Pierre Jacob

Why Verbal Understanding is Unlikely to be an Extended Form of Perception

Abstract: Millikan's teleosemantic approach constitutes a powerful framework for what evolutionary biologists call an "ultimate" (as opposed to a "proximate") explanation of the continued reproduction and proliferation of intentional conventional linguistic signs. It thereby aims at explaining the stability of human verbal ostensive communication. This evolutionary approach needs to be complemented by particular proximate psychological mechanisms. Millikan rejects the kind of mentalistic psychological mechanisms posited by the Gricean tradition in pragmatics, according to which the task of the hearer is to recognize the speaker's intentions. Instead Millikan has persistently argued that verbal understanding is an extended form of perception. My paper is a critical assessment of Millikan's thesis that verbal understanding of a speaker's utterance enables the hearer to perceive whatever the speaker's utterance is about. I argue that Millikan's thesis rests on two fundamental assumptions. First, Millikan's notion of extended perception of the world is itself an extension of her semiotic approach according to which the process of ordinary perception (in humans and non-human animals) involves the translation of what she calls locally recurrent natural signs. Secondly, Millikan argues that only humans have the further capacity for flexible extended perception of what she calls detached signs, as opposed to attached signs or location-reflexive signs.

Keywords: attached (vs. detached) sign, direct (vs. derived) proper function, natural sign, intentional representation, extended perception, translation, flexible perception.

In this paper, I want to probe Millikan's provocative thesis that human verbal understanding is just direct perception of the world being spoken of by the speaker. As she puts it repeatedly, "during normal conversation, it is not language that is most directly perceived by the hearer, but rather the world that is perceived through language" (Millikan, 2012) or else "understanding language is simply another form of sensory perception of the world" (Millikan, 2004).[1] Of

1 "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. Distal states of affairs are

https://doi.org/10.1515/[eISBN_pdf]

course, not all human speech acts are *descriptive* or have a mind-to-world direction of fit. If a speaker utters a *prescriptive* request for action with a world-to-mind direction of fit, then the hearer's task is to form a desire or an intention to perform an action, not a perceptual task. Millikan's perception thesis narrowly construed should be restricted to a hearer's verbal understanding of descriptive utterances. In what follows, I will only consider a recipient's response to a speaker's testimony at the expense of a speaker's request.

Millikan's thesis directly challenges three *mentalistic* tenets of the approach to human ostensive communicative actions inspired by the Gricean pragmatic tradition:[2]

(1) A hearer's first task is to recognize the speaker's *communicative* intention, namely her higher-order intention that he recognizes her lower-order informative intention that she wants him (she intends to cause him) to acquire a new belief.[3]

(2) Secondly, the hearer's contribution to the success of the speaker's act involves two separable psychological steps (or processes): *understanding* and *acceptance*. The hearer understands the speaker's utterance if (and only if) he fulfills the speaker's communicative intention and thereby recognizes her informative intention. To accept the speaker's testimony is to further fulfill the speaker's informative intention, which the hearer will do only if he takes the speaker to be sufficiently trustworthy or competent on the topic at hand. Thus, the hearer's understanding of the speaker's testimony is *not* a sufficient condition for his acceptance, i.e. for his endorsement of the new belief that the speaker wishes him to accept.[4]

(3) Thirdly, the interests of the speaker and the hearer of a speaker's testimony overlap to a large extent, but they are *not* strictly identical. Furthermore, they clearly face different options. While the speaker faces the basic choice between speaking *truthfully* or not, the hearer faces the basic choice between *trusting* the speaker or not (cf. Sperber, 2001). The speaker, whose main goal is to cause the hearer to accept a new belief, will be generally better off if her hearer trusts her rather than not, whether she is truthful or not. But the hearer, whose

---

perceived through speech sounds just as they may be perceived, for example, through the medium of structured light during normal vision." (Millikan, 2005, p. 207).

2 In the following, I spell out the mentalistic assumptions characteristic of the Gricean tradition in the terminology of Sperber and Wilson's (1986) relevance-based framework. For Grice's own views, see Grice (1957, 1969, 1989).

3 For purposes of clarification, I refer to the speaker as 'she' and the hearer as 'he'.

4 This second ingredient of what I broadly call the Gricean tradition in pragmatics is spelled out in greater detail in Sperber et al. (2010).

goal is to receive truthful information from the speaker, will be generally better off if the speaker is truthful rather than not, whether he is trustful or not.

I accept the mentalistic framework inspired by the Gricean pragmatic tradition and I wish to examine Millikan's challenge.

