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.

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, according 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 teleofunctional) 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.

- 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-

---

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.

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) interrelated 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 Milikan'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 *cooperative* 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 twofold 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-

dressee 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.

Origgi & 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]

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-

---

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.

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 co-ordination. 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 responsivity. 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 co-operative 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

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

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

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

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

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

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language, 25* (2), 141-168. 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

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

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

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
——— (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
——— (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

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.
——— (1969). Utterer's meaning and intentions. *The Philosophical Review, 78* (2), 147-177. 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

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

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

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

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

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

Mayr, E. (1961). Cause and effect in biology. *Science*, *134* (3489), 1501-1506. 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

——— (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

Origgi, G. & Sperber, D. (2000). Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (Eds.) *Evolution and the human mind: Language, modularity and social cognition* (pp. 140-169). Cambridge, 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

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

——— (2007). Millikan's theory of signs. *Philosophy and Phenomenological Research*, *75* (3), 674-681. 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

Senju, A., Southgate, V., Snape, C., Leonard, M. & Csibra, G. (2011). Do 18-mont-olds really attribute mental states to others? A critical test. *Psychological Science*, *22* (7), 878-880. 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

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

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.

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

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

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

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

Warneken, F. & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311* (5765), 1301-1303. 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

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

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

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

Wright, L. (1973). Functions. *The Philosophical Review*, *82* (2), 139-168.