entirely non-intentional and/or non-epistemic commodity. But as Millikan (2004, pp. 32–34) argues, if signal s could not carry information about F unless the probability that F is instantiated when s is tokened were 1 (in accordance with some natural law), then no animal could ever learn about F from perceiving tokens of s .

In her 2004 book, Millikan elaborates a notion of natural sign that is more “user-friendly” precisely because “it is at root an epistemic notion” (Millikan 2004, p. 37). On Millikan's (2004) account, natural signs (e.g., tracks made by quail) are locally recurrent signs within highly restricted spatial and temporal domains: relative to one local domain, such tracks are natural signs of quail. Relative to a neighboring domain, the very same tracks are made by pheasants and are therefore natural signs of pheasants, not quail. Locally recurrent signs afford knowledge of the world for animals who can learn how to track the circumscribed domains relative to which they carry reliable information. Furthermore, locally recurrent natural signs can form transitive chains (or be productively embedded) within circumscribed do-

⁴ For significant discussion and defense of the view that it is the etiological function of mental representations to carry information, in response to Millikan's criticisms, cf. Neander (1995, 2007), Godfrey-Smith (2006) and Shea (2007). Cf. the recent exchange between Neander (2011) and Millikan (2011) For a criticism of Millikan's view, cf. Pietroski (1992) and see Millikan's (2000) reply.

mains. For example, retinal patterns can be a natural sign of tracks in the ground, which in turn may be a natural sign of quail within a circumscribed local domain. Perception is what enables non-human animals and humans alike to track the meanings of locally recurrent natural signs in their circumscribed domain of validity and thereby to acquire knowledge of the world (Millikan 2004, Ch. 4).

The application of Millikan's (1984, 2004) teleosemantic framework to the meanings of intentional *conventional* signs results from the combination of three related ingredients: (i) the etiological view of functions; (ii) acceptance of the sender-receiver structure as a necessary condition on the contents of intentional representations; and (iii) a naturalistic account of the reproduction of conventions.

3.2 The etiological conception of functions

As I said above, on the etiological view, the function of some trait is its selected effect that explains the continued reproduction of past tokens of this trait. This is what Millikan (1984) calls a device's *direct* proper function. But a device may also have what she calls a *derived* proper function. For example, it is the direct proper function of the mechanism of color change in the skin of chameleons to make them undetectable from the local background by predators. It is a derived proper function of this mechanism in a particular chameleon, Sam, at a particular place and time, to make the color of its skin match the color of its particular local background at that time so as to make it undetectable by predators there and then.

While Millikan's teleofunctional framework based on the etiological approach to functions primarily fits biological traits, it applies equally to non-biological items such as non-bodily tools—including public-language forms. For example, screwdrivers have the direct proper function of turning (driving or removing) screws. This is the effect of screwdrivers that explains their continued reproduction. Clearly, a screwdriver may also be intentionally used for the purpose of driving a screw with a particular metallic

structure, length, and diameter into a particular wooden material at a particular time and place. If so, then driving this particular screw into this particular wooden material at a particular time and place will be the derived proper function that this particular screwdriver inherits from the agent's intention.

3.3 The sender-receiver framework

According to the sender-receiver framework, a sign or signal *R* can be an *intentional* representation (as opposed to a *natural* sign) only if it is a *relatum* in a three-place relation involving two systems (or mechanisms), one of which is the sender (who produces *R*), the other of which is the receiver (who uses *R*). By application of the etiological view of functions, the sender (or producer) and the receiver (or consumer) have co-evolved so that what Millikan (1984, 2004) calls the *Normal* conditions for the performance of the function of one depends on the performance of the other's function and vice versa. In a nutshell, the producer and the consumer are *co-operative* devices, whose interests overlap and whose activities are beneficial to both. Thus, the cooperative ternary sender-receiver structure naturally applies to the contents of intentional mental representations that mediate between cognitive mechanisms located within a single organism.⁵

In virtue of the fact that intentional mental representations can have two basic directions of fit, the evolved cooperation between the producer and the consumer can take two basic forms. If and when the representation is *descriptive* or has a mind-to-world direction of fit, the producer's function is to make a sign *R*, whose content matches some state of affairs *S*, for the purpose of enabling (or helping) the consumer to perform its own task when and only when *S* obtains. If and when the representation is *directive* (or prescriptive) or has a world-to-mind direction of fit, the producer's function is to produce a representation whose content will guide the consumer's action, and it is the con-

⁵ Cf. Godfrey-Smith (2013) and Artiga (forthcoming) for further elaborate discussion of the requirement of cooperation as a condition on application of the sender-receiver structure.

sumer's function to make the world match the content of the sign by its own activities. Furthermore, Millikan (1995, 2004) has argued that the most primitive kinds of intentional mental representations (shared by humans and non-human animals) are what she calls *pushmi-pullyu* representations, which are at once descriptive and prescriptive, with both a mind-to-world and a world-to-mind direction of fit.

3.4 Conventional patterns

The third component of Millikan's teleosemantic approach to the meanings of intentional conventional signs involves her (1998) naturalistic account of *conventions*. On her account, so-called *natural conventionality* rests on two elementary characteristics: first, natural conventions are patterns that are *reproduced* (or that proliferate). Second, they are reproduced (or "handed down") "owing to precedent determined by historical accident, rather than owing to properties that make them more intrinsically serviceable than other forms would have been" (Millikan 2005, p. 188). The fact that conventions rest on historical precedent to a large extent accounts for their arbitrariness.⁶ On the basis of her naturalistic account of the continued reproduction of natural conventions, Millikan further offers a purportedly naturalistic account of the continued reproduction of conventional public-language signs, whose function is to coordinate the transfer of information between speakers and hearers. She thereby extends the cooperative ternary sender–receiver structure to the meanings of intentional conventional signs (or public-language forms) that mediate cognitive mechanisms located within pairs of distinct organisms.

Conventional public-language forms are tools or *memes* in Dawkins's (1976) sense: they have been selected and have accordingly been reproduced because they serve *coordinating* functions between a sender (the speaker) and a receiver (the addressee), whose interests overlap. But like any other tool, in addition to its direct *memetic* (or "stabilizing") function (which ex-

plains its continued reproduction), a particular token of some public-language form may also have a derived function or purpose, derived from the purpose of the speaker who produced it at a particular place and time. Thus, a token of a public-language form has two kinds of purposes: a memetic purpose and the speaker's purpose, which may or not coincide (cf. Millikan 1984, 2004, 2005).

4 Is verbal understanding an extended form of perception?

4.1 Perceiving the world through language

One basic problem raised by Millikan's account of the proliferation of intentional conventional signs is that one and the same linguistic form detached from its context of use may belong to different *memetic* families (or chains of reproductive events). In the reproductive process, what gets copied from one pair of sender-receivers to the next is not merely a linguistic form (e.g., "clear"), but the *use* of a linguistic form embedded in a particular *context*. This is why on Millikan's (2005, Ch. 10, section 3) view, the boundary between semantics and pragmatics is *blurry* and the process whereby a hearer tracks the memetic lineage of a conventional sign is a pragmatic process. On the teleosemantic approach, the hearer's task is to retrieve the appropriate context necessary for recognizing the correct memetic family (or lineage) to which a particular conventional sign belongs. In a nutshell, the hearers' task is to track the *domains* of intentional *conventional* signs.

Thus, it would appear that the hearer's task is quite similar to what is involved in tracking the restricted domain over which the information carried by a locally recurrent *natural* sign (e.g., tracks made either by quail or by pheasants) is valid. Since tracking the local domains over which the information carried by locally recurrent natural signs is a *perceptual* task, it is not surprising that Millikan has persistently urged that "in the most usual cases understanding speech is a form of direct perception of whatever speech is *about*. Interpreting

⁶ Including the arbitrariness of the relation between particular word-types and what they mean (sense and/or reference).

speech does not require making any inferences or having any beliefs about words, let alone about speaker intentions” (Millikan 1984, p. 62).⁷ Millikan (2004, p. 122) nicely illustrates her view that verbal understanding is an extended form of perception:

rain does not sound the same when heard falling on the roof, on earth, on snow, and on the water, even though it may be directly perceived as rain through any of these media. Exactly similarly, rain has a different sound when the medium of transmission is the English language (“It’s raining!”). And it sounds different again when the medium of transmission is French or German.

In a nutshell, “during Normal conversation, it is not language that is most directly perceived by the hearer but rather the world that is most directly perceived *through* language” (Millikan 2005, p. 207).

Furthermore, both ordinary and extended perception rest on *translation*, not inference: “the first steps in perception involve reacting to natural signs of features of the outer world by translating them into inner intentional representations of these outer features, for example, of edges, lines, angles of light sources in relation to the eye” (Millikan 2004, p. 118). In *normal* verbal communication, translation plays a two-fold role in mediating transfer from the speaker’s belief to the addressee’s belief. First, the speaker of a descriptive utterance translates her belief into a sentential conventional sign. Secondly, the addressee translates the content of the speaker’s utterance into his own new belief (Millikan 1984, 2004, 2005).

4.2 Ordinary and extended perception

Clearly, Millikan’s thesis that verbal understanding is an extended form of perception is not consistent with the Gricean thesis that verbal understanding is an exercise in *mindreading*. But on the face of it, the thesis

that verbal understanding is an extended form of perception (of whatever speech is *about*) itself is puzzling for at least three related reasons.⁸ First of all, as Millikan (2004, Ch. 9) herself recognizes, there is a major difference between the content of a perceptual representation of some state affairs and the verbal understanding of the content of another’s testimony about the very same state of affairs. At an appropriate distance and in good lighting conditions, one could not perceive a cup resting on a table without also perceiving its shape, size, color, texture, content, orientation, and spatial location with respect to the table, to any other object resting on the table, and especially to oneself. As Millikan (2004, p. 122) recognizes, unlike the content of testimony, the content of ordinary perception can be put at the service of action precisely because it provides information about the agent’s spatial relation to an object that is potentially relevant for action. But if an addressee located in a room next to the speaker’s room understands the content of the latter’s utterance of the sentence “There is a cup on the table”, he may endorse the belief that there is a cup on the table without having any definite expectation about the shape, size, color, texture, content, orientation, and spatial location of the cup with respect to himself, the table, or anything else.

Second, the thesis that verbal understanding is an extended form of perception ought to be restricted to the hearer’s verbal understanding of the meanings of *descriptive* utterances of indicative sentences with a mind-to-world direction-of-fit, which describe *facts* (or *actual* states of affairs). It cannot without further modifications be directly applied to the hearer’s verbal understanding of the meanings of *prescriptive* utterances of *imperative* sentences whereby a speaker *requests* an addressee to *act* so as to turn a *possible* (non-actual) state of affairs into a fact (or an actual state of affairs). Prescriptive utterances, which have a world-to-mind direction of fit, fail to describe any fact that could be directly perceived at all. So the question arises whether Millikan would be willing to en-

⁷ Cf. Millikan (2000, Ch. 6), Millikan (2004, Ch. 9), Millikan (2005, Ch. 10).

⁸ Cf. Recanati (2002) for a defense of Millikan’s thesis.

dorse the revised two-tiered thesis that (i) a verbal understanding of a speaker's descriptive utterance is the perception of whatever the utterance is about and (ii) a verbal understanding of a speaker's prescriptive utterance is to intend to perform whatever action is most likely to comply with the speaker's request.

Finally, testimony enables a speaker to convey beliefs whose contents far outstrip the perceptual capacities of either the speaker or her addressee. For example, an addressee may understand that the speaker intends to verbally convey to him her belief that there is no greatest integer, that democracy is the worst form of government except all those other forms that have been tried from time to time, or that religion is the opium of the people. But it does not make much sense to assume that either the speaker or her addressee could perceive what the speaker's utterance is about.

4.3 Tracking the domains of intentional conventional signs

Furthermore, the thesis that verbal understanding is an extended form of perception clearly rests on the assumption that the process whereby the hearer of a speaker's utterance tracks the memetic family of the intentional conventional sign used by the speaker is basically the same as the process whereby human and non-human animals track the meanings of locally recurrent natural signs in their circumscribed domain of validity. As I mentioned above, Millikan (2004) argues that perception is the basic process whereby animals track the meanings of locally recurrent natural signs in their circumscribed domain of validity. Crucially, one can track the meanings of locally recurrent natural signs within their circumscribed domain of validity *without* representing an agent's psychological state. So the question arises whether a hearer of a speaker's utterance could *always* track the memetic family of the intentional conventional signs used by a speaker *without* representing any of the speaker's psychological states.

In particular, as Recanati (2007) has argued, the question arises for descriptive utter-

ances containing at least four kinds of conventional expressions considered by Millikan (2004, Chs. 10–12): so-called unarticulated constituents in Perry's (1986) sense, incomplete definite descriptions, quantifiers, and possessives. Consider first an utterance of (1):

(1) It is raining.

It is unlikely that by an utterance of (1) a speaker means to assert that it is raining somewhere or other at the time of utterance. Instead, she is likely to mean that it is raining at the time of utterance and at the place of utterance (which remains unarticulated in the sentence). If by an utterance of (1), the speaker could *only* mean that it is raining at the place of utterance, then Millikan's claim that a hearer need not represent any of the speaker's psychological states for the purpose of tracking the local domains of intentional conventional signs might be vindicated. However, by an utterance of (1) on the phone, a speaker located in Paris may mean that it is raining in Chicago, not in Paris. Similarly, a French speaker located in Paris may use the incomplete description "the President" to refer, not to the French President, but instead to the President of the US.

For the purpose of understanding an utterance of a sentence containing a universal quantifier, as shown by example (2), the hearer must be able to properly restrict the domain of the quantifier:

(2) Everyone is asleep.