In the language of evolutionary biologists and philosophers of biology,[5] Millikan's thesis that verbal understanding of a speaker's testimony consists in the hearer's perception of what the speaker's testimony is about can be construed as a purported *proximal* psychological mechanism, whereby human recipients contribute to the success of human verbal communicative actions. Millikan's proposal is that a speaker's communicative action is successful if and only if the hearer can *perceive* the state of affairs the speaker is talking about. This purported proximal mechanism is meant to supplement Millikan's teleosemantic approach to the continued proliferation of conventional linguistic signs, which in turn can be construed as a potential *ultimate* explanation of human verbal communication. In my opinion, Millikan's teleosemantic account of the re-production of conventional linguistic signs is one of the outstanding landmarks of naturalistic philosophy of mind and language of our time. It purports to answer such questions as: Why do humans engage in verbal communication at all? What are the biological and/or the cultural functions of human verbal communication? What selectional advantages does the capacity to perform verbal communicative actions confer to human agents and recipients?

This paper is comprised of four main sections followed by a short concluding section. In the first section, I briefly sketch Millikan's insightful teleosemantic account of the continued proliferation of human conventional linguistic signs, which in turn stands as the background to her thesis that verbal understanding is an extended perception of the world. As I explain in the second section, much of Millikan's argument for the view that verbal understanding is an extended form of perception involves three crucial assumptions, the first of which is that both ordinary perception and verbal communication rest on processes of *translation*. Her second assumption is that the relation of being a *natural sign* of something is *transitive*. Her third heterodox assumption is that, not just intentional signs but natural signs as well, may have *constituent structure*. In the third and penultimate section, I examine her responses to an obvious pair of objections to her thesis that verbal understanding is an extended form of perception. Finally, in the fourth main section, I examine a pair of arguments used by Millikan to undermine the mentalistic picture of verbal understanding inherited from the Gricean tradition.

---

5 Cf. Scott-Philipps et al. (2011).

1

Millikan's teleosemantic account rests on her adoption of a cooperative sender-receiver framework, according to which a sign can be an *intentional representation R* only if it is a *relatum* in a three-place relation involving two cooperative mechanisms, one of which (the *sender* mechanism) produces *R*, and the other of which (the *receiver* mechanism) consumes *R*.[6] The producer and the consumer mechanisms are further taken to have co-evolved so that the *Normal* conditions for the performance of the *proper* function of one are also parts of the *Normal* conditions for the performance of the *proper* function of the other.[7] On this approach, an intentional representation *R* has a *derived* function, i.e. derivative from the respective proper functions of the sender (or producer) and the receiver (or consumer). In short, the derived function of an *inner* (*mental*) intentional representation is to achieve the *coordination* between the producer and the consumer mechanisms when they are located *within* the brain of a *single* organism. The derived function of a *linguistic* (*conventional*) intentional representation is to achieve the *coordination* between the producer and the consumer mechanisms when they are located in the brains of *distinct* organisms (Millikan, 1984; 2004). Although the derived function of an intentional representation (whether mental or conventional) is to achieve the coordination between the producer and the consumer mechanisms, Millikan takes the content of an intentional representation to reflect primarily the needs of the consumer at the expense of the capacities of the producer.[8]

Arguably there are two relevant differences between the proliferation of inner (mental or non-conventional) representations and the proliferation of linguistic (conventional or non-mental) representations, the first of which is perhaps underestimated by Millikan. First, only speakers and hearers (i.e. distinct individuals with distinct brains) face the choice between being respectively truthful or not and trustful or not. A producer mechanism does not face the choice between deceiving or not the consumer mechanism if and when the two mechanisms belong to a single brain. Nor does the consumer mechanism face the choice between trusting or not the producer mechanism when both mechanisms belong to a single brain.

---

6 As I will explain soon, unlike an intentional representation, a natural sign is not a *relatum* of a three-place relation.

7 On Millikan's (1984) etiological approach, the proper function of a mechanism is one of its selected effects.

8 Cf. Neander (1995; 2017), Pietroski (1992), Jacob (1997) for critical discussions.

Secondly, Millikan has offered a powerful two-tiered naturalistic account of *conventional patterns*, which of course does not apply to the reproduction of inner mental representations. A pattern is conventional if it is the output of a continued process of reproduction (or replication). What makes it conventional (and to a large extent arbitrary) is that the reproduction is "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).

Given this framework, conventional linguistic forms turn out to be 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.[9] But like any other tool, in addition to the *direct memetic* (or 'stabilizing') proper function of its *type*, which explains the continued reproduction of its tokens, a particular *token* of some public language form may also have a *derived* function or purpose, i.e. derived from the purpose (or intention) of the speaker who produced it at a particular place and time. The direct or memetic purpose and the speaker's derived purpose may or not coincide (cf. Millikan, 1984, 2004, 2005 and Jacob, 2016 for further discussion).

