Croatian Journal of Philosophy

Vol. XIX · No. 57 · 2019

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XIX · No. 57 · 2019

Book Reviews

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Some Reflections on Conventions

CARLO PENCO and MASSIMILIANO VIGNOLO
University of Genoa, Genoa, Italy

In Overlooking Conventions *Michael Devitt argues in defence of the traditional approach to semantics. Devitt's main line of argument is an inference to the best explanation: nearly all cases that linguistic pragmatists discuss in order to challenge the traditional approach to semantics are better explained by adding conventions into language, in the form of expanding the range of polysemy or the range of indexicality (in the broad sense of linguistically governed context sensitivity). In this paper, we discuss three aspects of a draft of Devitt's* Overlooking Conventions*, which was discussed at a conference in Dubrovnik in September 2018. First, we try to show that his rejection of Bach's distinction between convention and standardization overlooks important features of standardization. Second, we elaborate on Devitt's argument against linguistic pragmatism based on the normative aspect of meaning and show that a similar argument can be mounted against semantic minimalism. While Devitt and minimalists have a common enemy, they are not allies either. Third, we address a methodological difficulty in Devitt's view concerning a threat of over-generation and propose a solution to it. Although this paper is the result of collaboration the authors have written different parts. Carlo Penco has written part 1, Massimiliano Vignolo has written part 2 and part 3.*

Keywords: Convention, incompleteness, minimalism, normativity, semantics, standardization.

1. *Conventions and the problem of standardization*

One of Devitt's main claims against linguistic pragmatism (or contextualism) is that many examples intended as cases of meaning underdetermination fall under a more *general* mechanism of meaning formation that Devitt calls 'metaconventions' governing polysemy. However, polysemy is a battlefield among different approaches: cognitive approaches, psycholinguistic approaches, synchronic and diachronic approaches and computational approaches, with no real consensus on the

status of polysemy itself (see for instance Ravin et al. 2000, Nerlich et al. 2003, Vanhove 2008, Falkum, Vicente 2015).

For instance, there is no agreement on whether to treat a linguistic phenomenon as polysemy or semantic generality. The Russian verbs *plavat'* and *plyt'* are both used do designate multidirectional or monodirectional motion in water. In English we have three verbs for motion in water representing passive motion ('float'), self-propelling motion of animated individuals ('swim') and motion of vessels and people aboard ('seal'). We may claim (a) *plyt'* actually distinguishes the three different meanings depending on context, and we may distinguish three different lexical units (or conceptual units), or we may claim that (b) *plyt'* is semantically general and does not distinguish among float, swim and sail (Koptjevskaja-Tamm 2008: 8–9).

There are various tests for distinguishing semantic generality and polysemy, but this distinction is really 'a tricky business' because it often depends on the question under discussion in a specific theoretical settings (see Koptjevskaja-Tamm 2008: 10–13). We accept here some common results in the recent discussion on polysemy and will not enter the debate of polysemy vs. semantic generality. Neither we will follow Grundzinska 2011, who claims, contrary to Devitt's view, that considering polysemy a semantic phenomenon and not a pragmatic one leads to blurring the distinction between semantics and pragmatics and to meaning eliminativism.

We claim that Devitt's insistence on the role of metaconventions for grounding polysemy does not lead to such an undesirable consequence if a more restricted view of conventional meaning is adopted, avoiding a too generalized use of polysemy. Our discussion points to a distinction between what we may call 'strong' and 'weak' metaconvention. Such distinction might be helpful to cope with the alleged shortcomings of Devitt's liberal use of polysemy.

In *Overlooking Conventions,*[1] Devitt employs the notion of metaconvention to address the problems raised by Nunberg (1979: 149–150), who suggests solving some ambiguities of meaning with a pragmatic account of deferred reference and 'explain polysemy without having to introduce any linguistic conventions.' Nunberg was intoducing one of the most debated examples in the literature on meaning underdetermination:

(0) The ham sandwich is sitting at table 20.