By an utterance of (2), the speaker presumably means to assert, not that everyone in the universe is asleep, but that everyone in some restricted domain (e.g., a relevant household) is asleep.⁹ The relevant restricted domain is the domain the speaker has in mind. Finally, by using the possessive construction "John's book", the speaker may have in mind many different relations between John and the book: she may mean the book written by John, the book

⁹ A nice example suggested by a referee is "There is no beer left", where the audience does not take the speaker to mean that there is no beer left in the universe, but instead in some properly restricted domain (e.g., some relevant fridge).

read by John, the book bought by John, the book sold by John, the book John likes, the book John dislikes, the book John just referred to in the conversation, the book John lost, the book John gave to the speaker, the book the speaker gave to John, the book the hearer gave to John, and so on. Unless the hearer hypothesizes what relation the speaker has in mind, he will fail to understand what the speaker means by her utterance of “John’s book”. In none of these four cases does it seem as if the hearer could recognize the memetic family of intentional *conventional* signs, i.e., track their relevant domains—unless he could represent the contents of some of the speaker’s *beliefs* or *assumptions*.

5 Conventions and belief-desire psychology

5.1 Teleosemantics and the separability thesis

Millikan’s thesis that verbal understanding is an extended form of perception is meant as an alternative to the Gricean thesis that verbal understanding is an exercise in mindreading. The further question arises to what extent Millikan’s teleosemantic account of the proliferation of public language conventions is consistent with the Gricean separability thesis, i.e., the distinction between verbal understanding and either acceptance (belief) or compliance. I will first argue that there is a restricted sense in which Millikan’s teleosemantics seems to be consistent with the separability thesis. But I will further argue that in a broader sense Millikan’s rejection of the mindreading thesis undermines the separability thesis.

On Millikan’s teleosemantic account, for a speaker’s descriptive utterance of an indicative sentence to meet the requirement of cooperation (and mutual interest) between the sender (or producer) and the receiver (or consumer), its direct proper function must be to cause the addressee to form a (true) belief. For a speaker’s *prescriptive* utterance of an imperative sentence to meet the requirement of cooperation, its direct proper function must be to cause the ad-

dresser to act in compliance with the content of the speaker’s request.

In the terminology of the relevance-based framework, a speaker who utters a descriptive utterance makes manifest to her addressee her communicative intention to make manifest her informative intention to make some fact manifest to him. The addressee may fulfil the speaker’s communicative intention by recognizing her informative intention and yet fail to fulfil her informative intention by resisting endorsing the relevant belief. A speaker who utters a prescriptive utterance makes manifest to her addressee her communicative intention to make manifest her informative intention to make manifest to him the desirability of turning some possible state of affairs into a fact by his own action. The addressee may fulfil the speaker’s communicative intention by recognizing her informative intention and yet fail to fulfil the speaker’s informative intention by resisting endorsing the intention to act in accordance with the speaker’s request.

Origi & Sperber (2000, pp. 160–161), who subscribe to the Gricean thesis of the separability between verbal understanding and acceptance or compliance, have argued that the direct proper function of either a descriptive utterance or a prescriptive utterance could not be to reliably elicit the addressee’s response “at the level of belief or desire formation” (i.e., “the cognitive outputs of comprehension”), but instead “at an intermediate level in the process of comprehension”. Millikan might reply that according to her teleosemantic framework, an utterance may have a direct proper *function* and yet remain *unfulfilled*. If so, then the fact that an addressee may fulfil the speaker’s communicative intention (by recognizing her informative intention) and yet fail to fulfil the speaker’s informative intention seems entirely compatible with the teleosemantic framework.

However, to the extent that Millikan explicitly rejects the mindreading thesis, which is presupposed by the separability thesis, it is unlikely that she would find the separability thesis itself acceptable. On the relevance-based approach, it is a sufficient condition for securing what Austin (1975) called the “up-

take” (or success) of a communicative act (or speech act) that the speaker causes the addressee to fulfil the speaker’s communicative intention by recognizing her informative intention. It is not necessary that the addressee further fulfil the speaker’s informative intention. Successful communication does not require the addressee to accept either a new belief or a new desire, in accordance with the speaker’s informative intention. But on Millikan’s teleosemantic framework, failure of the addressee to comply with the speaker’s goal of causing the addressee to accept either a new belief or a new desire looks like a failure of the addressee to cooperate with the speaker’s conventional action, and therefore like a breakdown of the speaker’s communicative action. In fact, Millikan (2000, 2004, 2005) has offered two broad grounds for rejecting the mindreading thesis, both of which make it unlikely that she would support the separability thesis; the second of which is based on developmental evidence. I start with the non-developmental argument.

5.2 Cooperation and social conformity

First, Millikan (2004) rejects the mindreading thesis as part of her criticism of the reasoning that leads to the separability thesis: she rejects the joint assumptions that human predictions of others’ behavior are based on mindreading and that cooperation in human verbal communication is vulnerable to the risks of deception. On the one hand, she argues that “most aspects of social living involve cooperation in ways that benefit to everyone [...] for the most part, social cooperation benefits both or all parties. There is nothing mysterious about its evolution in this respect” (Millikan 2004, pp. 21–22). In a nutshell, Millikan argues that the urge to explain how the benefits of human communication are not offset by the risks of deception is misplaced on the grounds that the interests of speakers and hearers are sufficiently similar, if not identical.¹⁰

¹⁰ As Godfrey-Smith (2013, p. 45) observes, sameness of interests in human cooperation can be safely assumed in small contemporary communities, but not on a large scale, and nor in an evolutionary context.

On the other hand, she argues that we use belief-desire psychology, not for prediction, but “for explanation after the fact” (Millikan 2004, p. 22). This is consonant with her (1984, pp. 67–69) earlier claim that while human adults have the ability to *reflect* on a speaker’s communicative intention if the automatic flow of conversation is interrupted for one reason or another, *normal* verbal understanding does *not* require representing the speaker’s communicative intention. Instead, *normal* verbal understanding should be construed as a *conventional* transfer of information whereby the speaker *translates* her belief into an utterance, whose meaning is in turn *translated* back by the addressee into a newly acquired belief.

Thus, Millikan rejects two of the major assumptions on which the separability thesis rests. She underestimates the gap between the interests of speakers and hearers in human communication and she minimizes the role of belief-desire psychology in the prediction of others’ behavior. Interestingly, her rejection of both assumptions rests in turn on her own competing account of communicative acts. As she puts it, “a surprise of this analysis of the conventional nature of the information-transferring function of the indicative is that believing what you hear said in the indicative turns out to be a conventional act, something one does in accordance with convention” (Millikan 2005, p. 46).¹¹ First of all, Millikan (2004, p. 23) argues that humans expect others to behave in conformity with social conventions, not on the basis of others’ beliefs and desires. Second, she further speculates that the conventional behaviors that are caused by a disposition to social conformity may derive from natural selection the memetic function of serving a coordinating function (*ibid.*).

Clearly, being disposed to social conformity and expecting others to be similarly disposed may help solve coordination problems (as shown by driving on one side of the road). However, being disposed towards social conformity is not sufficient to comply with social conventions. Compliance requires *learning*, i.e., the ac-

¹¹ Note that this quote seems to presuppose the negation of the separability thesis.

quisition of relevant true *beliefs* about the contents of social conventions. Thus, the basic challenge for Millikan's claim that humans expect others to behave, not so much in accordance with the contents of their beliefs and desires, but in conformity with social conventions, is to offer an account of how humans come to learn and thereby know what it takes to act in conformity with social conventions.

5.3 Counterpart reproduction

This issue has been highlighted by the exchange between Tomasello's (2006) comments on Millikan's (2005) book and Millikan's (2006) response, which focuses on Millikan's (1998) thesis that many conventions, whose function it is to solve *coordination* problems, are reproduced by what she calls *counterpart* reproduction (or nuts and bolt reproduction). Typical coordination problems involve at least two partners, who share a common purpose that can be achieved only if each partner plays its assigned role, where both partners can be required to perform either the same act or two distinct complementary acts. In counterpart reproduction, when the respective roles of each partner require them to perform two different complementary acts, one typically adjusts her behavior to the other's and vice-versa. Counterpart reproduction is exemplified by, e.g., handshake reproduction, the reproduction of the respective postures assumed by men and women in traditional dancing, the reproduction of social distances appropriate for conversation, or the reproduction of the use of chopsticks for eating. Similarly, Millikan (2005, 2006) argues that counterpart reproduction also underlies the continued reproduction of conventional public-language signs.

Millikan (2005) further mentions open, partially or completely blind, conventional *leader-follower* co-ordinations involved in joint actions based on shared goals, whereby one agent (the leader) introduces a component of a pattern whose completion requires her partner (the follower) to perform a complementary component (*ibid.*, pp. 12–14). One example of open conventional leader-follower coordination is the

pattern whereby one agent selects her seat at an arbitrary table in a restaurant and her partner follows suit and selects his accordingly. One example of a partially blind conventional leader-follower coordination is the couch-moving pattern whereby the leader affords the follower anticipatory cues of her next move by ostensibly exaggerating her own movements, where the follower's familiarity with the pattern enables him to recognize the leader's ostensive cues and thereby to reproduce the complementary portion of the joint action. Another of Millikan's examples of a partially blind conventional leader-follower coordination is the US mailbox-flag convention, whereby the leader puts up a flag after she has placed mail in the mailbox and the postman picks up the mail after perceiving the flag.

Much comparative work by Tomasello and colleagues (reported by Tomasello et al. 2005 and summarized by Tomasello 2006, 2008) shows that while most communicative gestures in chimpanzees are learnt by ontogenetic ritualization, most communicative behaviors in human infants are acquired by imitative learning. As Tomasello (2006) argues, Millikan's own requirement that the reproduction of a conventional pattern depends on "the weight of precedent", not on its perceived intrinsic superior ability to produce a desired result, seems better fulfilled by a process of imitative learning than by a process of trial and error whereby one individual adjusts her behavior to another's. There seems to be nothing arbitrary (as there should if it were conventional) about an individual's adjusting her behavior to another's. While Tomasello (2006) does not deny that counterpart reproduction plays a significant role in cultural transmission, he disputes the claim that the output of counterpart reproduction qualifies as conventional.

Part of the gap between Millikan and Tomasello lies in what they take to be the proper unit for the analysis of the mechanism underlying the continued reproduction of conventional patterns involved in solving problems of coordination. While Tomasello focuses on the learning capacities of single individual minds, Millikan focuses on what can be achieved by the

reciprocal adjustments of pairs of cooperative partners. For example, when Millikan (2005, 2006) argues that counterpart reproduction underlies the continued proliferation of the custom of using chopsticks for eating in some cultures, she construes the convention of using chopsticks as a solution to the problem of coordination between pairs of partners, some of whom buy chopsticks and use them for eating and others who manufacture chopsticks. The latter would not manufacture chopsticks unless the former bought them and used them for eating. Conversely, the former would not buy them and use them for eating unless the latter manufactured them.

But of course, as Millikan is aware, this leaves open the question of how young children learn to use chopsticks for eating. As Millikan (2006, pp. 45–46) rightly observes, young human children understand their native language long before they can speak it. Nor can they learn to understand by imitating mature speakers: as she puts it, “they don’t watch how other people understand and then copy”. She further argues that young children would never understand their native tongue unless “their teachers” spoke to them, but “their teachers” would never speak to young children unless “they had had some reasonably successful experience” with previous listeners. This makes the continued reproduction of conventional public-language signs fit the pattern of counterpart reproduction. But still the question arises: how do young children learn to produce words of their native tongue? Vocal imitative learning may well play an important role (cf. Hauser et al. 2002). In a nutshell, according to Millikan the function of conventions is to solve coordination problems. She offers an elegant account of the proliferation of conventions based on counterpart reproduction. Her account must make room for the role of imitative learning in the way young human children learn either to use chopsticks for eating or to produce (and not just understand) words of their native tongues. As I shall argue in section 6.2, evidence shows that imitative learning in young children rests on their ability to construe the model’s demonstration as an ostensive communicative action. If so, then Mil-

likan’s view that counterpart reproduction underlies the proliferation of conventions must make room for the role of children’s ability to recognize the model’s communicative intention.

6 Teleosemantics and the puzzles of early human social cognition

6.1 Millikan’s developmental puzzle

To further undermine the mindreading thesis, Millikan (1984, 2000, 2004, 2005) has also appealed to findings from the developmental psychological investigation of early human social cognition, showing that “children younger than about four, although fairly proficient in the use of language, don’t yet have concepts of such things as beliefs, desires, and intentions” (Millikan 2005, p. 204). If such children do not have such concepts, then, unlike adults, they cannot reflectively engage in tasks of mindreading, i.e., in tracking the contents of others’ intentions, beliefs, and desires. To the extent that they can engage in verbal understanding, this further shows that verbal understanding cannot rest on mindreading (or belief–desire psychology).

As Millikan emphasizes, much developmental evidence shows that before they are at least four years old the majority of human children systematically *fail* elicited-response false-belief tasks. (In the terminology of developmental psychologists Baillargeon et al. 2010, *elicited-response* tasks are tasks in which a participant is requested to generate an explicit answer in response to an explicit question.) For example, in the Sally-Anne test, after Sally places her toy in the basket, she leaves. While Sally is away, Anne moves Sally’s toy from the basket to the box. When Sally returns, participants, who know the toy’s actual location, are explicitly asked to predict where Sally (who falsely believes her toy to be in the basket) will look for her toy. The evidence shows that the majority of three-year-olds, “although quite proficient in the use of language” (in Millikan’s terms, Millikan 2005, p. 204), typically point to the box (i.e., the toy’s actual location), not to the basket where the agent falsely believes her toy to be (cf. Wimmer & Perner 1983, Baron-

Cohen et al. 1985 and Wellman et al. 2001 for a meta-analysis).