2

Millikan's thesis that understanding another's testimony is an extended perception of the world is itself an extension of her *semiotic* approach to the ordinary perception of what she calls *locally recurrent natural signs* (Millikan, 2004). Unlike an inner or a conventional representation, a natural sign lacks a *function*. It carries information about what it is reliably correlated within a highly restricted (i.e. *local*) spatial and temporal *domain*. For example, in one geographical area, tracks made by quail are locally recurrent natural signs of quail. But in a neighboring spatial area, visually indistinguishable tracks made by pheasants are locally recurrent natural signs of pheasants, not of quail (Millikan, 2004, p. 32). Locally recurrent natural signs also carry information over limited *temporal* domains. For example, the position of the needle of a gas gauge in a particular car is a locally recurrent natural sign of the amount of gas in the same car "from

9 As I noticed earlier, overlapping interests are not identical interests.

the time it is initially installed until the time it first breaks down" (Millikan, 2004, p. 51).

Ordinary visual perception is the process of interpreting locally recurrent natural signs by *tracking* their informational source through a semiotic cascade generated by the *transitivity* of the relation of being a locally recurrent natural sign of a state of affairs over a restricted spatial and temporal domain. For example, retinal patterns on a human (or non-human) eye can be locally recurrent natural signs of distinctive shapes, colors and textures at a location, which in turn are locally recurrent natural signs of earlier fresh goose droppings at this location, which in turn are locally recurrent natural signs of geese flying over this location at the time of the droppings, which in turn are locally recurrent natural signs of upcoming Winter into this location (Millikan, 2004, pp. 54-55). As Millikan (2004, p. 55) puts it, in virtue of its transitivity, the natural sign relation can be "interpreted at any level of embedding or at more than one level of embedding."

On Millikan's teleosemantic account of the proliferation of conventional linguistic signs, the task of the hearer of a verbal communicative act is parallel to a perceptual task to the following extent: the hearer must *track* the correct memetic family (lineage or type) to which a particular conventional sign (e.g. 'clear' or 'the dog') belongs. Thus, both ordinary perception and verbal understanding turn out to be processes of *translation* (*not inference*). While the former is guided by the capacity to track the spatial and temporal domain over which a natural sign can be deemed to be locally recurrent, the latter is guided by the capacity to track the spatial and temporal domain of the memetic family or type of conventional signs to which a particular token belongs.

Arguably Millikan's assumption that perception is a process of *translation* sheds light on her puzzling statement that "the perceptual and cognitive systems of every animal are deeply dependent on the local information found both in the environment and within the organism itself. Without information there could not be any intentional signs or intentional information" (Millikan, 2004, p. 32). This is puzzling in light of Millikan's repeated rejection of informational teleosemantics on the grounds that it could not be the etiological function of an intentional representation to carry information (cf. Millikan, 1989a; 2004; 2013). What she has in mind is not that it is the function of an intentional representation to carry information, but instead that it is the task of perception (and cognition) to build intentional representations by means of translating natural signs.

Clearly, whether translation counts as an inferential process or not, the idea of translation paradigmatically applies to the interpretive process whereby a structured sequence of *conventional* signs uttered by a speaker for the purpose of expressing her thought is being mapped onto another structured sequence of

conventional signs. The types of conventional signs used by the speaker belong to one natural language. The complex meaning of the sequence of conventional signs uttered by the speaker depends on the meanings of its constituents and the syntactic rules of combination. The speaker uses the complex compositional meaning of the sequence of conventional signs from her language to express the propositional content of her thought. The translation process maps the sequence of conventional signs used by the speaker onto a sequence of distinct conventional signs whose types belong to a different natural language. The mapping is expected to preserve enough of the complex meaning of the sequence of conventional signs used by the speaker so that the translation can count as an alternative expression of the propositional content of the speaker's thought. Upon understanding the complex meaning of the translation of the speaker's initial utterance, the hearer is likely to entertain a thought that appropriately resembles the speaker's own thought.

Millikan construes Normal verbal communication between a sender and a receiver as a two-step translation process involving one and the same structured sequence of conventional signs produced by the speaker. First, the speaker translates her belief into a sentential conventional sign, which is uttered by the speaker. Secondly, the hearer (who speaks the same language as the speaker) translates the content of the speaker's utterance of conventional signs into his own new belief. By proposing to assimilate perceptual processes to processes of translation, Millikan means to reject an *inferential* model of perception, whereby she seems to assimilate inferential processes and processes of deliberate reasoning from explicitly entertained premises to conclusions via explicitly known rules of inference. As Millikan (2017, p. 186) puts it, "reading a sign does not require understanding why it corresponds to its signified but only how it corresponds."