Given that sandwiches are inanimate things, they are not agents of actions. The predication 'is sitting' constrains a shift of the meaning of 'ham sandwich' into something that accepts the predicate 'sitting'. In this case the person who ordered the ham sandwich. Nunberg introduces here a pragmatic mechanism, analogous to a metonymical

[1] Given that we refer here to the incomplete draft, dated 7/9/2018, our critical remarks are not strictly directed to the forthcoming book, which might have a different take on the problem discussed here.

transfer from a part to the whole. With the idea of metaconventions Devitt suggests that we have *general* rules for defining different types of conventional meaning with the same lexical entry. They are typically presented in the following form:

if a word refers to things of type X will also refer to things of type Y

Examples are:

> A *count noun* for an organism yields a *mass noun* for its skin (rabbit, crocodile…)
> A word for a *physical entity* yields a word for its *content* (book, television…)
> A word for a *location* yields a word for its *legal entity* or its *people* (state, city…)

In linguistic literature these kinds of expressions are defined as 'dot-objects' or 'dual kinds terms'. They are expressions that can refer to different types: for instance, 'book' or 'television' may refer to a physical or an information entity, 'house' or 'room' may refer to the building or to the location; 'meal' or 'breakfast' may refer to an event or to food. What constrains the choice of the relevant type is the predicative phrase, with a mechanism called 'dot exploitation'. Dot-exploitation is a light form of coercion[2] that consists in exploiting one aspect of the dot-type expression, by predicating only that aspect. In 'The meal was heavy', the predicate 'heavy' constrains the type 'food eaten', while in 'the meal lasted one hour' the predicate 'lasted' constrains the type 'event'.

This particular way to constrain the choice of the type also helps distinguishing regular or logical polysemy from irregular or accidental or idiosyncratic polysemy.[3] Regular or logical polysemy relies on lexical rules, while accidental polysemy is a kind of lexical ambiguity that depends on context. Two basic criteria for distinguishing logical or proper polysemy from accidental polysemy are the Test of Anaphoric Cotenability and the Co-predication Test.[4] The anaphora test is easily exemplified:

(1) That book is boring. Put *it* on the shelf.

Here apparently the anaphora refers to a book as physical object, while the first occurrence of 'book' refers to an informational object.

[2] See Pustejowsky-Jezek (2008); Asher (2011). On coercion see also Asher (2015).

[3] The distinction is not always clear. Apresjan (1974), after distinguishing regular and irregular polysemy, considers the example of the ham sandwich as a case of regular polysemy, something that has been put in doubt later (see Asher 2011, 2015).

[4] Copredication is a topic of interest since Montague 1975 and has raised many problems and tentative solutions in logic and linguistics (see. e.g. Barhamian et al. 2017). Here we are only interested in using it to challenge the idea of too an easy generalization of proper polysemy. We do not discuss tests for distinguishing polysemy from generality or indeterminacy, a topic on which Devitt just raised some doubts and did not elaborate in the draft discussed here.

The copredication test is given as follows: we are in front of a proper polysemy when, in the same sentence, we can attribute to an expression different predicates, concerning different types of things the expression stands for.

Examples are:

(2) Mary picked up and mastered three books on mathematics. *[the two predications refer to a physical object and to its content]*

(3) The city mainly voted democrat and passed a progressive law *[the two predications refer to population and legal entity]*

The copredication test helps understanding the way in which we disambiguate, following the mechanism of dot-exploitation mentioned above. The choice of the type depends on the lexicon used for the predication because the kinds of predications *constrain* the type. In example (2), the predicate 'pick up', a verb for physical activities, constrains the expression 'book' to be intended as a physical object. The predicate 'mastered', a verb for capacities and abilities, constrains the expression 'book' to be intended as an informational object. The choice of meaning therefore depends on the relations among types in the lexicon, which can be viewed as an expression of ontological relations embedded in the lexicon. 'Dual kind terms' are a perfect exemplification of Devitt's examples.

The above criteria for 'proper' polysemy put some worries on a generalized use of polysemy to widen the number of conventional meanings. Devitt presents his answer to Nunberg's example as a consequence of a general metaconvention prompting the following conventional reading:

(4) A word for ordered food yields (in restaurants at least) a word for who ordered it.[5]

Here we face a problem. Actually, it seems that 'ham sandwich' is not a *typical* case of polysemy. There are at least two reasons: as Asher remarks, sentences (0) and (5) seem to stand at different levels: a sentence like (0) is more difficult to process than a sentence like (5):

(0) The ham sandwich is sitting at table 20.
(5) I'm parked out back.