Millikan assumes that the failure of most three-year-olds in such elicited-response false-belief tasks demonstrates that they lack what she calls a “representational theory of mind”. In a nutshell, she assumes that success at elicited-response false-belief tasks is a necessary condition for crediting an individual with a representational theory of mind (i.e., the ability to track the contents of others’ false beliefs). Acceptance of this assumption gives rise to Millikan’s developmental puzzle, which is “to understand how very young children can be aware of the intentions and of the focus of attention of those from whom they learn language without yet having this sort of sophisticated theory of mind” (Millikan 2005, p. 205). Before explaining why Millikan’s assumption is contentious, I shall briefly examine Millikan’s solution to her own puzzle.

Millikan’s solution involves three related ingredients, the most important of which is her thesis that normal verbal understanding is an extended form of perception (which does not require thinking about a speaker’s intention at all). Second, she argues that young children can understand the goal-directedness of a speaker’s communicative action without tracking the content of her communicative intention. Third, she argues that young children can understand the referential focus of a speaker’s attention without having a sophisticated theory of mind. As I understand it, much of the argument for the possibility of understanding the referential focus of a speaker’s attention without having a sophisticated theory of mind rests on the thesis that verbal understanding is an extended form of perception. As I have already expressed doubts about the thesis that verbal understanding is an extended form of perception, I shall now briefly examine the second thesis: that young children could understand the goal-directedness of a speaker’s communicative without tracking the content of her communicative intention.

Millikan (2005, pp. 206–207) offers two main reasons for granting young children the ability to recognize the goal-directedness of a speaker’s communicative action without grant-

ing them a full representational theory of mind. First, she argues that the evidence shows that mammals (dogs and cats and non-human primates, presumably, as well) lack a representational theory of mind but have the ability to recognize the goal-directedness of each other’s behavior. So by parity, very young children should also be granted the ability to recognize the goal-directedness of others’ actions, including speakers’ communicative actions. Second, she argues that communicative actions are cooperative actions. When young children are engaged in some cooperative action (including a communicative action) with a caretaker, they can easily keep track of the shared goal of the cooperative action, while tracking the focus of the speaker’s visual attention, without having a full representational theory of mind.

On the one hand, there is evidence that non-human primates recognize the goals of conspecifics engaged in the execution of *instrumental* actions (Call & Tomasello 2008). On the other hand, there is also evidence that non-human primates—and birds as well—can discriminate *knowledgeable* agents (who know about, e.g., food from visual perception) from *ignorant* agents (who don’t know about food because their line of vision is obstructed) in *competitive* situations (Bugnyar 2011; Call & Tomasello 2008; Dally et al. 2006; Hare et al. 2001; Tomasello et al. 2003). But the question raised by Millikan’s puzzle is to understand what enables very young human children to make sense jointly of a speaker’s goal and the focus of her visual attention, when the speaker is performing a *communicative* action, not an *instrumental* action, in a *cooperative*, not a *competitive*, context. The fact that non-human primates can represent the goal of an agent’s *instrumental* action and discriminate a knowledgeable from an ignorant agent in a *competitive* context falls short of providing the required explanation.

Furthermore, two of Millikan’s assumptions are contentious in light of recent findings from developmental psychology. One is her assumption that young children could recognize the goal-directedness of speakers’ communicative actions without a representational theory of mind. The other is her assumption that success

at elicited-response false-belief tasks should be taken as a criterion for having the ability to track the contents of others' false beliefs (and therefore having a representational theory of mind). I shall start with the former, which amounts to denying the asymmetry between instrumental and communicative agency—which I earlier dubbed the thesis of the *ostensive* nature of human communicative agency.

6.2 The puzzle of imitative learning

The first relevant developmental finding, reported by [Gergely et al. \(2002\)](#), shows that approximately one-year-old human children (fourteen-month-olds) selectively imitate an agent's odd action. First, infants were provided with ostensive cues whereby an agent made manifest her intention to convey some valuable information by looking into the infants' eyes and addressing them in motherese. She then told the infants that she felt cold and covered her shoulders with a blanket. She finally performed an odd head-action whereby she turned a light box in front of her by applying her head, in two slightly different conditions. In the hands-occupied condition, she used her hands in order to hold the blanket around her shoulder while she executed the head-action. In the hands-free condition, she ostensibly placed her free hands on the table while she executed the head-action. [Gergely et al. \(2002\)](#) found that while 69% of the children replicated the head-action in the hands-free condition, only 21% did in the hands-occupied condition. In the hands-occupied condition, the majority of children used their own hands to turn the light box on. [Csibra & Gergely \(2005, 2006\)](#) further report that the asymmetry between infants' replication of the model's odd head-action in the hands-free and hands-occupied conditions vanishes if the model fails to provide infants with ostensive cues.

[Gergely & Csibra \(2003\)](#) have reported evidence that twelve-month-olds expect agents engaged in the execution of *instrumental* actions to select the most efficient action as a means towards achieving their goal (or goal-state), in the context of relevant situational constraints. So

the findings on imitation reported by [Gergely et al. \(2002\)](#) raise the following puzzle. Many more infants replicated the agent's head-action when the teleological relation between the agent's means and the agent's goal was opaque (in the hands-free condition) than when it was transparent (in the hands-occupied condition). Why did infants reproduce the agent's head-action more when it was a *less* efficient means of achieving the agent's goal of switching the light box on?

The Gricean thesis about the *ostensive* nature of communicative agency and the asymmetry between instrumental and communicative agency is relevant to answering this puzzle. Arguably, reception of ostensive signals prepared the infants to interpret the agent's action as a communicative, not an instrumental, action. It made manifest to the infants that the agent intended to make something novel and relevant manifest to them by her subsequent non-verbal communicative action. In the hands-occupied condition, the infants learnt how contact was necessary in order to turn on the light bulb, which was part of an unfamiliar device. Since the model's hands were occupied, the infants whose own hands were free assumed that that they were free to select the most efficient means at their disposal to achieve the same goal as the model. In the hands-free condition, the model could have used her hands, but she did not. So the infants learnt from the model's non-verbal demonstration that they could turn the light on by applying their own heads.

On the one hand, the evidence shows that infants construe imitative learning as a response to an agent's communicative action and that they selectively imitate a model's action as a function of what they take to be relevantly highlighted by the model's communicative act (cf. [Southgate et al. 2009](#)). On the other hand, further evidence shows that newborns prefer to look at faces with direct gaze over faces with averted gaze. Right after birth, they display sensitivity to eye-contact, infant-directed speech or motherese, and infant-contingent distal responsiveness. If preceded by ostensive signals, an agent's gaze shift has been shown to generate in preverbal human infants a referential expecta-

tion, i.e., the expectation that the agent will refer to some object (Csibra & Volein 2008, cf. Csibra & Gergely 2009, and Gergely & Jacob 2013, for review).

One further intriguing piece of evidence for the early sensitivity of human toddlers to the ostensive nature of human communicative agency is offered by experiments that shed new light on the classical A-not-B perseveration error phenomenon first reported by Piaget (1954). Infants between eight and twelve months are engaged in an episodic hide-and-seek game in which an adult repeatedly hides a toy under one (A) of two opaque containers (A and B) in full view of the infant. After each hiding event, the infant is allowed to retrieve the object. During test trials where the demonstrator places the object repeatedly under container B, infants continue to perseveratively search for it under container A where it had been previously hidden. Experimental findings reported by Topal et al. (2008) show that minimizing the presence of ostensive cues results in significant decreases of the perseverative bias in ten-month-olds. This finding is consistent with the assumption that infants do not interpret the hide-and-seek game as a game, but instead as a teaching session about the proper location of a toy.

All this evidence strongly suggests that human infants are prepared from the start to recognize nonverbal ostensive referential signals and action–demonstrations addressed to them as encoding an agent’s communicative intention to make manifest her informative intention to make some relevant state affairs manifest to the addressee. But of course this raises a puzzle: how could preverbal infants recognize an agent’s communicative intention to make manifest her informative intention? A novel approach to this puzzle has been insightfully suggested by Csibra (2010). According to Csibra, very young infants might well be in a position similar to that of a foreign addressee of a verbal communicative act, who is unable to retrieve a speaker’s informative intention for lack of understanding of the meaning of the speaker’s utterance. Nonetheless, the foreign addressee may well recognize being the target of the speaker’s communicative intention on the basis of the speaker’s ostensive

behavior. Furthermore, ostensive signals to which preverbal human infants have been shown to be uniquely sensitive can plausibly be said to *code* the presence of an agent’s communicative intention. If this is correct, then little (if any) further work would be left for preverbal infants to infer the presence of a speaker’s communicative intention after receiving ostensive signals.

6.3 The puzzle about early false-belief understanding

As Millikan has emphasized, much developmental psychology has shown that the majority of three-year-olds fail *elicited-response* false-belief tasks. For example, when asked to predict where an agent with a false belief will look for her toy, most three-year-olds who know the toy’s location point to the toy’s actual location, and not to the empty location where the mistaken agent believes her toy to be. However, in the past ten years or so, developmental psychologists have further designed various *spontaneous-response* false-belief tasks, in which participants are *not* asked any question and therefore *not* requested to produce any answer. Typical spontaneous-response tasks involve the use of the violation-of-expectation and anticipatory-looking paradigms, which involve two steps. In habituation or familiarization trials, participants are first experimentally induced to form expectations by being repeatedly exposed to one and the same event. Second, in test trials of violation-of-expectation experiments, participants are presented with either an expected or an unexpected event. By measuring the time during which participants respectively look at the expected vs. the unexpected event, psychologists get evidence about the nature and content of the infants’ expectations formed during the habituation or familiarization trials. Psychologists can also use the anticipatory-looking paradigm and experimentally determine where participants first look in anticipation of the agent’s action, thereby revealing their expectation about the content of the agent’s belief.

Thus, in a seminal study based on the violation-of-expectation paradigm by Onishi & Baillargeon (2005), fifteen-month-olds saw an

agent reach for her toy either in a green box or in a yellow box when she had either a true or a false belief about her toy's location. Onishi and Baillargeon report that fifteen-month-olds looked reliably longer when the agent's action was incongruent rather than congruent with the content of either her true or false belief. In a study based on the anticipatory looking paradigm, twenty-five-month-olds were shown to look correctly towards the empty location where a mistaken agent believed her toy to be, in anticipation of her action (Southgate et al. 2007). Many further subsequent studies show that toddlers and even preverbal human infants are able to track the contents of others' false beliefs and expect others to act in accordance with the contents of their true and false beliefs.

In a classical experiment by Woodward (1998), six-month-olds were familiarized with an agent's action, who repeatedly chose one of two toys. In the test trials, the spatial locations of the toys were switched and the infants either saw the agent select the same toy as before at a new location or a new toy at the old location. Six-month-olds looked reliably longer at the former than at the latter condition. Luo & Baillargeon (2005) further showed that infants do not look reliably longer at a change of target if, in the familiarization trials, the agent repeatedly reached for the same object, but there was no competing object (for further discussion cf. Jacob 2012). This result has been widely interpreted as showing that six-month-olds are able to ascribe a preference to an agent. Luo (2011) further found that ten-month-olds who know that an agent is in fact confronted with only *one* object (not two) ascribe a preference to the agent if she *falsely believes* that she is confronted with a pair of objects, but *not* if the agent knows (as the infants do) that she is confronted with only one object.

Thus, the psychological investigation of early human social cognition is currently confronted with a puzzle different from that confronted by Millikan: on the one hand, robust findings show that the majority of three-year-olds fail elicited-response false-belief tasks such as the Sally-Anne test. On the other hand, more recent findings based on spontaneous-response

tasks show that preverbal infants expect others to act in accordance with the contents of their true and false beliefs. The puzzle is: how do we make sense of the discrepancy between both sets of experimental findings?

So far, psychologists have offered two broad strategies for this, one of which assumes (as Millikan does) that success at elicited-response false-belief tasks is a necessary condition of the ability to ascribe false beliefs to others, which is taken to be the output of "a cultural process tied to language acquisition" (Perner & Ruffman 2005, p. 214). Their burden is to explain away the findings about preverbal infants without crediting them with the ability to track the contents of others' false beliefs. Thus, the majority of "cultural constructivist" psychologists offer low-level associationist accounts of the findings about preverbal infants based on spontaneous-response tasks. Other psychologists (including Baillargeon et al. 2010; Bloom & German 2000 Leslie 2005; Leslie et al. 2004; Leslie et al. 2005; Scott et al. 2010) argue that the findings about preverbal infants show that they can track the contents of others' false beliefs. Their burden is to explain why elicited-response false-belief tasks are so challenging for three-year-olds. The prevalent non-constructivist explanation is the processing-load account offered by Baillargeon and colleagues.

The core of the associationist strategy is to account for findings about preverbal human infants based on spontaneous-response tasks on the basis of a three-way association between the agent, the object, and its location. It postulates that infants will look longer in the test trials at events that depart more strongly from the three-way association generated by the familiarization trials. For example, in the test trials of Onishi & Baillargeon (2005), infants should look longer when the agent reaches for her toy in the yellow box if in the familiarization trials the agent placed her toy in the green box on three repeated occasions.

The main obstacle for the associationist path is a recent study by Senju et al. (2011) based on the anticipatory-looking paradigm. In the familiarization stage, eighteen-month-olds experience the effect of wearing either an

opaque blindfold through which they cannot see or a *trick* blindfold through which they can see. In the first trials of the test phase, the children are familiarized to seeing an agent retrieve her toy at the location where a puppet has placed it in front of her. The agent's action is always preceded by a pair of visual and auditory cues. In the last test trial, the agent first sees the puppet place the toy in one of the two boxes; she then ostensibly covers her eyes with a blindfold, and finally the puppet removes the toy. After the puppet disappears, the agent removes her blindfold and the cues are produced. Using an eye-tracker, [Senju et al. \(2011\)](#) found that only infants who had experienced an opaque blindfold, not infants who had experienced a trick see-through blindfold, reliably made their first saccade towards the empty location in anticipation of the agent's action.