In other words, when Millikan (2017, pp. 185-186) rejects the inferential approach to visual perception, what she really objects to is the claim that successful visual perception requires the ability to understand and reason explicitly about the causal mechanisms of visual processing: the light being reflected by a distal stimulus hits the retina where it is converted into electrical impulses. Visual information is carried from the retina through the optic nerve to various dedicated areas of the visual cortex, where it is ultimately transformed into a unified visual percept. Thus, it looks as if Millikan takes this perceptual process to be a process of translation, not inference, because she assimilates inference to conscious reasoning (or even theorizing) from explicit premises to explicit conclusion via explicitly known rules of inference. In a sense, Millikan's skeptical attitude with respect to the role of inferential (or computational) processes in perceptual psychology goes hand in hand with her skeptical attitude with respect to the role

assigned to mindreading (i.e. the attribution of mental states) in verbal understanding by the neo-Gricean tradition.

Millikan takes perceptual processes to be translation processes unfolding within a single brain and mapping locally recurrent natural signs onto inner mental representations of what the signs are signs of. Not only does the transitivity of the relation of being a natural sign of something makes natural signs interpretable at any level of embedding, but locally recurrent natural signs also have an unexpected feature that Millikan takes to support her thesis that perceptual processes are translation processes. While natural signs (unlike conventional signs) have no function, like conventional linguistic signs, they have *constituent structure*: they exhibit significant variables (or determinables). For example, if tracks in the mud are locally recurrent natural signs, then not only will they be locally recurrent natural signs of e.g. pheasants (not quail), at a determinate location and a determinate time, but the size of the tracks will further be a natural sign of the size of the pheasants that caused them and the distance between the tracks will also be a natural sign of how fast the pheasants were moving (Millikan, 2004, pp. 47-48). Mapping chains of locally recurrent natural signs onto a perceptual representation of some distal state of affairs may involve filling in a determinate value for the significant variables or determinables that are parts of the constituent structure of the natural signs. By stressing the constituent structure of natural signs, Millikan intends to close much of the gap between natural signs and intentional representations and to hereby pave the way for her semiotic thesis that both perception and verbal understanding are processes of translation.

On Millikan's approach, an organism's sensory (e.g. visual) mechanisms could not efficiently translate deeply embedded natural signs into ordinary perceptual intentional representations of distal states of affairs unless the organism were able to reliably *track* the spatial and temporal domains over which the relevant natural signs are indeed locally recurrent natural signs of what they are signs of. What makes tracking these spatial and temporal domains reliable in turn is the invariance of the physical, chemical and neurophysiological laws governing the sensory mechanisms of animals (including humans) endowed with visual perceptual capacities on the surface of the Earth. Light is reflected by the surface of objects onto the retina of the perceiver's eye where the energy of photons is converted into electrical impulses, which are carried to the visual cortex via the optic nerve and so on. Thus, the capacity to reliably track the spatial and temporal domains of natural signs relevant for ordinary perception is likely to be built in the brains of humans and non-human animals by biological evolution by natural selection.

Tracking the memetic family of types of linguistic signs, however, is an entirely different matter. Humans at the surface of the Earth speak different languages comprised of different memetic families of conventional signs that are not natural signs. According to Millikan, linguistic signs have proliferated via processes of *conventional* re-production. Thus, the reliability of the human capacity to track the memetic family of types of conventional signs cannot be taken to reflect the invariance of deep physical, chemical and neurophysiological laws. What Chomsky (2000) calls "universal grammar" may have been built into the brains of humans by evolution by natural selection, but what Millikan calls the ability to track the memetic families of types of conventional signs used by different human groups in different geographical areas cannot be built into the brains of humans as a result of evolution by natural selection. For example, tacit knowledge of universal grammar by itself cannot provide children with the knowledge that what is called "dog" in one language community is called "chien" in the next.

3

On the face of it, Millikan's thesis that verbal understanding is a form of perception faces an obvious pair of objections. There is a major difference between the content of a visual representation of some fact or actual state of affairs and the verbal understanding of the content of another's testimony describing the very same state of affairs. At an appropriate distance and in good lighting conditions, one could not see a cup resting on a table without also seeing its shape, size, color, texture, orientation and spatial location with respect to the table, to any other object resting on the table, and especially to oneself. Moreover, the spatial relation of a perceived object to the self is likely to change over time as one moves with respect to the object and it must be updated especially if one uses the visual information for acting on a seen object. However, if an addressee located in the next room understands the content of the speaker's utterance of the sentence "There is a cup on the table," he may endorse the belief that there is a cup on some salient table without having any definite expectation about the shape, size, color, texture, orientation and spatial location of the cup with respect to the table or anything else, and especially himself. Nor does the addressee need update his understanding of what the speaker said about there being a cup in the next room as he moves around inside his own room.

Furthermore, unlike the content of an individual's perceptual experience, the content of an agent's testimony is not restricted to objects, events and properties with which the hearer stands in a direct causal relation. It is precisely a distinctive purpose of human testimony that it enables a hearer to learn about things that are *not* directly observable by him because either he is not at the right place at the right time or else he simply could not perceive them at all. For example, numerals can, but numbers cannot, be perceived at all. A speaker's testimony can be about e.g. abstract numbers or theoretical entities posited by scientific theories. Numbers are not observable at all. Theoretical entities posited by scientific theories (e.g. quarks) can only be indirectly tested through long chains of reasoning and complex measuring instruments, neither of which might be accessible to the hearer. The point here is that some topics that are not open to an individual's perceptual experience can be conveyed by speaker's verbal testimony for the benefit of a recipient.