There is a standard metaconvention according to which the word referring to a private vehicle of transportation is often referred to with the word for the owner. I may say: 'my car is parked out back', but nobody would have any difficulty to understand my using (5) as referring to my car (Asher 2011: 250–251). Sentence (0) seems missing this easy interpretation. Second, and even more important, (0) presents some problems about copredication, making some sentences awkward or contradictory:

[5] As in Devitt (draft: 143).

(6) The ham sandwich went away and then he came back and paid for it#.[6]

(7) The ham sandwich left without paying and I have eaten it#.

(8) The ham sandwich that hasn't been eaten is on the counter#.[7]

These problems make the example (0) difficult to be solved in a semantic framework. Even Stanley (2005b: 225), one of the strongest antagonist of contextualists, recognized that problems coming from examples like (0) are genuinely worrying for a semanticist: on the one hand, we recognize that the intuitive truth conditions involve a person rather that a sandwich. But Stanley continues: 'Yet it's not clear that a process that maps ham sandwiches onto persons counts as genuinely semantic.' However, also Recanati's contextualist solution is not without problems. If we take (8) we might interpret it with the reading that the eater of the ham sandwich that hasn't been eaten is on the counter; but why couldn't we interpret it with the reading on which the ham sandwich itself has been put back on the counter? With a general definition of transfer, Asher (2011: 69) claims, 'there are no constraints on when a sense transfer function can be introduced at all (…). Why should we make the transfer in some cases and in other we couldn't?' Transfer function simply runs the risk of overgeneration of meanings.[8]

Devitt implicitly gives a suggestion for an answer. Metaconventions have typically this form:

> "A word for a *physical entity* yields a word for its [informational] *content*",

Differently from the general form of metaconventions, example (0) seems to require a specification:

(4) A word for ordered food yields (*in restaurants at least*) a word for who ordered it.

Our italics makes it apparent that there is some contextual restriction that does not appear in more general metaconventions linked to dual kinds terms expressing polysemy and able to pass the copredication test.

Nunberg's example—example (0)—reminds us that we have an analogous problem with numbers. Certainly there is a general conven-

[6] Suggestion by Belen Soria.

[7] The Example is given in Asher (2011: 65). For a more detailed discussion of similar examples see Asher (2015: 68, 77).

[8] Somebody might use the idea of metonymy. However, cases of this kind are not exactly cases of metonymy because they do not represent a part for a whole, or, better, the part for the whole is *highly theoretically construed and strongly context dependent*: the food for the eater, the chair for the person who should be sitting on the chair, the number for the person somehow linked with the number in a certain situation. Recanati (2010: 167) acknowledges the problem of the dual interpretation: "The ham sandwich stinks' can be so understood, in a suitable context, even though the property of stinking potentially applies to sandwiches as well as to customers'. In this way transfer is not a linguistically controlled process, but it is mere pragmatics, depending freely on intentions and context.

tion to use numbers to refer to everything, and in particular to tables where waiters serve customers, or to customers themselves:

(9) (Customer at table) number 7 left without paying.

However, also numbers seem not to pass copredication tests:

(9) Number 7 went away and then he came back and paid for it#.
(10) Number 7 went away and left itself completely empty#.

There is an obvious convention to use numbers to refer to people in restaurants. The convention is fairly general and works in many other contexts, as, for example, in chess competitions ('Number 7 ended the game'), at the post office ('Number 23 go to the cashier'), at the Hospital ('Please pay attention to number 25'). However, there is no *general* convention for which *kind* of object a number represents: a customer, a chess player, a patient, a bed, a table or what else. The convention is restricted, every time, by a specific setting and by previous agreement on the use of one part of lexicon. In case of restaurant, numbers and food may be used to refer to the person sitting at a table or ordering food. But we need a specific context and a specific agreement among waiters at the restaurant. It cannot be generalised.

Let us make a further example: the expression 'chair' is used at conferences to refer to the chairperson. It seems, again, that there are problems with copredication. We doubt that we can properly accept:

(11) The chair (referring to the chairperson) is not here yet and it (referring to the piece of furniture) is empty#.