[Senju et al.'s \(2011\)](#) findings are inconsistent with the associationist strategy: since all infants saw exactly the same events, they should have formed exactly the same threefold association between the agent, the toy, and the location, and on this basis they should have gazed at the same location in anticipation of the agent's action. But they did not. Only infants whose view had been previously obstructed by an opaque blindfold, not those whose view had not been obstructed by a trick blindfold, expected the blindfolded agent to mistakenly believe that the object was still in the opaque container after the puppet removed it.

The evidence against the associationist strategy is also evidence against the assumption (accepted by Millikan) that success at elicited-response false-belief tasks is a necessary condition for having a representational theory of mind and being able to track the contents of others' false beliefs. But this assumption is unlikely to be correct if, as several critics of the cultural constructivist strategy have argued, the ability to ascribe false beliefs to others is not a sufficient condition for success at elicited-response false-belief tasks. As advocates of the processing-load account ([Baillargeon et al. 2010](#)) have argued, an agent could have the ability to ascribe false beliefs to others and still fail elicited-response false-belief tasks for at

least three reasons: she could fail to understand the meaning of the linguistically-encoded sentence used by the experimenter to ask the question. She could fail to select the content of the agent's false belief in the process whereby she answers the experimenter's question. She could fail to have the executive-control resources necessary to inhibit the prepotent tendency to answer the question on the basis of the content of her own true belief. I will now argue that solving the puzzle about early belief-understanding may well depend on acceptance of the Gricean thesis of the ostensive nature of communicative agency and the asymmetry between instrumental and communicative agency.

I now want to offer a speculative solution to the puzzle about early false-belief understanding based on two related Gricean assumptions. The first is the asymmetry between the non-ostensive nature of instrumental agency and the ostensive nature of human communicative agency. The second related assumption is that the human ability to track the content of the false belief of an agent of an instrumental action must be a by-product of the ability to deal with deception (e.g., lying) in the context of human communicative agency.

In the typical Sally-Anne elicited-response false-belief task, participants are requested to make sense of two actions performed by two different agents at the same time: they must track the contents of the motivations and epistemic states of a mistaken agent engaged in the execution of an instrumental action (Sally) and they must also make sense of the communicative action performed by the experimenter who asks them "Where will Sally look for her toy?" The findings based on spontaneous-response tasks strongly suggest that much before they become proficient in language use, young human children are able to spontaneously track the contents of the false beliefs of agents of instrumental actions. So the question is: what is it about the experimenter's question that biases them towards pointing to the toy's actual location?

In [Helming et al. \(2014\)](#), we have argued that two biases are at work, one of which is a referential bias and the other of which is a co-

operative bias. The referential bias itself turns on two components. On the one hand, the experimenter could not ask the question “Where will the agent look for her toy?” unless she referred to the toy. On the other hand, the experimenter shares the participants’ correct epistemic perspective on the toy’s location. In answering the experimenter’s question, participants have the option of mentally representing either the toy’s actual location or the empty location (where the mistaken agent believes her toy to be). The experimenter’s question may bias young children’s answer towards the actual location by virtue of the fact that the experimenter both referred to the toy (whose actual location they know) and shared the participants’ correct epistemic perspective on the toy’s actual location (at the expense of the mistaken agent’s incorrect perspective on the empty location). What we further call the cooperative bias is the propensity of young children to help an agent with a false belief about her toy’s location achieve the goal of her instrumental action by pointing to the actual location (cf. Warneken & Tomasello 2006, 2007; Knudsen & Liszkowski 2012), in accordance with their own true belief about the toy’s actual location. If so, then young children might interpret the prediction question “Where *will* Sally look for her toy?” as a normative question: “Where *should* Sally look for her toy?” Of course, the correct answer to the normative question is the toy’s actual location, not the empty location where the mistaken agent believes her toy to be.

7 Conclusion

The goal of this paper was to assess the gap between Millikan’s particular views about some of the proximate psychological mechanisms underlying human communication and three core assumptions of the Gricean approach: the mindreading thesis, the separability thesis, and the ostensive nature of communicative agency. I have criticized five of Millikan’s basic claims about psychological mechanisms: (i) verbal understanding is best construed as an extended form of perception; (ii) hearers can track the domains of intentional conventional signs

without representing any of the speaker’s psychological states; (iii) the overlap between the interests of speakers and hearers undermines the separability thesis; (iv) humans can predict others’ behavior out of social conformity; (v) developmental psychology supports the view that neither verbal understanding nor language acquisition requires a representational theory of mind.

Millikan’s major teleosemantic contribution has been to open an entirely novel approach to the continued reproduction of intentional conventional public-language signs. As was shown by the discussion of whether her view of the proper function of descriptive and prescriptive utterances is consistent with the separability thesis, there is room for disagreement about particular psychological mechanisms within a teleosemantic approach. I do not think that Millikan’s teleosemantic framework for addressing the continued reproduction of intentional conventional signs mandates the particular choice of proximate psychological mechanisms that she recommends. One of the major challenges for the scientific investigation of cultural evolution is to make sure that the proximate psychological mechanisms that underlie the continued reproduction of human cultural conventions are supported by findings from experimental psychological research, in particular developmental psychology.

Acknowledgements

I am grateful to the editors for inviting me to write this essay, and to Ned Block, Carsten Hansen, Georges Rey, Dan Sperber and two anonymous reviewers for their comments.

References

- Artiga, M. (forthcoming). *Signaling without cooperation*.
- Austin, J. (1975). *How to do things with words*. Oxford, UK: Oxford University Press.
- Baillargeon, R., Scott, R. M. & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14 (3), 110-118. [10.1016/j.tics.2009.12.006](https://doi.org/10.1016/j.tics.2009.12.006)
- Baron-Cohen, S., Leslie, A. & Frith, U. (1985). Does the autistic child have a “theory of mind?”. *Cognition*, 21 (1), 37-46. [10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Bloom, P. & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77 (1), 25-31. [10.1016/S0010-0277\(00\)00096-2](https://doi.org/10.1016/S0010-0277(00)00096-2)
- Bugnyar, T. (2011). Knower-guesser differentiation in ravens: others’ view points matter. *Proceedings of the Royal Society of London B: Biological Sciences*, 278 (1705), 634-640. [10.1098/rspb.2010.1514](https://doi.org/10.1098/rspb.2010.1514)
- Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12 (5), 187-192. [10.1016/j.tics.2008.02.010](https://doi.org/10.1016/j.tics.2008.02.010)
- Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language*, 25 (2), 141-168. [10.1111/j.1468-0017.2009.01384.x](https://doi.org/10.1111/j.1468-0017.2009.01384.x)
- Csibra, G. & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13 (4), 148-153. [10.1016/j.tics.2009.01.00](https://doi.org/10.1016/j.tics.2009.01.00)
- Csibra, G. & Volein, A. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, 26 (1), 1-11. [10.1348/026151007X185987](https://doi.org/10.1348/026151007X185987)
- Dally, J. M., Emery, N. J. & Clayton, N. S. (2006). Food-caching western scrub-jays keep track of who was watching when. *Science*, 312 (5780), 1662-1665. [10.1126/science.1126539](https://doi.org/10.1126/science.1126539)
- Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford University Press.
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- (1990). Reply to reviewers of explaining behavior: Reasons in a world of causes. *Philosophy and Phenomenological Research*, 50 (4), 819-839.
- (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, 7 (7), 287-292. [10.1016/S1364-6613\(03\)00128-1](https://doi.org/10.1016/S1364-6613(03)00128-1)
- (2005). The social construction of the cultural mind: imitative learning as a mechanism of human pedagogy. *Interaction Studies*, 6 (3), 463-481. [10.1075/bct.4.19ger](https://doi.org/10.1075/bct.4.19ger)
- (2006). Sylvia’s recipe: the role of imitation and pedagogy in the transmission of cultural knowledge. In S. Levenson & N. Enfield (Eds.) *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229-255). Oxford, UK: Berg Publishers.
- Gergely, G., Bekkering, H. & Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature*, 415 (6873), 755-755. [10.1038/415755a](https://doi.org/10.1038/415755a)
- Gergely, G. & Jacob, P. (2013). Reasoning about instrumental and communicative agency in human infancy. In J. B. Benson, F. Xu & T. Kushnir (Eds.) *Rational constructivism in cognitive development* (pp. 59-94). Waltham, MA: Academic Press.
- Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In G. MacDonald & D. Papineau (Eds.) *Teleosemantics* (pp. 42-68). Oxford, UK: Oxford University Press.
- (2013). Signals, icons, and beliefs. In D. Ryder, J. Kinsbury & K. Williford (Eds.) *Millikan and her critics* (pp. 41-58). Oxford, UK: Blackwell.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66 (3), 377-388. [10.2307/2182440](https://doi.org/10.2307/2182440)
- (1969). Utterer’s meaning and intentions. *The Philosophical Review*, 78 (2), 147-177. [10.2307/2184179](https://doi.org/10.2307/2184179)
- (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Hare, B., Call, J. & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61 (1), 139-151. [10.1006/anbe.2000.1518](https://doi.org/10.1006/anbe.2000.1518)
- Hauser, M., 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. [10.1126/science.298.5598.1569](https://doi.org/10.1126/science.298.5598.1569)
- Helming, K.A, Strickland, B. & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, 18 (4), 167-170. [10.1016/j.tics.2014.01.005](https://doi.org/10.1016/j.tics.2014.01.005)
- Jacob, P. (1997). *What minds can do*. Cambridge, UK: Cambridge University Press.
- (2011). Meaning, intentionality and communication. In C. Maierborn, K. Heusinger & P. Portner (Eds.) *Semantics: an international handbook of natural language meaning* (pp. 11-24). Berlin, GER: Walter de

- Gruyter.
- (2012). Sharing and ascribing goals. *Mind and Language*, 27 (2), 200-227. [10.1111/j.1468-0017.2012.01441.x](https://doi.org/10.1111/j.1468-0017.2012.01441.x)
- Knudsen, B. & Liszkowski, U. (2012). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, 17, 672-691.
- Krebs, J. R. & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.) *Behavioral ecology* (pp. 380-402). Sunderland, MA: Sinauer Associates.
- Leslie, A. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, 9, 459-462.
- Leslie, A. M., Friedman, O. & German, T. P. (2004). Core mechanisms in “theory of mind”. *Trends in Cognitive Sciences*, 8 (12), 528-533.
- Leslie, A. M., German, T. P. & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, 50, 45-85.
- Luo, Y. (2011). Do 10-month-old infants understand others’ false beliefs? *Cognition*, 121 (3), 289-298. [10.1016/j.cognition.2011.07.011](https://doi.org/10.1016/j.cognition.2011.07.011)
- Luo, Y. & Baillargeon, R. (2005). Can a self-propelled box have a goal? *Psychological Science*, 16 (8), 601-608. [10.1111/j.1467-9280.2005.01582.x](https://doi.org/10.1111/j.1467-9280.2005.01582.x)
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134 (3489), 1501-1506. [10.1126/science.134.3489.1501](https://doi.org/10.1126/science.134.3489.1501)
- Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.
- (1989a). Biosemantics. *The Journal of Philosophy*, 86 (6), 281-297.
- (1989b). In defense of proper functions. *Philosophy of Science*, 56 (2), 288-302.
- (1990). Seismograph reading for “Explaining Behavior”. *Philosophy and Phenomenological Research*, 50 (4), 807-812.
- (1995). Pushmi-Pullyu Representations. *Philosophical Perspectives*, 9, 185-200.
- (1998). Conventions made simple. *The Journal of Philosophy*, 95 (4), 161-180.
- (2000). *On clear and confused ideas: An essay about substance concepts*. Cambridge, UK: Cambridge University Press.
- (2004). *Varieties of meaning*. Cambridge, MA: MIT Press.
- (2005). *Language: A biological model*. Oxford, UK: Oxford University Press.
- (2006). Reply to Tomasello. In M. Mazzone (Ed.) *Symposium on language: A biological model by Ruth Millikan*,. SWIF Philosophy of Mind Review.
- (2011). Reply to Neander. In D. Ryder & K. Williford (Eds.) *Millikan and her Critics* (pp. 37-41). Oxford, UK: Blackwell.
- Neander, K. (1991). Functions as selected effects. *Philosophy of Science*, 58 (2), 168-184.
- (1995). Misrepresenting & malfunctioning. *Philosophical Studies*, 79 (2), 109-141. [10.1007/BF00989706](https://doi.org/10.1007/BF00989706)
- (2004). Teleological theories of mental content. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/content-teleological/>
- (2007). Biological approaches to mental representation. In M. Matthen & C. Stephens (Eds.) *Handbook of the philosophy of science: Philosophy of biology* (pp. 548-565). Elsevier.
- (2011). Towards an informational semantics. In D. Ryder, J. Kinsbury & K. Williford (Eds.) *Millikan and her critics* (pp. 21-36). Oxford, UK: Blackwell.
- Onishi, K. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308 (5719), 255-258. [10.1126/science.1107621](https://doi.org/10.1126/science.1107621)
- Origg, 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, UK: Cambridge University Press.
- Perner, J. & Ruffman, T. (2005). Infants’ insight into the mind: How deep? *Science*, 308 (5719), 214-216. [10.1126/science.1111656](https://doi.org/10.1126/science.1111656)
- Perry, J. (1986). Thought without representation. *Proceedings of the Aristotelian Society*, 60 (137), 137-152.
- Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.
- Pietroski, P. (1992). Intentionality and teleological error. *Pacific Philosophical Quarterly*, 73 (3), 267-282.
- Recanati, F. (2002). Does communication rest on inference? *Mind and Language*, 17 (2), 105-126. [10.1111/1468-0017.00191](https://doi.org/10.1111/1468-0017.00191)
- (2007). Millikan’s theory of signs. *Philosophy and Phenomenological Research*, 75 (3), 674-681. [10.1111/j.1933-1592.2007.00103.x](https://doi.org/10.1111/j.1933-1592.2007.00103.x)
- Scott, R. M., Baillargeon, R., Song, H. & Leslie, A. (2010). Attributing false beliefs about nonobvious properties at 18 months. *Cognitive Psychology*, 61, 366-395.