One interesting way Millikan proposes to bridge the gap between perception and verbal understanding is by arguing that humans have a distinctive capacity for flexibly *perceiving* things and events *without* encoding their direct spatial and temporal relations to the self, i.e. to the spatial and temporal location of the perceiver's own body. Millikan (2017) deals with this putatively distinctive human capacity for flexible extended perception in terms of what she insightfully calls "detached signs," as opposed to "attached signs" or "location-reflexive" signs. Ordinary perception (available to both humans and non-human animals) involves the translation of "location-reflexive" (or attached) natural signs into intentional representations, where the attached or location-reflexive signs carry information about the spatial and temporal relation between what they are signs of and the perceiver's own spatial and temporal context. For example, when the surface of an object reflects light onto an animal's retina, the retinal image is a location-reflexive or attached sign: it carries information about the relation between the spatial and temporal location of the reflecting surface and the spatial and temporal location of the perceiver's own body. This is why an inner perceptual representation that is mapped by translation from a chain of attached (or location-reflexive) signs can be a representation of an *affordance*, i.e. why it can guide the perceiver's *action*.

By contrast, on Millikan's (2017) account, flexible perception by verbal understanding is typically translation of *detached* signs, where a detached natural sign does *not* carry information about the relation between the spatial and temporal location of the things or event that the sign is about and the spatial and temporal location of the hearer's own body. Millikan (2004, 2017) further argues

that there are cases of typically human flexible perception other than verbal understanding, whereby humans can perceive events by translating detached signs that fail to carry information about the relation between the spatial and temporal location of the signed events and the spatial and temporal location of the perceiver's own body.

For example, humans can visually derive information from the image of an object that is reflected by a mirror. Light reflected from the mirror to the perceiver's eye directly carries information about the relation between the spatial location of the mirror and that of the perceiver's body. But it carries only *indirect* information about the relation between the spatial location of the reflected object and that of the perceiver's body. Nonetheless a human agent can learn to use information derived from an image of herself reflected by a mirror for the purpose of combing her hair or shaving (Millikan, 2004, p. 122). A human driver can also learn to use information derived from an image of a blue car that was initially seen through the rear-view mirror of her car to re-identify the blue car that just passed her own car on the left and is now moving in front of her own car (*Ibid*., p. 132). Similarly, humans can visually extract from a photograph the mental representation that there once existed sometime and somewhere or other e.g. an apple looking so and so, while the relation between the spatial and temporal location of the apple and the spatial and temporal location of the perceiver's body is not encoded at all by the photograph.[10] Humans can also see events depicted on a television screen: the relation between the spatial and temporal location of the depicted events and that of the perceiver's body is not encoded either by the pictures on television. In a nutshell, images reflected by a mirror, photographs and pictures on television are what Millikan (2017) calls *detached* natural signs, not attached (or location-reflexive) natural signs.

Millikan takes it that ordinary perception, which is based on the capacity to track the relevant domains of *attached* signs, is shared by human and non-human animals. However, Millikan (2004, pp. 122-124) strongly suggests that unlike ordinary perception, flexible extended perception is like verbal understanding in being uniquely human: both extended perception and verbal understanding rest not only on "a marvellous flexibility in accommodating new semantic functions, but also [on] the capacity mentally to represent... information that does not include the relations to you of the things the information is about." Millikan (2004,

10 Arguably, non-human animals can discriminate pictures of living entities from pictures of non-living entities, but Millikan would likely argue that this is part of ordinary non-flexible perception.

pp. 122-124) further assumes that flexible perception, including verbal understanding, is no more a mentalistic task than ordinary perception is. In other word, the hearer of another's testimony must be able in *all* cases to track the relevant appropriate memetic family of the type of conventional linguistic signs used by the speaker *without* representing any of her psychological states.