Our suggestion is that we are in front of kinds of conventions that, being restricted to specific cognitive contexts, should be distinguished from the standard production of 'conventional meanings' via polysemy. We might call them 'restricted' or 'weak' conventions.

To sum up, these kinds of cases (i) don't appear to be subject to copredication and therefore they cannot be counted as 'dual kinds terms' like 'city', 'book', 'lunch' etc. and (ii) are more 'localised' or linked to specific *cognitive contexts.* Saying that they are 'localized' we mean that they require also a very specialised 'mutual understanding' in local environment (waiters in the restaurant, participants to a conference, and so on). All these cases are not easily treated inside Devitt's framework of metaconventions explaining disambiguation of conventional meanings. Furthermore, they seem to be a good approximation of what Bach meant by 'standardization', which is connected with some kind of *weaker* metaconventions insofar as it requires 'online' inferential processes (restricted to local or specialised cognitive contexts).

The two main ideas supporting standardization are (i) mutual beliefs and (ii) streamlining or default inferences. It is true that Bach's standardization is something not clearly defined and with no sharp and neat difference from convention. However, there is an interesting aspect of Bach's definition of conventionalization as based on '*general* mutual belief', while standardization does not entail such thing (Bach

1995: 683). The implicit suggestion, I suggest, is that linguistic conventions based on general mutual beliefs should be contrasted with conventions based on some *particular* mutual beliefs: a convergence of beliefs grounded on some particular contextual or cognitive settings. We may say that there is no *general* linguistic convention for treating 'ham sandwich', 'number 7', 'bed 25', and 'the chair' for a *specific* kind of object, but only a general strategy of *online adjustments* to recover different kinds of objects depending on the specific or particular contexts.

A useful distinction might be the following: some basic linguistic (meta) conventions are disambiguated *by linguistic context* via type selection constrained by the lexicon. These are typical cases of conventional meanings. Other more specialized cases are disambiguated *by specific cognitive contexts* and require more 'streamlining' inferential processes. Are these cases of standardization? We are content to point out some interesting aspects of Bach's idea of standardization. Not every disambiguation comes from metaconventions, as Devitt (draft: 143) recognises:

> 'Metaphors, Metonymy, synecdoche, yield other examples of polysemous phenomena which often become conventionalized, yielding ambiguities. These processes leading to new meanings are to some extent "rule-governed, and predictable", although not to the extent of those covered by meta-conventions'.

Therefore, Devitt himself acknowledges that there are different kinds of conventions, some generate conventional meanings from polysemy and some are less generalized. We tried to show the difficulty of a too hasty generalization of the idea of meta-conventions supporting different conventional meanings given by polysemy. Shall we be obliged to accept underdetermination of meaning? Sometimes, probably, yes.

2. *The normativity of meaning and minimalism*

Devitt's main line of argument against linguistic pragmatism is based on an inference to the best explanation. Semantics in the traditional approach and linguistic pragmatism agree that their principal theoretical goal is to explain the literal truth conditional content of utterances of sentences. Linguistic pragmatism disagrees with traditional semantics on the idea that all context sensitivity is morpho-lexico-syntactically triggered, either in the form of a plurality of related conventional meanings (polysemy) or in the form of conventions of saturation (indexicality in Devitt's broad sense of linguistically governed context sensitivity). According to linguistic pragmatists, semantic conventions provide at most propositional schemata (propositional radicals) that lack determinate truth conditions. Even in cases in which a sentence possesses determinate truth conditions by semantic conventions alone, there is very often a mismatch between the truth conditions so determined and the truth conditions of the utterances of the sentence. The conclusion drawn by linguistic pragmatists is that the truth conditions

of utterances are underdetermined by their narrow and broad semantic properties and a new theoretical approach introducing truth conditional roles for pragmatic properties is called for.