- Scott-Philipps, 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. [10.1177/1745691610393528](https://doi.org/10.1177/1745691610393528)
- Senju, A., Southgate, V., Snape, C., Leonard, M. & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, 22 (7), 878-880. [10.1177/0956797611411584](https://doi.org/10.1177/0956797611411584)
- Shea, N. (2007). Consumers need information: supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research*, 75 (3), 404-435. [10.1111/j.1933-1592.2007.00082.x](https://doi.org/10.1111/j.1933-1592.2007.00082.x)
- Sober, E. (1984). *The Nature of selection: Evolutionary theory in philosophical focus*. Chicago, IL: Chicago University Press.
- Southgate, V., Senju, A. & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, 18 (7), 587-592. [10.1111/j.1467-9280.2007.01944.x](https://doi.org/10.1111/j.1467-9280.2007.01944.x)
- Southgate, V., Chevallier, C. & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 12 (6), 1013-1019. [10.1111/j.1467-7687.2009.00861.x](https://doi.org/10.1111/j.1467-7687.2009.00861.x)
- Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.) *Metarepresentations: A multidisciplinary perspective* (pp. 117-137). Oxford, UK: Oxford University Press.
- (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29 (1/2), 401-413. [10.5840/philtopics2001291/215](https://doi.org/10.5840/philtopics2001291/215)
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010). Epistemic vigilance. *Mind and Language*, 25 (4), 359-393. [10.1111/j.1468-0017.2010.01394.x](https://doi.org/10.1111/j.1468-0017.2010.01394.x)
- Sperber, D. & Wilson, D. (1986). *Relevance, communication and cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2006). Conventions are shared. In M. Mazzone (Ed.) *Symposium on language: A biological model by Ruth Millikan*. SWIF Philosophy of Mind Review. <http://lgxserver.uniba.it/lei/mind/swifpmr.htm>
- (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
- Tomasello, M., Call, J. & Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences*, 7 (4), 153-156. [10.1016/S1364-6613\(03\)00035-4](https://doi.org/10.1016/S1364-6613(03)00035-4)
- Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675-691.
- Topal, J., Gergely, G., Miklosi, A., Erdohegyi, A. & Csibra, G. (2008). Infant perseverative errors are induced by pragmatic misinterpretation. *Science*, 321 (1269), 1831-1834. [10.1126/science.1176960](https://doi.org/10.1126/science.1176960)
- Warneken, F. & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311 (5765), 1301-1303. [10.1126/science.112144](https://doi.org/10.1126/science.112144)
- (2007). Helping and cooperation at 14 months of age. *Infancy*, 11 (3), 271-294. [10.1111/j.1532-7078.2007.tb00227.x](https://doi.org/10.1111/j.1532-7078.2007.tb00227.x)
- 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. [10.1111/1467-8624.00304](https://doi.org/10.1111/1467-8624.00304)
- Wilson, D. & Sperber, D. (2004). Relevance theory. In L. Horn & G. Ward (Eds.) *The handbook of pragmatics*. Oxford, UK: Blackwell.
- 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. [10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69 (1), 1-34. [10.1016/s0010-0277\(98\)00058-4](https://doi.org/10.1016/s0010-0277(98)00058-4)
- Wright, L. (1973). Functions. *The Philosophical Review*, 82 (2), 139-168.

Communicative Agency and *ad hominem* Arguments in Social Epistemology

A Commentary on Pierre Jacob

Marius F. Jung

A central point in Jacob's paper focuses on the incompatibility of Grice and Millikan's account of communicative agency. First, the Gricean mindreading thesis is incompatible with Millikan's direct perception account. Second, the account of cooperative devices, defended by Millikan, contradicts the Gricean separability thesis in a broad sense. While I agree with Jacob that these positions are indeed incompatible, I will shift focus and concentrate on issues concerning social epistemology with regard to communicative agency. A main issue in social epistemology concerns the accessibility of the speaker's reliability. How could the hearer remain epistemically vigilant without using fallacious reasoning? (i) I argue that the hearer, in order to be epistemically vigilant, could commit a local *ad hominem* attack, a process of inductive Bayesian reasoning which is an epistemic tool for assessing the speaker's reliability. (ii) Compared to this, a global *ad hominem* attack is a fallacious kind of reasoning, because it undermines knowledge transmission and it cannot be calculated in *Bayes' Theorem*. (iii) The account of a local *ad hominem* attack fits with Grice's mindreading thesis, which is incompatible with Millikan's account of direct perception. (iv) The Gricean separability thesis could better explain occurrences of *ad hominem* attacks than Millikan's assumption that speaker and hearer are cooperative devices.

Keywords

Bayesian reasoning | Communicative intention | Cooperative devices | Direct perception | Epistemic injustice | Epistemic vigilance | Global *ad hominem* argument | Informative intention | Local *ad hominem* argument | Mindreading | Personal attack | Positive reasons | Separability thesis

1 Introduction: Grice's individualistic account of meaning and epistemic trustworthiness

One of the main findings of Jacob's paper is a detailed elaboration of the differences between Millikan's (1984, 2004, 2005) communicative agency and the Gricean (Grice 1957, 1969, Sperber & Wilson 1986) account of speaker's meaning and intention. Jacob argues that the Gricean mindreading thesis, the separability

thesis, and the ostensive nature of communication are not supported by Millikan's account of the direct perception of speaker's intention, which supports a non-inferential model of the understanding of intentional signs. Furthermore, the Gricean account is incompatible with Millikan's claim that speaker and hearer are co-

Commentator

Marius F. Jung

mjung02@students.uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

Target Author

Pierre Jacob

jacob @ ehess.fr
Institute Jean Nicod
Paris, France

Editors

Thomas Metzinger

metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt

jennifer.windt@monash.edu
Monash University
Melbourne, Australia

operative devices; the claim that the prediction of another's behavior could be explained through reliance on socially established conformities and conventions and that modern developmental psychology could get along without any theory of mind.

A very influential account of naturalizing the content of intentional mental representations is Millikan's teleosemantic framework (Jacob 2010). According to this view, the content of intentional mental representations can best be naturalized by relying on the history of the biologically-selected functions of these representations, namely the *direct proper functions* (cf. Millikan 1984, 1989). Interestingly, Jacob focuses on Millikan's concept of communicative agency, which is strongly connected to the teleosemantic framework, and argues that there are several aspects of the Gricean individualistic account of meaning that are more plausible than Millikan's when it comes to explaining social communicative agency. The most illuminating finding of Jacob's paper is the modern and precise presentation of the actuality of the Gricean separability thesis and the mindreading account, because it is explanatorily fruitful not only for philosophy of language, but also for social cognition, social epistemology, informal logics, and the relation between these different studies.

I generally agree with Jacob's main findings, nevertheless I will address some further issues of Millikan and Grice's account with regard to philosophical problems in social epistemology. Before I respond to these in detail, I first focus on the Gricean account and its implications for social epistemology.

The well-known Gricean (Grice 1957) account of the meaning of an utterance focuses on analysis of the speaker's meaning in a conversation. A speaker *S* means something *unnatural* if she intends something by the utterance of a sentence.¹ Let us suppose that the speaker is a politician with a specific agenda and with a propensity for aggressive propaganda. She utters the following:

(1) Speaker: "Our party will ensure that taxes go down".

The speaker *S* means (1) iff *S* utters (1) with the intention that a hearer *H* will gain the belief that *S*' party will make sure that taxes go down, if (i) the hearer *H* recognizes the speaker's intention (1) and (ii) because of that she gains the belief that *S*' party will make sure that taxes go down, (iii) since she recognizes that the speaker's intention is exactly that (cf. Grice 1957, 1969).

Since Gricean meaning is individualistic and subjective it is important to note which underlying cognitive states constitute this meaning. As Jacob puts it—relying on Sperber & Wilson's (1986) interpretation of the Gricean account²—there are three main assumptions upon which the psychological theory of meaning is based, namely the separability thesis, the mindreading thesis, and the asymmetry between an informative and communicative intention. Together with Jacob I shall focus on Sperber and Wilson's account, which argues that the Gricean theory can be summarized as a reciprocal process of intentions. The *informative intention* is an intention of a speaker who wants to inform a hearer about some state of affairs. In order to be successful, the speaker has also the intention that the hearer recognize the informative intention (Grice 1957). This means that at first the hearer has to understand the informative intention. If she understands it, the *communicative intention* of the speaker has been fulfilled. But the informative intention will only be fulfilled, if the speaker is trustworthy: a necessary condition for accepting a speaker's utterance. In effect, the hearer gains a new belief. The assessment of her trustworthiness depends on the hearer of that intention (Sperber & Wilson 1986). She must admit that the speaker has to be reliable in order to be trustworthy, or, to put it in Jacob's (this collection, p. 4) words, "the addressee must further accept the speaker's epistemic or practical authority". But a question arises: on which kind of epistemic practices does the hearer have to rely in order to accept the

¹ Grice (1957) distinguishes between a *natural* and an *unnatural meaning*. Unnatural meaning is always characterized by the speaker's intention. The natural meaning of a sign characterizes meaning that is independent of a speaker.

² I follow Jacob in relying partly on Sperber & Wilson (1986) when I talk about the Gricean account.

speaker as an epistemic authority? I will address this question in the following commentary.

In order to answer it, I will argue that (i) a hearer could commit a *local ad hominem attack*, a process of inductive *Bayesian reasoning* that secures epistemic vigilance. Roughly, an *ad hominem* attack is an argument that considers rather personal properties of an utterer than the argument itself. (ii) A fallacious kind of the personal attack is the *global ad hominem attack*, which undermines every testimony of a speaker because of its personal traits. (iii) The Gricean account of mindreading could better account for an inductive inference model than Millikan's direct-perception-account. (iv) Practices of *ad hominem* attacks, I will argue, support the Gricean separability thesis, while Millikan's cooperative devices account is less plausible.

The structure of this commentary will be as follows: first, I focus on Grice and Millikan's framework and its implications for social epistemology, namely the problem of epistemic reliability (cf. section 2). In section 3 I shall present Lackey's account of a social epistemological dualism, a hybrid theory in which Lackey tries to connect the most plausible findings of social epistemological reductionism and anti-reductionism. Then I argue that the Gricean account of informative intentions and Lackey's *positive reason component* could lead to the *personal attack* or *ad hominem argument* (cf. section 4). In section 5 I argue that there are two possible commitments of *ad hominem* arguments, to be specific, the *global* and the *local ad hominem attack* (cf. section 5.1, section 5.2). In the Gricean account of the *mindreading thesis* is compatible with the drawn picture of our social epistemological practices, because it supports the inductive inference model, while Millikan's account of *direct perception* could not account for this. The Gricean *separability thesis* fits nicely with the positive reasons component and the reliance on *ad hominem* arguments, while

Millikan's account of speaker and hearer as *co-operative devices* is less plausible (cf. section 6).

2 Epistemic intentions and epistemic reliability

The utterance of a speaker depends on two directions of fit. The first can be characterized as a *mind-to-world-relation*. Here, the speaker has the intention of conveying some states of affairs about the actual world. This direction of fit implies that the speaker wants to share some epistemic notions. If she is successful in doing so, the hearer will gain a true belief. This class of utterances is descriptive.

The second is a *world-to-mind direction* of fit of the speaker's utterance. Here the speaker wants to convey some of her desires to the hearer, who acts in a particular way in order to fulfill the speaker's desire. The intention is fulfilled if the hearer gains a new desire to act in order to fulfill the speaker's desire (Sperber & Wilson 1986). This kind of direction of fit is unimportant for the following account. Here I shall focus on descriptive utterances.

Before I address some implications of epistemic intentions, I shall focus on the *separability thesis*. This thesis addresses the problem of an asymmetry of interests between hearer and speaker. Since the interests are not identical, the speaker could deceive the hearer. And the other way around: the hearer could distrust the speaker even though she utters a true sentence. Sperber et al. (2010) claim that some amount of distrust is a stabilizer in the evolution of human communication, which they call *epistemic vigilance*. Imagine that humans believed almost everything they were told. Since not every speaker has the propensity to speak the truth, hearers would have a lower amount of knowledge because they would have no tool for distinguishing a reliable testimony from a non-reliable one. Communication would be very imprecise, because knowledge agency would be less successful. Hence, epistemic vigilance is a precondition for cooperative communication, because both speaker and hearer check the reliability of knowledge transition. I agree with Sperber et al. (2010) that epistemic vigilance is a

feature of a *source* and the *content* of information.³

Let me go back to the mind-to-world-relation of fit. If the twofold account given by [Sperber & Wilson \(1986\)](#) is correct, the hearer has gained a true belief (about some state of affairs). To count as knowledge, we have to ask whether the true belief is justified.⁴ What could count as a justification? Some social epistemologists would say that the testimony of the speaker is sufficient to count as a justification. This is the thesis of an *anti-reductionism* in social epistemology which contains the claim that the speaker must not rely on other sources of knowledge such as perception, inference or memory to justify her belief ([Coady 1992](#)). A *reductionist* would say that the testimony cannot count as knowledge without relying in addition upon other sources of knowledge ([Fricker 1995](#)).