The challenges for this assumption seem, however, quite overwhelming. For example, Millikan claims that there are many ways to flexibly recognize rain, all of which are perceptual. There is a way that rain feels on one's skin and a way it looks when one sees it fall out the window. There are distinct ways it sounds when falling either on the rooftop or on the ground. There is still another way it sounds when falling on English speakers: 'It's raining!' But this last statement cannot be strictly true for this English sentence could clearly be used by an English speaker who is talking about a raining event somewhere in the universe where no English speaker is present. Similarly, it is quite unclear how a hearer could understand what a speaker means by her utterance of the definite description 'the dog' or some universally quantified sentence e.g. 'Everybody is asleep' or the possessive 'John's book' without representing the speaker's beliefs and intentions. Only if the speaker's recipient is part of a narrow circle of well-known relatives (e.g. one's spouse and/or children) could the contextual common ground for such utterances be taken for granted by the speaker without requiring the hearer to represent the speaker's beliefs and intentions. However, human verbal communication is not so restricted to a narrow circle of in-group members. A speaker's use of these linguistic forms (possessives, definite descriptions or universally quantified expressions) is not restricted to a narrow circle of in-group recipients. When a speaker uses such expressions, her recipients may be foreigners and members of cultural communities far different from hers. It is likely that in many cases, a recipient could only make sense of the speaker's utterance involving such expressions if he was able to retrieve the speaker's communicative intention. Thus, Millikan's picture of verbal understanding seems to rest on a *parochial* view of human verbal communication, where common grounds between a speaker and a recipient can be taken for granted.

4

I now turn to two separate arguments used by Millikan to undermine the neo-Gricean mentalistic picture of human verbal communication, the first of which

rests on what it takes to achieve a task of flexible extended perception. Her second argument rests on findings from the experimental psychological (especially developmental) investigation of false-belief understanding.

Millikan's first argument rests on a questionable analogy between ordinary and flexible extended perception. One stage of a rabbit's ordinary visual perception of a fox involves the processing of retinal images of the fox, which are natural signs of the fox. Seeing the fox, however, does not require the rabbit to form intentional representations *of* her retinal images. Millikan makes a somewhat similar claim about flexible extended perception. For example, she claims that the successful use of binoculars or televisions and the extraction of abstract information from photographs do not require a flexible perceiver to "understand the innards" of binoculars, televisions or cameras. She concludes that a hearer need not either represent the speaker's beliefs and intentions in order to perceive what the speaker's testimony is about.

There is arguably a stronger and a weaker version of Millikan's premise that successful flexible extended perception does not require any understanding of the relevant visual tools. On the stronger reading, Millikan denies that an individual must have scientific knowledge of the natural laws underlying its construction for successfully enrolling a visual tool in a task of flexible extended perception. On the weaker reading, she denies that an individual could successfully enroll a visual tool (such as binoculars) *without* making *any* assumption or other about them.

While I agree with Millikan's rejection of the stronger claim, I don't think that she can easily reject the weaker claim. A human perceiver could hardly calibrate the bilateral acuity of her binoculars, switch her television on or extract relevant abstract information from a photograph without making some assumptions or other about the function of binoculars, televisions and cameras. For example, a human perceiver could not switch a TV on unless he or she assumed that it is an electrical appliance. He or she could not calibrate the acuity of binoculars unless he or she assumed that it is an optical device whose function is to enhance the visual perception of objects at a distance. Nor could he or she extract from a photograph the information that there once existed sometime and somewhere or other an apple looking so and so unless he or she knew something about how a camera can be used to take a picture of an apple. Arguably, what the weaker claim highlights is that flexible extended perception rests on human cultural learning, which in turn rests on human mindreading capacities, in accordance with the neo-Gricean mentalistic picture of human verbal communication. If so, then acceptance of the weaker claim about flexible extended perception would seem to vindicate, not to undermine, the neo-Gricean picture.

Finally, I turn to Millikan's repeated thesis that experimental developmental evidence shows that young children reach proficiency in tasks of verbal communication much before they are able to read others' minds, which she also takes to undermine the neo-Gricean assumptions. As she nicely puts it, "the infants learn what kitties look like in various postures, what they feel like, the sounds they make and what they sound like through language. There seems no reason why this last would require that the infants employ a theory of mind or concepts of mental states." As Millikan recognizes, human adults and older children have the reflective resources needed to represent a speaker's mental states. However, if infants don't need to represent a speaker's beliefs and intentions, then neither does an adult hearer in a normal conversation. Only if the normal flow of verbal communicative information breaks down and the meaning of the speaker's utterance is puzzling for one reason or another must an adult hearer reflect upon the speaker's beliefs and intentions.

Millikan's two-pronged strategy against the neo-Gricean picture rests on two intimately related assumptions about mindreading (or theory of mind), the first of which is that the experimental developmental investigation of false-belief understanding has established that young children still lack a theory of mind (or the ability to read others' minds) when they are already proficient in tasks of verbal communication. Her second related assumption is that for human adults, mindreading is a demanding (or effortful) cognitive task, which requires reflective thinking.

Taking Millikan's second assumption first, it may seem as if what makes her diagnosis that mindreading, unlike perception (including extended flexible perception), is effortful for human adults is that mindreading has a *metarepresentational* architecture. However, the fact that the best scientific (or proto-scientific) characterization of some cognitive capacity has a complex structure does not entail that the use of this cognitive capacity is effortful. For example, the best current scientific characterization of human vision is complex. But this does not entail that visual processing is effortful for humans.