According to Devitt, the explanation of truth conditions supplied by semantics on the traditional approach is superior to the explanation supplied by linguistic pragmatism because the former accounts for the normative aspect of meaning while the latter does not. Normativity is constitutive of the notion of meaning. If there are meanings, there must be such things as going right and going wrong with the use of language. The use of an expression is right if it conforms with its meaning, and wrong otherwise. If literal contents of utterances are thought of in truth conditional terms, conformity with meaning amounts to constraints on truth conditions. In case of polysemous expressions the speaker undertakes the semantic burden of selecting a convention that fixes a determinate contribution to the truth conditional contents expressed by utterances of sentences. In case of expressions governed by conventions of saturation, the speaker undertakes the semantic burden of loading the demanded parameters with contextual values.

Devitt says that the problem for linguistic pragmatism is to provide an account of how the conventional meanings of expressions constrain truth conditional contents of utterances, if the composition of truth conditions is not governed by linguistic conventions, and how, lacking such an explanation, linguistic pragmatism can preserve the distinction between going right and going wrong with the use of language. In the following we will elaborate on Devitt's argument against linguistic pragmatism based on the normative aspect of meaning and show that semantic minimalism suffers from a similar difficulty. It is difficult for minimalists to explain the normative aspect of meaning.

Semantics on the traditional approach, which Devitt defends, and linguistic pragmatism agree on the view that the goal of semantics is to explain the literal contents of utterances of sentences. They both agree that there must be a close explanatory relation between the meaning encoded in a sentence S and the semantic contents of utterances of S. One corollary of this conception is that if a sentence S is systematically uttered for expressing different contents at different contexts, some expression occurring in S must be context sensitive. As said, the point of disagreement is that semantics on the traditional approach explains context sensitivity by pluralities of conventions and by conventions of saturation, whereas linguistic pragmatism explains it in terms of modulation (optional pragmatic processes).

The debate between Devitt and linguistic pragmatists takes for granted from the start the explanatory connection between meanings and contents of speech acts. Semantic minimalists (Borg 2004, 2012, Cappelen and Lepore 2005, Soames 2002) instead reject such explanatory connection. On their view, semantics is not in the business of explaining the contents of speech acts performed by utterances of sentences. Minimalists work with a notion of semantic content that does

not play the role of (direct) speech act content. According to minimalists the semantic content of a sentence is a full truth conditional content that is obtained compositionally by the syntactic structure of the sentence and the semantic values of the expressions in the sentence that are fixed by conventional meaning. Moreover, minimalists say that the set (the *Basic Set*) of genuinely context sensitive expressions, which are governed by conventions of saturation, comprises only overt indexicals, demonstrative, tense markers and a few other words. Minimalists call the semantic content of a sentence its *minimal proposition*.

The above statement that minimal propositions are not contents of speech acts requires qualification. Cappelen and Lepore (2005) argue indeed for Speech Act Pluralism. They argue that speech acts have a plurality of contents and the minimal proposition of a sentence is always one the many contents that its utterances express. In order to protect Speech Act Pluralism from the objection that very often speakers are not aware of having made an assertion with the minimal proposition as content, and, if speakers were asked, they would deny to have asserted the minimal proposition, Cappelen and Lepore argue that speakers can sincerely assert a proposition without believing it and without being aware of having asserted it.

Semantic minimalists oppose linguistic pragmatism and argue that their examples conflate minimal propositions with speech act contents. Although Devitt and semantic minimalists have a common enemy, they are not allies because they disagree on the theoretical goals of semantics and, consequently, their respective notions of semantic content diverge. In the remainder of this section we will argue that semantic minimalism suffers from a difficulty about the normative aspect of meaning no less than linguistic pragmatism does.

The difficulty for semantic minimalism is brought to light by incompleteness arguments. An incompleteness argument shows that there is no invariant proposition that a sentence S expresses in all contexts of utterance. For example, with respect to the sentence 'Mary is ready' an incompleteness argument starts from the observation that if the sentence is taken separately from contextual information specifying what Mary is ready for, people are unable to evaluate it as true or false. This evidence leads to the conclusion that there is no proposition—that Mary is ready (*simpliciter*)—that is invariant and is semantically expressed by 'Mary is ready' in all contexts of utterance.

Minimalists have responded to incompleteness arguments in two ways. Cappelen and Lepore (2005) accept the premises of incompleteness arguments, i.e. that people are unable to truth evaluate certain sentences, but argue that from these premises it does not follow that minimal propositions do not exist. Borg (2012) adopts a different strategy. Borg tries to block incompleteness arguments by rejecting their premises and explaining away people's inability to truth evaluate the sentences in question. We will argue that both manoeuvres fail.