In [Millikan's \(2005\)](#) account of an intentional conventional sign (which is the content of an intentional mental representation), she assumes that speaker and hearer are cooperative devices that have co-evolved. The relationship between sender and receiver can be characterized as beneficial in the long term. Millikan proposes a framework, in which the descriptive representations describe a mind-world-direction, whereas the *directive* representation describes the world-to mind relation. The long term beneficial communicative agency between sender and receiver characterizes a function of reproduction of conventional signs. The *direct proper function* in this particular case is that the hearer gains a new belief. [Millikan \(1984\)](#) is well known for this teleosemantic account that can deal with misrepresentations. In such a case, the proper function remains unfulfilled. It is unfulfilled if the speaker fails to cause a new belief in the hearer ([Millikan 1984, 2005](#)). But the hearer could also be responsible for the unfulfilled proper function if she mistakenly judges the speaker to be untrustworthy. The hearer is also

a constitutive part of the cooperative devices that establish the direct proper function ([Millikan 1984, 2005](#)).

3 Social epistemology: Lackey's dualism

Before I present a more detailed account of *ad hominem arguments*, I will say a few words about social epistemology and the position that is presupposed in this commentary. [Jennifer Lackey's \(2006\)](#) account of social epistemology relies upon a kind of dualism, in which she combines anti-reductionism and reductionism. According to her, social epistemology has made the mistake of addressing the debate between reductionism and non-reductionism unilaterally. *Reductionism* takes epistemic responsibility and the rationality of the hearer far too seriously, because the hearer has to rely upon other sources of knowledge like perception, memory, deductive inferences, etc. The claim here is that testimony is not a source of knowledge in the first place, because a hearer could never know the intentions of a speaker who held accidentally or intentionally false beliefs. In contrast, *anti-reductionism* always focuses on the speaker's perspective and her propensity for credible testimonials. Proponents of anti-reductionism claim that a large amount of our knowledge depends on testimonials. We would know almost nothing if we were as restrictive as the reductionist claims ([Coady 1992](#)). Lackey wants to combine these two accounts in a kind of dualism. Her dualism contains the presupposition of the reliability of the speaker along with *positive reasons* to accept the speaker's testimony, evaluated from the hearer's perspective. If the speaker utters a true sentence and the hearer has positive reasons to trust the speaker, then knowledge from testimony is possible. [Lackey](#) argues for the following conditional, which contains three necessary conditions:⁵

For every speaker A and hearer B, B justifiably believes that p on the basis of A's testimony that p only if: (1) B believes that p on the basis of the content of A's

³ I will address this topic with respect to the *ad hominem* argument in section 5.

⁴ I will not address Gettier cases with regard to social epistemology. For the sake of this commentary, I will use the term *knowledge* as meaning justified true belief. Issues concerning the *ad hominem* fallacy will concern the justification-condition.

⁵ Lackey claims that dualism accounts only for necessary conditions for a source of knowledge.

testimony that p , (2) A's testimony that p is reliable or otherwise truth conducive, and (3) B has appropriate positive reasons for accepting A's testimony that p . (2006, p. 170)

For the present account it is important that a testimony, given by A, qualifies as a source of knowledge that depends on the hearer having positive reasons to think that A's testimony is reliable. Recall the informative intention and the mind-to-world-direction of fit, which is fulfilled if the speaker causes a new belief in the hearer. The direct proper function of the co-operation between speaker and hearer in Milikan's (1984, 2005) account would be fulfilled. But according to Lackey's condition (3), the achievement of a new belief is only justified if there are various *positive reasons* that account for the reliability of the speaker's testimony. Consider the account of Sperber et al. (2010, p. 379) that "the filtering role that epistemic vigilance [...] in the flow of information in face-to-face interaction" is an important feature of communicative agency. But which kind of filtering do they mean? In other words, what are positive reasons, exactly? Could they be past experiences about the reliability of the speaker or even a group to which the speaker belongs?

4 *Ad hominem* arguments and epistemic injustice

I claim that the Gricean account of communication supports our social practices of committing *ad hominem* arguments. The committing of *ad hominem* attacks in communicative agency becomes patent when you look at *positive reasons* in more detail. During past events of communicative agency, a hearer has tested the trustworthiness of several speakers on the basis of her personal properties and the context to which these properties have been related (Lackey 2006; Fricker 2007). The ability of being epistemically vigilant emerges very early in human development. At the age of three years, infants already prefer testimony from a reliable source (cf. Clément 2010). From that age on, infants develop a "cognitive filter that enables children

to take advantage of testimony without the risk being completely misled" (Clément 2010, p. 545).⁶

Now back to the example: suppose that the hearer in question has been confronted with the testimony of politicians in the past. She then hears the following sentence from speaker S :

(1) "Our party will make sure that taxes will go down".

Would you, as a hearer, believe her? Consider past cases of political propaganda and ask yourself how reliable the politician, the speaker, really is. At first, let us assume you do not. You are a very skeptical person, especially when it comes to political issues. Is it rational to be skeptical, so are you guilty of prejudice? Let us assume that the speaker is surprisingly reliable. She speaks the truth. The party wins and reduces taxes. Have you treated the politician in an epistemic inequitably way or was it the only way to remain epistemic vigilant? These questions will be addressed in the following sections.

The hearer who does not believe S ' utterance (1) has committed an *ad hominem* argument, a personal attack against the speaker. According to Walton (2008, p. 170) "[t]he *argumentum ad hominem*, meaning 'argument directed to the man', is the kind of argument that criticizes another argument by criticizing the arguer rather than his argument." A hearer takes some personal properties, such as being a politician, and infers that the expressed sentence is false. It is *prima facie* irrelevant to consider personal traits as indicators of a false testimony t (Yarp 2013; Walton 1998).⁷ Keeping Walton in mind, we are able to generalize *ad hominem* attacks as follows:

⁶ There is further evidence in developmental psychology that speaks of very early acquisition and practice of epistemic vigilance (cf. Clément 2010; Sperber et al. 2010; Mascaro & Sperber 2009). Another issue with regard to the positive reasons component in social epistemology is the so-called *infant/child-objection*. This concerns the hearer's competence in evaluating the speaker's reliability that small children lack, which is often construed as an argument against reductionism (Lackey 2006). For a general discussion see Lackey (2005).

⁷ As will be seen in section 5.1, there are some exceptions where personal properties are relevant.

Ad hominem attack

- (1) Speaker *S* gives a testimony *t*.⁸
 (2) The speaker's *S* property φ is a negative property with regard to trustworthiness.
 (3) Speaker *S* has a negative property φ , which is ascribed as relevant for her testimony *t* by hearer *H*.
-
- (C) The testimony *t*, uttered by speaker *S*, is false as assessed from hearer *H*.⁹

The arguments of the speaker (implicitly represented or explicitly formulated) have not been challenged seriously by the hearer. She just considers personal traits sufficiently to reject the given argument or proposition.¹⁰ The allegedly suboptimal personal characteristics of the person do not provide any evidence for rejecting the proposition *p*. The hearer neither shows that the deduction of the speaker includes fallacious reasoning nor that the premises on which her proposition is based are wrong. Informal logic does not support the hearer in this situation (Groarke 2011). Has the speaker been treated inequitably? Miranda Fricker (2007) tries to answer this question and introduces the notion of *epistemic injustice*. She generalizes the notion as follows:

Any epistemic injustice wrongs someone in their capacity as a subject of knowledge, and thus in a capacity essential to human value; and the particular way in which testimonial injustice does this is that a hearer wrongs a speaker in his capacity as a giver of knowledge, as an informant. (Fricker 2007, p. 5)

⁸ I assume that a testimony *t* expresses an argument that contains the relevant proposition *p*.

⁹ One could of course distinguish between a testimony and an argument. Here I presuppose that a testimony is somehow a conclusion of an argument. Fricker (2007, p. 61) supports this view as follows: "One might be inclined to put a familiar picture of justification to the fore and argue that in order to gain knowledge that *p* from somebody telling her that *p*, the hearer must in some way (perhaps very swiftly, perhaps even unconsciously) rehearse an argument whose conclusion is *p*."

¹⁰ For the purpose of this commentary I will defend a weak view of propositions. The utterance of a speaker expresses a proposition that is true if it represents a state of affairs. There is of course an asymmetry between the propositional content of an utterance and the propositional content of the speaker's belief or thought. I agree with Jacob (1987) that it is sufficient to assume similarity between the two.

It fits Fricker's generalization that the capacity of a speaker to convey true beliefs is undermined. The positive reasons clause of Lackey's dualism also supports this step of reasoning because the character or the identity of a speaker could be relevant for her evaluation of trustworthiness in epistemic contexts (cf. Fricker 2007; Lackey 2006). Crucially, stereotypes and prejudices—based upon *ad hominem* arguments—are paradigmatic cases of epistemic injustices (cf. Fricker 2007). But is it not rational for a hearer to distrust our politician? Jacob (this collection, pp. 4–5) claims that "not every speaker is (or should be) granted equal epistemic or practical authority on any topic by every addressee." Remember that the hearer's positive reason component is a remainder of the reductionist account with regard to testimony as a justifier of knowledge. Is it not a necessary condition for the positive reason component to remain vigilant in such contexts? If epistemic vigilance does not play a role in this context, then Lackey's suggestion of the necessary condition of positive reasons on the hearer's side is implausible. In the following I shall argue that epistemic vigilance is very important and that the dualistic account could account for it. Nevertheless, one has to accept what I call a local *ad hominem* argument in order to be epistemically vigilant in our particular case.

5 Two kinds of *ad hominem* attack

In the following section I make a suggestion in order to disarm the problem of the *ad hominem* argument with regard to the *positive reason* component. Does an *ad hominem* attack always include fallacious reasoning? Walton (2008, p. 170) claims that "the *argumentum ad hominem* is not always fallacious, for in some instances questions of personal conduct, character, motives, etc., are legitimate and relevant to the issue." Even though cases of *ad hominem* arguments might sometimes be informally fallacious, there are some highly relevant cases in which a particular *ad hominem* attack could be committed in order to remain epistemic vigilant. Since you, as a hearer, have a set of positive reasons—for instance being aware of the usual verbal

espousals of politicians during an election campaign—you are forced to commit a personal attack. Did your reliability assessment rely on fallacious reasoning? In the literature there is a common distinction between three types of the *ad hominem* argument: the *abusive*, the *circumstantial*, and the *tu quoque argument* (Groarke 2011; Walton 1998, 2008).¹¹ These kinds of personal attacks describe various pragma-dialectical reasoning in interpersonal communicative relationships. Below (cf. section 5.1, 5.2) I want to draw a further distinction between two kinds of *ad hominem* attack that are closely connected to communicative situations explicitly involving knowledge transmission. Hence I want to provide a framework that fits well with social epistemological dualism.

Before this, I want to address the *Bayesian argumentation model*, which is presupposed by the following account of *ad hominem* arguments. Few things have been said about the inductive reasoning model which is the underlying mechanism of an *ad hominem* attack. Roughly, one has to consider past experiences with regard to reliability, constituted by contexts and speaker properties, to adjust this experience for future communication. Harris et al. (2012) provide an account that fits well with an inductive model of reasoning, because, potentially evidence is not provided by deductive reasoning. They claim that the evaluation of a proposition or a given testimony is based on an individual's probabilities, which could be described formally using *Bayes' Theorem*. A big advantage of this account is that it describes our subjective evaluations in daily experiences very well. Often, when we are asked “do you believe *S*?” we are inclined to say something like “I am not sure. I guess not”. This could be well explained with the Bayesian model, where an individual's belief does not have a truth-value of 0 or 1. The relevant belief is instead estimated in one's subjective degree of that belief, as a probability between 0 and 1. Let us embed this in our current considerations. The utterance type of the politician is already embedded in a

kind of bias or in posterior beliefs about politicians in general: this is called the hypothesis *h*, and has a particular probability $P(h)$ in isolation from evidence *e*. Evidence *e* in this particular case is constituted by personal characteristics, properties, and circumstances of the utterance. The receiving of the new evidence *e* should update $P(h)$, the probability of the hypothesis. Individuals ought, according to the normative stipulation, if they receive any evidence, to let it influence the probability of the proposition: “[this] normative procedure by which individuals should update their degree of belief in a hypothesis *h* upon receipt of an item of evidence *e* is given by Bayes' Theorem:

$$P(h|e) = \frac{P(h)P(e|h)}{P(e)} \quad \text{“(Harris et al. 2012, p. 316)”}$$

$P(h|e)$ describes the conditional probability of a hypothesis being true after one has received evidence *e*. $P(e|h)$ terms the conditional probability of receiving evidence *e*, given hypothesis *h*. $P(e)$ just describes the evidence in isolation from the truth-values of the hypothesis *h* (cf. Harris et al. 2012).

5.1 Local *ad hominem* attack

I am now able to distinguish two kinds of *ad hominem* attack, a local and a global one. I will thus sketch out some considerations with regard to the presented Bayesian framework.

The positive reason component describes the practice of a speaker's credibility assessment. This process of credibility assessing could lead to what I will call a *local ad hominem* attack. Roughly, one commits a local *ad hominem* argument if one acknowledges someone to be trustworthy in general, but with some exceptions in particular cases. If you ask the politician what time it is or the straightest way to the subway station, it is very unlikely that she would have the intention of deceiving you (Spencer 2001). Hence, one would count her as a trustworthy person. Nonetheless, given the particular information about her party and the reduction of taxes, you might find it unlikely that she is telling the truth. As long as you do not dismiss her in general as an eligible bearer of

¹¹ For some very interesting empirical investigations with regard to these three kinds of *ad hominem* arguments, see van Eemeren et al. (2000) and van Eemeren et al. (2008).

knowledge, it is vigilant in a rational way to distrust her in this case. This view could be described as a subclass of epistemic vigilance, because it is an evolved tool that minimizes the risks of deception which is—according to Sperber et al. (2010)—a condition for cooperative communicators.