In fact, some empirical evidence suggests that human adults perform tasks of mindreading, if not automatically, at least spontaneously. For example, in the context of a psychophysical study, Kovács (2010) and colleagues found evidence that adults automatically compute the content of a protagonist's false expectation about the presence of a ball behind an occluder on a computer screen, even if the protagonist is a blue smurf and his expectation is irrelevant to their psychophysical task, which is to press a button as fast as possible if they detect the ball behind the occluder. In a (2010) study on Level 1 visual perspective-taking in

adults, Samson and colleagues also found that participants automatically compute the number of dots that an avatar can see on the walls of a room, as shown by the fact that they were slower to respond and made more mistakes about the number of dots that *they themselves* could see, if the number of dots that they could see was different from, rather than equal to, the number of dots that the avatar could see.

I finally turn to Millikan's appeal to the developmental investigation of false-belief understanding in human childhood. Millikan appeals to evidence based on explicit change-of-location false-belief tasks, in which participants who know the location of a mistaken agent's toy are asked to predict where the mistaken agent will look for her toy. This evidence shows that most 3-year-olds fail and incorrectly point to the toy's actual location. Not until they are 4,5 years of age do most children pass these tests (cf. Wimmer and Perner, 1983; Baron Cohen et al., 1985; Wellman et al., 2001). However, much recent evidence based on implicit change-of-location false-belief tasks, in which participants are not asked any question, also shows that preverbal infants expect an agent to act in accordance with the content of her (true or false) belief (Onishi and Baillargeon, 2005; Baillargeon et al., 2010). The puzzle is: how to reconcile these discrepant findings?

So far as I am aware, everything Millikan has written on this topic suggests that she accepts the assumption that only success on explicit false-belief tasks can demonstrate false-belief understanding. I think, however, that this assumption is demonstrably false. If only success on explicit false-belief tasks could demonstrate false-belief understanding, then conversely false-belief understanding should be sufficient for success on explicit false-belief tasks. But this is clearly not true: a monolingual adult Russian speaker might fail to answer the canonical English sentence "Where will Sally look for her toy?" and still be fully able to ascribe false beliefs to others.

Suppose on the contrary that we take the findings based on implicit false-belief tests at face value as showing that preverbal infants can represent the contents of others' false beliefs. The question then is: why do most 3-year-olds find explicit false-belief tasks so challenging? Here is one possible pragmatic answer. In order to correctly *predict* where a mistaken agent will look for her toy, it is sufficient to know where she last placed it. In typical explicit false-belief tasks, however, participants are provided by the experimenter with much information about the relocation of the toy and its actual location, which is strictly irrelevant to the prediction task. Preschoolers should simply ignore this irrelevant information and focus on the only relevant information, namely where the mistaken agent last placed her toy. But they are likely to find it difficult to ignore irrelevant information provided by an adult that looks like a benevolent and competent speaker.

Thus, one possible way preschoolers might make this irrelevant information relevant is if they turn the experimenter's *prediction* question "Where *will* Sally look for her toy?" into the *normative* question "Where *should* Sally look for her toy?" If they do, then the correct answer to this normative question is the toy's actual location, which is exactly where most preschoolers point in answer to the experimenter's question. If this explanation of the findings based on explicit false-belief tasks is on the right track, then preschoolers' failure in these tasks is consistent with preschoolers' (and also infants') ability to ascribe false belief to others.

5

As I have argued in the first section of this paper, Millikan's major teleosemantic contribution has been to open an entirely novel approach to the continued reproduction of intentional conventional public-language signs. This contribution is best construed as a potential *ultimate* explanation of human ostensive verbal communication, i.e. as an answer to the question "Why do humans engage in ostensive verbal communicative actions?" Millikan's teleosemantic approach to the proliferation of intentional conventional public-language signs must be supplemented by some *proximate* psychological mechanism. Millikan proposes that the basic proximate psychological mechanism whereby a human recipient understands and accepts an agent's verbal testimony is *perception*.

As I have argued in the second section of this paper, Millikan makes three fundamental assumptions about ordinary perception. She assumes that ordinary perception is a process of translation mapping natural signs onto intentional representations. She further assumes that the relation of being a natural sign is transitive. Finally, she assumes that natural signs, just like intentional representations, also have constituent structure. In a nutshell, Millikan's semiotic approach to perception paves the way for her thesis that verbal understanding is an extended form of perception.

As I have explained in the third section of the paper, in response to obvious objections to her thesis that verbal understanding is ordinary perception, Millikan argues that humans have a distinctive capacity for flexible extended perception that enables them to flexibly perceive things and events (e.g. on a television screen) without encoding their direct spatial and temporal relations to the self, i.e. to the spatial and temporal location of the perceiver's own body. In particular, Millikan claims that successful flexible extended perception through visual tools

(e.g. binoculars) does not require the flexible perceiver to have any understanding of the relevant visual tools. By parity of reasoning, she further claims that the success of verbal understanding does not either require a recipient to recognize a speaker's informative intention, let alone to fulfill it.

Finally, Millikan has repeatedly argued against the mentalistic Gricean approach to human verbal understanding that the developmental evidence about false-belief understanding shows that human children can reach proficiency in tasks of verbal communication much before they are able to read others' minds. In the fourth section of the paper, I have distinguished between a stronger and a weaker reading of her claim that successful flexible extended perception through visual tools (e.g. binoculars) does not require the flexible perceiver to have any understanding of the relevant visual tools. I have argued that she cannot easily reject the weaker version of the claim that successful flexible perception requires some assumptions about the function of visual tools. I have also taken issue with Millikan's interpretation of the developmental evidence about false-belief understanding and I have argued that the failure of preschoolers on explicit false-belief tasks is compatible with their mentalistic capacity to attribute false beliefs to others. My overall conclusion is that verbal understanding of a speaker's testimony is unlikely to be an extended form of perception of what the speaker is talking about.[11]

References

Baillargeon, R., Scott, R. M., & He, Z. (2010) False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110-118.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985) Does the autistic child have a "theory of mind?" *Cognition*, 21 (37), 46.

Chomsky, N. (2000) *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Grice, H.P. (1957) Meaning. *Philosophical Review*, 66, 377–88.

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

11 I am grateful to Christoph Limbeck-Lilienau for inviting me to the 40th International Wittgenstein Symposium "The Philosophy of Perception and Observation" as well as for his comments on a draft of this paper, and to members of the audience for their questions and comments. I acknowledge support from the European Research Council (ERC) under the European Union Seventh Programme (FP/2007-2013)/ERC Grant 609819.

Grice, H.P. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
Jacob, P. (1997) *What Minds Can Do*. Cambridge: Cambridge University Press.
Jacob, P. (2016) Millikan's teleosemantics and communicative agency. In Metzinger, T. & Windt, J.M. (eds.) *Open Mind, Philosophy and the Mind Sciences in the 21st Century*. Cambridge, MA: MIT Press (pp. 767-788).
Kovács, Á., Téglás, E., and Endress, A. (2010) The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330, 1830-1834.
Millikan, R.G. (1984) *Language, Thought and Other Biological Categories*. Cambridge, MMA: MIT Press.
Millikan, R.G. (1989a) Biosemantics. *The Journal of Philosophy*, 86, 281-197.
Millikan, R.G. (1989b) In defense of proper functions. *Philosophy of Science*, 56, 2, 288-302.
Millikan, R.G. (1990) Seismograph reading for "Explaining Behavior." *Philosophy and Phenomenological Research*, 50 (4), 807-812.
Millikan, R. (1995) Pushmi-Pullyu Representations, *Philosophical Perspectives*, 9:185-200.
Millikan, R.G. (1998) Conventions made simple. *The Journal of Philosophy*, 95(4), 161-180.
Millikan, R.G. (2000) *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
Millikan, R.G. (2004) *Varieties of Meaning*. Cambridge, MA: MIT Press.
Millikan, R.G. (2005) *Language: A Biological Model*. Oxford: Oxford University Press.
Millikan, R.G. (2010) Learning language. https://production.wordpress.uconn.edu/philosophy/wp-content/uploads/sites/365/2014/02/Spracherwerb-Learning-Language.pdf
Millikan, R.G. (2013) Reply to Neander. In Ryder, D., Kingsbury, J. & Williford, K. (eds.) *Millikan and her Critics*. Oxford: Blackwell.
Millikan, R.G. (2017) *Beyond Concepts, Unicepts, Language, and Natural Information*. Oxford: Oxford University Press.
Neander, K. (1995) Misrepresenting and Malfunctioning. *Philosophical Studies*, 79, 109-141.
Neander, K. (2017) *A Mark of the Mental, in Defense of Informational Semantics*. Cambridge, MA: MIT Press.
Onishi, K., Baillargeon, R. (2005) Do 15-Month-Old Infants Understand False Beliefs? *Science* 308 (5719): 255-258.
Pietroski, P. (1992) Intentionality and Teleological Error. *Pacific Philosophical Quarterly*, 73(3): 267-282.
Samson, D., Apperly, I., Braithwaite, J. J., Andrews, B. J. and Bodley Scott, S. E. (2010) Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266.
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.
Sperber, D. (2001) An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29, 401-413.
Sperber, D., & Wilson, D. (1986) *Relevance, Communication and Cognition*. Cambridge, MA: Harvard University Press.
Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., and Wilson, D. (2010) Epistemic Vigilance. *Mind and Language*, 25, 4, 359-393.
Wellman, H.M., Cross, D., & Watson, J. (2001) Meta-analysis of theory of mind development: the truth about false belief. *Child Development*, 72, 655-684.

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.