At this stage, subjective Bayesian probability comes into play. The general bias of your past experiences with regard to politicians, which enters the stage before you have received the evidence, could be described with (h) . The probability of the hypothesis $P(h)$ is the probability of your believing her without having received evidence e . Given (1), evidence e describes that the person is a politician during a campaign, which also has a particular probability, termed $P(e)$. The evidence condition is the part that divides the local *ad hominem* attack from the global, because the evidence is able to influence one's subjective degree of probability. The personal traits of the speaker as well as the context of utterance-use serve as evidence e . $P(h/e)$ is then the conditional probability of h , if the hearer H receives evidence e . Given the hypothesis h , the probability of receiving evidence e is described by $P(e/h)$. As presented above, this could be calculated within *Bayes' Theorem* (cf. Harris et al. 2012).

As you can see, the *content* and the *source* of information, which serve as evidence e , are both similarly relevant for a testimony (cf. Sperber et al. 2010). Buening (2005) calls this kind of reasoning *relevance-based* with regard to the relevant circumstances that could invoke *ad hominem* attacks. The context and content of the utterance or proposition in question is important for assessment. Walton (1998) calls such a context and content-related *ad hominem* attack a *credibility function*. The proposition in question undergoes an *ethotic rating*, a kind of evaluation of a person's epistemic input value (of her testimony), which can go up or down. When committing a local *ad hominem* attack, the rating goes down. Hence the credibility of the speaker is undermined in a specific case that affects the proposition, which fits with the account of Bayesian argumentation. The local *ad hominem* argument is an example of a non-

fallacious *ad hominem* attack because it is *content* as well as *context-related* and epistemically equitable. The normative stipulation of the Bayesian account that a hearer “should update [her] probabilistic degrees of belief in a hypothesis in accordance with the prescriptions of Bayes' Theorem” (Harris et al. 2012, p. 316) is fulfilled in the local version. The take home message of this passage could be presented in a more simplified way:

Local *ad hominem* attack (non-fallacious)

- (1) Speaker S gives a testimony t .
- (2) Speaker's S property ϕ is a negative property with regard to trustworthiness.
- (3) Speaker S has a negative property ϕ that is relevant evidence e for a hypothesis h with regard to the content of the particular testimony t by hearer H .

(C) The testimony t , uttered by speaker S , is probably false as assessed by hearer h .¹²

5.2 Global *ad hominem* attack

I call the opposite kind of reasoning the *global ad hominem argument*, which I claim, is fallacious. The hearer commits a global attack if she does not believe the speaker in general. The hearer discredits her any kind of trustworthiness. Consider some stereotypes and prejudices that could suffice for such a radical conclusion.¹³ This behavior is clearly irrational, since it undermines any testimony of a speaker in every situation. As you can see, this kind of fallacy is neither a *content-related* nor *context-related ad hominem* argument. In distinction from the global attack, here the content does not play any role in the evaluation of the speaker's reliability. First, the evidence e only includes per-

¹² As Jacob (1987) suggests, beliefs are shared in different communities with different ideological backgrounds that are themselves constitutive of belief-formation. One could defend that the local *ad hominem* attack is an important tool for running a communicative society. If so, we would be using local *ad hominem* attacks as a form of self-deception, which would then be somehow an instance of a shared optimism bias. I will not discuss this phenomenon any further, because it is not a tool or cognitive filter that improves knowledge transmission. Hence, positive local *ad hominem* attacks, one could argue, have at least a propensity for being epistemically unvigilant mechanisms.

¹³ Yarp (2013) suggests that *ad hominem* fallacies like prejudices could be unconscious or at least not transparent to the hearer's reasoning.

sonal traits and not the circumstances of the utterance. Second, the evidence e does not influence the degree of belief that the hypothesis h is true. This is the reason why the belief, formed via the process of a global *ad hominem* fallacy, could not be calculated in *Bayes' Theorem*. The normative stipulation that evidence e should affect the probability of a hypothesis h is not satisfied.

Another reason why this type of personal attack is fallacious is because it includes irrelevant circumstances and personal traits as the basis of the speaker's evaluation (Buenting 2005). The speaker, as she is assessed, is not in any way disposed to maximizing the hearer's set of true beliefs.¹⁴ In other words, the hearer undermines any potential benefit she may gain through any of the speaker's testimony (Sperber 2001), which could be described as a paradigmatic case of epistemic injustice (Fricker 2007). Hence, a general assessment of the speaker's credibility has nothing to do with the process of positive reason formation. In other words, the global *ad hominem* attack is not an epistemic tool, because it lacks any *credibility function*, because the speaker assesses the testimony as necessarily false, because of her personal properties. Since it is not an epistemic tool, this kind of reasoning is not epistemically vigilant (Sperber et al. 2010). To summarize:

Global *ad hominem* attack (fallacious)

- (1) Speaker S gives a testimony t .
- (2) Speaker's S property φ is a negative property with regard to trustworthiness.
- (3) Speaker S has a negative property φ that is ascribed as relevant for every testimony t by hearer H .

(C) The testimony t , uttered by speaker S , is necessarily false as assessed by hearer H .

6 *Ad hominem* arguments and communicative agency

I agree with Jacob that some aspects of Grice's theory of meaning are in a broad sense incompatible with Millikan's account of communicat-

ive agency. The focus of this commentary so far has been communicative agency in epistemic contexts and its implications, and in particular the personal attack. I will now evaluate whether Millikan's account of direct perception or Grice's account of mindreading could account for *ad hominem* arguments in epistemic contexts. My answer is that the Gricean mindreading thesis is more plausible. I then compare the separability thesis with the cooperative devices. The separability thesis fits best with the practices of *ad hominem* fallacies. The presupposition of cooperative devices is less plausible.

6.1 Mindreading vs. direct perception

Recall from section 1 that the mindreading thesis relies on the twofold account of informative and communicative intention. First, the speaker has to recognize or understand the speaker's informative intention, which is the speaker's communicative intention. For the *fulfillment* of the informative intention, the trustworthiness of the speaker has to be accepted. In other words, in order to *fulfill* the informative intention, the hearer commits neither a local nor a global *ad hominem* argument (or she would not accept it). But in order to be an epistemically vigilant agent, the hearer has to make some further inferences, which are inductive (as well as the *ad hominem* fallacies). This inductive inference model involves some kind of mindreading that could affect the reliability judgment.¹⁵ Millikan claims that the acceptance of a given testimony as a source of knowledge is a form of direct perception without any kind of inference (Millikan 1984; Sperber et al. 2010). She talks about *translation* instead of inference. The hearer translates the utterance via direct perception into a new belief (Millikan 2004):

Forming a belief about where Johnny is on the basis of being told where he is I just as direct a process (and just as indirect) as forming a belief about where Johnny is on

¹⁴ I will not consider the ethical implications of this view any further in this commentary.

¹⁵ Unfortunately, I cannot address in this paper which kind of mindreading is supported by this view and how it could perhaps be related to social cognition and mirror-neurons. For a general discussion see Jacob (2008, 2013).

the basis of seeing him there. (Millikan 2004, p. 120) There is no reason to suppose that any of these ways of gaining the information that Johnny has come in requires that one perform inferences. (Millikan 2004, p. 125)

It is doubtful that these circumstances explain our everyday communicative agency, especially with regard to epistemic conversations. According to Millikan, the acceptance of a new belief does not involve any representation of the speaker's intention. But in order to assess the reader as benevolent and competent (or reliable), one has to rely—as argued in section 5—on inductive inferences which are of course derived representations manifested in beliefs about the speaker's intention.

In epistemic contexts of communication only the mind-to-world direction is involved, *qua* descriptive utterances. One criticism offered by Jacob is that Millikan's account of perception could only account for descriptive utterances, hence only for the mind-to world direction. Another issue is closely related to this kind of criticism. It concerns testimony that has very little to do with perceptual capacities. With regard to very complex utterances like (1), I agree with Jacob (this collection, p. 9) that “it does not make much sense to assume that either the speaker or her addressee could perceive what the speaker's utterance is about.” Consider the nature of testimonial reports. Even some direct perception of a testimony about some state of affairs is perceptually impoverished compared to directly perceiving the state of affairs in question. Imagine some testimonial reports that have been heard through the radio. In such a case, you are not in a perceptually close relationship to the reported state of affairs. If you evaluate the credibility of the speaker, it is very likely that you would run through different processes of inductive inference in order to commit an *ad hominem* argument or avoid one. The more abstract the testimony, the more implausible it becomes that it has anything to do with direct perception. It becomes even more complicated with complex indexical utterances or a group of different but

equally eligible interpretations of a particular testimony. Consider again example (1). Here it is very likely that a hearer represents some intentions of the speaker that are linked to her psychological states. If one representation is that the speaker could deceive the hearer in particular circumstances, the hearer will probably commit a local *ad hominem* attack. To sum up: *Ad hominem* arguments are ascriptions that result from inductive inferences that also depend on belief-desire psychology, because the hearer gains a representation of the second-order representation of the sentence expressed by the speaker. The representation of the hearer is a third-order representation of the second-order linguistic representation of the speaker (cf. Jacob 1987).¹⁶

6.2 Separability thesis vs. cooperative devices

The problems addressed so far are closely related to the separability thesis. The separability thesis is the claim that the hearer and the speaker could have different interests, which are causes of the informative intention remaining unfulfilled, because there are two cases that suggest that the interests of both parties fall apart. In the first, the hearer gains a new belief that is not true, because the speaker has the informative intention to deceive the hearer. So her informative intention has the aim that the speaker gains a false belief and not one about some states of affairs, as described in section 2. In the second, the sentence, uttered by the speaker, is true, but nonetheless denied by the hearer on the basis of an *ad hominem* argument (Sperber & Wilson 1986).¹⁷ These two cases do not support Millikan's (2005) claim that the interests of both speaker and hearer are balanced. If a hearer commits a global *ad hominem* argument, it is even harder to ascribe balanced interests to speaker and hearer. Sperber (2001) defends a plausible weak version of coincidence of interests. It is only necessary that they over-

¹⁶ According to Jacob (1987), a belief-ascription is not constitutive of the subject's belief in the first place.

¹⁷ There are, of course, plenty of other options for different interests (cf. Sperber 2001).

lap in the long term in order to establish successful practices of social knowledge transmission.

Cases of global *ad hominem* arguments could only occur if a speaker understands the informative intention which she combines with some particular personal properties of the speaker in order to reject the testimony in question. Hence the speaker succeeds in establishing the communicative intention, but fails to fulfill the informative intention. In Millikanean terms, the direct proper function of the speaker is that the hearer gains a new belief. If the hearer commits an *ad hominem* attack, the direct proper function remains unfulfilled. But the communicative intention is still fulfilled, and that is all that is required for successful communication according to the separability thesis of communicative intentions (Sperber & Wilson 1986). The hearer recognizes that the speaker wants to inform her of her informative intention, which means that the communicative intention has been fulfilled. But the informative intention—which is that the hearer gains some new information or a true belief—fails, because an *ad hominem* attack has been committed. This circumstance could be well explained with the separability thesis and the weak account of communication that we addressed in section 6.1. To sum up, and in agreement with Jacob, if an *ad hominem* attack has been committed, even a weaker version, communicative agency is violated in the Millikan (2004, 2005) framework because the cooperative conventional transmission is violated in the first place.¹⁸

The picture I draw with regard to the *ad hominem* arguments rests on the assumption that trustworthiness has to be assessed by the hearer in order to be counted as epistemic vigilant, which would be the reductionist component. On the other side, the speaker has to utter a true sentence to transmit knowledge to a speaker, which would be the anti-reductionist-component. This view is supported by Lackey's dualism. As can be seen, the establishment of a dualistic account and all its implications for the

inductive reasoning model can be better explained with the separability thesis.

7 Conclusion

In this commentary I have extended the refreshing account given by Jacob, who presents Grice's individual account of meaning and assesses its plausibility with regard to communication and knowledge transmission. I defended the view that one promising way to talk about testimony as a source of knowledge is offered by Lackey's dualism. Here, both speaker and hearer are the important in knowledge transmission. In order to secure this transmission, the speaker has to utter a true sentence and the hearer has to check the speaker's trustworthiness. I distinguished two kinds of personal attack:

(i) The local *ad hominem* argument, which is not fallacious, focuses on the proposition and personal properties of the speaker, and is a content-related, relevance-based attack based upon one's subjective probabilistic estimation of the speaker's reliability, which can be calculated in *Bayes' Theorem*.

(ii) The global *ad hominem* argument, which is fallacious, is an extreme prejudice that denies that the speaker is reliable in any case. It is not usable as a tool for knowledge transmission, because it violates the stipulation of the Bayesian argumentation in which one should include some evidence in the subjective probability estimation. This extreme kind of a personal attack could be racism or stigmatizing, for instance.

(iii) Here I argued—in agreement with Jacob—that the Gricean account of mindreading is more plausible than Millikan's account of direct perception. To use the inductive model of reasoning when evaluating a speaker's reliability, one also has to rely on the use of belief-desire psychology. It is important to *infer* the intentions of the speaker, and to think about whether she has good reasons or a general propensity to speak the truth. Millikan's framework of direct perception does not account for this, because the direct perceptibility of some

¹⁸ The question, of course, is in which sense it is violated, in detail, and how this affects Millikan's theory of language in general. However, these implications cannot be addressed here.

abstract cases of testimony and their evaluated reliability is very implausible.

(iv) Last, I argued that the description of speaker and hearer as cooperative devices is implausible, too. First, we know that the speaker could have deceptive intentions. Second, if somebody is committing a global *ad hominem* attack, the interests of the utterer and her addressee fall apart. Nonetheless, the communicative intention still holds. Both parties communicate successfully, even if the hearer does not gain a new belief. So we could conclude with Grice and Sperber and Wilson that the communicative intention is sufficient for a successful communication. If the addressee commits an *ad hominem* fallacy, the proper function is unfulfilled. But in this case the conventional speaker action, which is part of successful communication, has been violated. I have argued that the Gricean account could well explain our communicative practices regarding epistemic contexts.

In terms of future research, it would be very interesting to see, how the Gricean philosophy of a speaker's individual meaning and mindreading could be embedded in theories about social cognition, social epistemology and informal logics. Jacob has presented an illuminating account of how the Gricean philosophy could be embedded in modern philosophy of mind and the cognitive sciences. I propose that one should further reflect on Jacob's arguments and adopt his conceptual framework, which I think is very precise and explanatorily fruitful. But for my own proposal of a distinction between local and global *ad hominem* attacks, it will be important to flesh out these accounts with regard to Bayesian reasoning and argumentation in epistemic contexts. A good candidate for elaborating this kind of research is the recent account of *predictive processing*, which is also based on Bayesian probabilities.

In this commentary I have claimed that these personal attacks are inductive mechanisms. But much more could be said about their functionality or even their instantiation in a cognitive system. Then it would be interesting to see if non-human cognitive systems could commit these kinds of *ad hominem* attacks. How precise could they be in evaluating a

speaker's reliability? Are instances of *ad hominem* attacks bound to a specific type of brain through which the relevant representational and functional architectures are realized? How could such a phenomenon like the global and local *ad hominem* attack evolve in *homo sapiens*, and what are the deeper underlying cognitive mechanisms of such attacks? These questions need to be answered if we want to understand these important mechanisms and processes of social knowledge, as well as our communicative society as a whole.

Acknowledgements

First of all, I am appreciative for the illuminating target paper. In addition to that, I would like to thank the two anonymous reviewers and Thomas Metzinger and Jennifer M. Windt for their editorial reviews. They really helped to improve this paper. I am also grateful to Thomas Metzinger and Jennifer M. Windt for the opportunity to contribute to this project.

References

- Buenting, J. M. (2005). The rejection of testimony and the normative recommendation of non-fallacious 'ad hominem' arguments based on Hume's 'Of Miracles' and Canadian Law. *Auslegung*, 27 (2), 1-16.
- Clément, F. (2010). To trust or not to trust? Children's social epistemology. *Review of Philosophy and Psychology*, 1 (4), 531-549. [10.1007/s13164-010-0022-3](https://doi.org/10.1007/s13164-010-0022-3)
- Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford, UK: Oxford University Press.
- Fricker, E. (1995). Telling and trusting: Reductionism and anti-reductionism in the epistemology of testimony. *Mind*, 104 (414), 393-411.
- (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford, UK: Oxford University Press.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66 (3), 377-388. [10.2307/2182440](https://doi.org/10.2307/2182440)
- (1969). Utterer's meaning and intentions. *The Philosophical Review*, 78 (2), 147-177. [10.2307/2184179](https://doi.org/10.2307/2184179)
- Groarke, L. (2011). Informal logic. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/logic-informal/>
- Harris, A. J. L., Hsu, A. & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using ad hominem. *Thinking & Reasoning*, 18 (3), 311-343. [10.1080/13546783.2012.670753](https://doi.org/10.1080/13546783.2012.670753)
- Jacob, P. (1987). Thoughts and belief ascriptions. *Mind & Language*, 2 (4), 301-325. [10.1111/j.1468-0017.1987.tb00124.x](https://doi.org/10.1111/j.1468-0017.1987.tb00124.x)
- (2008). What do mirror neurons contribute to human social cognition? *Mind & Language*, 23 (2), 190-223. [10.1111/j.1468-0017.2007.00337.x](https://doi.org/10.1111/j.1468-0017.2007.00337.x)
- (2010). Intentionality. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/intentionality/#9>
- (2013). How from action-mirroring to intention-ascription? *Consciousness and Cognition*, 22 (3), 1132-1141. [10.1016/j.concog.2013.02.005](https://doi.org/10.1016/j.concog.2013.02.005)
- (2015). Millikan's teleosemantics and communicative agency. In T. Metzinger & J. M. Windt (Eds.) *Open MIND* (pp. 1-22). Frankfurt a. M., GER: MIND Group.
- Lackey, J. (2005). Testimony and the infant/child objection. *Philosophical Studies*, 162 (2), 163-190. [10.1007/s11098-004-7798-x](https://doi.org/10.1007/s11098-004-7798-x)
- (2006). It takes two to tango: Beyond reductionism and non-reductionism in the epistemology of testimony. In J. Lackey & E. Sosa (Eds.) *The epistemology of testimony* (pp. 160-189). Oxford, UK: Oxford University Press.
- Mascaro, O. & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, 112 (3), 367-380. [10.1016/j.cognition.2009.05.012](https://doi.org/10.1016/j.cognition.2009.05.012)
- Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.
- (1989). Biosemantics. *The Journal of Philosophy*, 86 (6), 281-297.
- (2004). *Varieties of meaning*. Cambridge, MA: MIT Press.
- (2005). *Language: A biological model*. Oxford, UK: Oxford University Press.
- Sperber, D. (2001). An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29 (1/2), 401-413. [10.5840/philtopics2001291/215](https://doi.org/10.5840/philtopics2001291/215)
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G. & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25 (4), 359-39. [10.1111/j.1468-0017.2010.01394.x](https://doi.org/10.1111/j.1468-0017.2010.01394.x)
- Sperber, D. & Wilson, D. (1986). *Relevance, communication and cognition*. Cambridge, MA: Harvard University Press.
- Van Eemeren, F. H., Meuffels, v. B. & Verburg, M. (2000). The (un)reasonableness of ad hominem fallacies. *Journal of Language and Social Psychology*, 19 (4), 416-435. [10.1177/0261927X00019004002](https://doi.org/10.1177/0261927X00019004002)
- Van Eemeren, F. H., Garrsen, B. & Meuffels, B. (2008). Reasonableness in confrontation: Empirical evidence concerning the assessment of ad hominem fallacies. In F. H. Van Eemeren & B. Garrsen (Eds.) *Controversy and confrontation: Relating controversy analysis with argumentation theory* (pp. 181-195). Amsterdam, NL: John Benjamins.
- Walton, D. (1998). *Ad hominem arguments*. Tuscaloosa, AL: The University of Alabama Press.
- (2008). *Informal logic: A pragmatic approach*. Cambridge, UK: Cambridge University Press.
- Yarp, A. (2013). Ad hominem fallacies, bias and testimony. *Argumentation*, 27 (2), 97-109. [10.1007/s10503-011-9260-5](https://doi.org/10.1007/s10503-011-9260-5)

Assessing a Speaker's Reliability Falls Short of Providing an Argument

A Reply to Marius F. Jung

Pierre Jacob

When confronted with a speaker's assertion, her addressee can either fulfill the speaker's informative intention and accept the new belief or not. If he does, he can either accept the new belief on the sole basis of the speaker's authority or not. If not, then the addressee can examine the reliability of the speaker's assertion. If he does, then he can either check the content of the speaker's assertion with the contents of his own beliefs or scrutinize the speaker herself as the source of the novel information. If the latter, then he can either examine the speaker's epistemic competence in the relevant domain of discourse or the speaker's moral benevolence (or both). None of the above processes amounts to the addressee producing an argument, let alone an *ad hominem* argument. Only if the speaker offers an argument to back her assertion could the addressee commit an *ad hominem* counter-argument in his attempt at rebutting the speaker's.

Keywords

Argument | Assessment of the reliability of a speaker's assertion | Authority | Benevolence | Competence | Fulfilling the speaker's informative intention | Knowledge

In my paper, I probed the gap between the Gricean approach and Millikan's approach to human communicative agency. In particular, I argued in favor of the Gricean separability thesis, i.e., the thesis that the process whereby an addressee fulfills an agent's communicative intention (by understanding or recognizing her informative intention) is distinct from the process whereby the addressee further fulfills (if and when he does) the agent's informative intention (by accepting either a new belief or a new desire for action). I am grateful to Marius F. Jung for his valuable comments on my paper, in which he tries to offer positive suggestions to-

wards bridging the gap between the Gricean separability thesis and (social) epistemology.

In particular, I agree with Marius F. Jung that the issues of whether and to what extent a communicative agent's testimony should or can be assessed as reliable and justified, and thereby construed as knowledge (and not as mere opinion) by her recipient, are of fundamental importance. I also agree with him that it is worthwhile to try and bridge the gap between the psychological investigation of the process whereby an addressee assesses the reliability of a speaker's testimony and the major divide between the reductionist and the anti-reduction-

Author

Pierre Jacob

jacob@ehess.fr

Institute Jean Nicod
Paris, France

Commentator

Marius F. Jung

mjung02@students.uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

ist perspectives in the epistemology of testimony. However, I still want to resist using the particular bridge (or bridges) Jung is building for me. In the following, I want to briefly explain why.

First of all, let us be clear that what we are dealing with here is the addressee's basic epistemic task of assessing the reliability of a communicative agent's (the speaker's) *testimony* or *assertion*, i.e., utterances with truth-conditional contents, because only assertions can be assessed for their reliability or believability. Only a speaker's assertions, not a speaker's requests, can directly enlarge her addressee's knowledge of the world. For the purpose of the discussion of Jung's epistemological project, we should simply ignore addressees' responses to speakers' utterances of requests, i.e., of utterances that lack truth-conditional contents. (I ignore here the fact that a speaker's request may enlarge an addressee's knowledge of the speaker's own character traits.)

Secondly, as I understand it, Jung would like to directly link the investigation of the addressee's task of assessing the reliability of a speaker's assertion to the dispute between the reductionist and the anti-reductionist perspective in the epistemology of testimony. I will reconstruct Jung's basic strategy by means of the six following assumptions.

- He construes the addressee's overall process of assessment of the reliability of a speaker's assertion as an *argument*.
- As I understand it, he also accepts Sperber et al.'s (2010) view that the overall process whereby an addressee assesses the reliability of a speaker's assertion can be divided into two component processes: the assessment of the *authority* of the speaker (who is the *source* of the testimony) and the assessment of the *content* of the speaker's assertion.
- He further focuses on the addressee's assessment of the *authority* of the speaker as the source of the testimony, at the expense of the assessment of the content of the speaker's assertion.

- He links the addressee's assessment of the authority of the speaker as the source of the testimony to *ad hominem arguments*.
- He draws a distinction between local and global *ad hominem* arguments.
- Finally, he argues that only *local*, not global, *ad hominem* arguments are valid methods whereby an addressee can assess the reliability of the speaker's assertion.

I want mainly to take issue with Jung's very first assumption: when assessing a speaker's assertion, the addressee is evaluating the reliability or believability of her utterance. He is *not* arguing with her and therefore not producing an *ad hominem argument*. (Construing the addressee's process of appraisal as an *attack* against the speaker seems far-fetched to me.) In accordance with Jung's second assumption (at least, on my reconstruction of his train of thought), the addressee's appraisal can in turn be seen as a two-fold process: the addressee can focus on either the *content* or the *source* of the speaker's utterance (or both). If the former, then the addressee's task can be construed as a *consistency* check: he checks the compatibility of the truth of the speaker's assertion with the truths of a relevant sub-set of his own beliefs. In the latter case, he scrutinizes some of the speaker's relevant moral or "personal" properties (to use Jung's own phrase). In particular, he will assess the personal authority of the speaker along two main dimensions: her epistemic competence (or knowledge) about the relevant domain of discourse and her moral honesty, i.e., her benevolence towards him.

Of course, the addressee's assessment of the speaker's reliability along these two dimensions is an inferential process, which builds on the addressee's beliefs about both the content of the speaker's assertion and the speaker's personal authority. In an informal sense, it is a reasoning process. But I want to resist the view that this process should be construed as an *argument*, let alone as an *ad hominem* argument. As Sperber et al. (2010) and Mercier & Sperber (2011) have interestingly argued (no pun intended), to *argue* is to try and cause an addressee to accept a new belief (to endorse the truth of

some proposition), by providing explicit *reasons* for it, i.e., by construing it as the *conclusion* of a set of premises from which it derives either deductively or inductively. In fact, arguments are devices used by a speaker in order to try to *overcome* her addressee's reluctance to fulfill her informative intention (i.e., his reluctance to accept a new belief in accordance with her informative intention), on the *sole* grounds of her authority. If so, then speakers (communicative agents) argue, but an addressee doesn't: an addressee *evaluates* the speaker's argument. Of course, an addressee who disagrees with a speaker's argument in favor of some proposition P can turn into a speaker and offer *counter-arguments* to try to cause his opponent to *change her mind* about the truth of P.

1 Conclusion

When a speaker makes an assertion, she commits herself to the truth of some proposition. She thereby knowingly takes the risk that her addressee examines the reliability of her assertion by either checking the content of the asserted proposition or by scrutinizing her epistemic and moral authority. The addressee's choice is to either fulfill the speaker's communicative intention or not. She can further do it on the sole ground of the speaker's authority or not. As I see it, the issue of whether the addressee could *wrong* the speaker by committing some epistemic injustice towards her cannot arise in the process whereby the addressee assesses the reliability of the speaker's mere assertion of P. It can only arise if and when the speaker offers some explicit *argument* in favor of proposition P, in the reasoning process whereby the addressee evaluates the speaker's explicit argument in favor of P, i.e., the link between P and the premises selected by the speaker to justify P. Only then could the addressee produce an *ad hominem* counter-argument (either local or global) meant to successfully or unsuccessfully rebut the speaker's argument for P.

References

- Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34 (2), 57-111.
[10.1017/S0140525X10000968](https://doi.org/10.1017/S0140525X10000968)
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, 25 (4), 359-393.
[10.1111/j.1468-0017.2010.01394.x](https://doi.org/10.1111/j.1468-0017.2010.01394.x)