Cappelen and Lepore (2005) raise the objection that incompleteness arguments try to establish metaphysical conclusions, for example about the existence of the property of being ready (*simpliciter*) as a building block of the minimal proposition that Mary is ready, from premises that concern psychological facts regarding people's ability to evaluate sentences as true or false. They rightly point out that psychological data are not relevant in metaphysical matters. Cappelen and Lepore say that people's inability to evaluate sentences like 'Mary is ready' as true or false independently of contextual information does not provide evidence against the claim that the property of being ready exists and is the semantic content of the adjective 'ready'. On the one hand, they acknowledge the problem of giving the analysis of the property of being ready as a very difficult one, but only for metaphysicians, not for philosophers of language or semanticists. On the other hand, they (2005: 164) argue that semanticists have no difficulty at all in stating what invariant minimal proposition is semantically encoded in 'Mary is ready'. The sentence 'Mary is ready' semantically expresses the minimal proposition that Mary is ready. There is no difficulty in determining its truth-conditions either: 'Mary is ready' is true if and only if Mary is ready.

Cappelen and Lepore address the immediate objection that if the truth conditions of 'Mary is ready' is represented by a disquotational principle like the one reported above, then nobody is able to verify whether such truth conditions are satisfied or not. If the premises of incompleteness arguments are taken at face value, as Cappelen and Lepore do, this fact is witnessed by people's inability to evaluate 'Mary is ready' as true or false independently of information specifying what Mary is ready for. Cappelen and Lepore (2005: 164–165) respond that it is not a task for semantics to ascertain how things are in the world. For example, it is not a task for semantics to say whether 'Mary is ready' is true or false. That a semantic theory for a language L does not provide L-speakers with a method of verification for L-sentences is not a defect of that semantic theory. Cappelen and Lepore say that those theorists who think otherwise indulge in verificationism.

Cappelen and Lepore's confidence in disquotational truth-conditions betrays their underestimation of the real nature of incompleteness arguments. Contrary to what they claim, the conclusion of an incompleteness argument is not a metaphysical conclusion about the existence of this or that entity. Rather, incompleteness arguments provide evidence against the possibility that certain entities get associated with certain expressions as their semantic contents. The conclusion of the incompleteness argument about the adjective 'ready' is not that the property of being ready does not exist because people are unable to evaluate 'Mary is ready' without considering contextual information. The real conclusion of the incompleteness argument is that a semantic theory that assigns the property of being ready to the adjective 'ready' as its semantic content is in tension with the normative aspect of meaning. The reason

why there is no minimal proposition that Mary is ready is not that there is no property of being ready. This is a metaphysical claim that does not follow from people's inability to truth evaluate 'Mary is ready' without taking into account what Mary is ready for. The reason why 'Mary is ready' does not express the minimal proposition that Mary is ready is that the property of being ready cannot be the semantic content of the adjective 'ready', even if one grants that it is a real property. In general, and contrary to Cappelen and Lepore's interpretation, the gist of incompleteness arguments is not that certain entities do not exist, and *a fortiori*, the minimal propositions having those entities as constituents cannot exist. Rather, the gist of incompleteness arguments is that such entities, if any, cannot be the semantic contents of words, because a semantic theory that assigns such entities to words as their semantic contents is incompatible with the normative aspect of meaning, that is with the idea that speaking a language entails being under the normative control of semantic rules. We shall elaborate on this point.

Let us examine the following example in order to better understand the strength of this objection against Cappelen and Lepore. Suppose that a semantic theory for English contains a disquotational clause like (A) below, which arguably captures the idea that Cappelen and Lepore have in mind when they say that the semantic content of the adjective 'ready' is the property of being ready (*simpliciter*):

(A) For any object *o* 'ready' applies to *o* if and only if *o* is ready.

Insofar as (A) is a semantic clause, it has a normative import. It establishes that it is right to apply the adjective 'ready' to all and only objects that are ready. In order for semantics to capture the normative aspect of meaning, clause (A) must exert its normative control over competent English speakers. Moreover, it must also be possible to explain how the adjective 'ready' arrived at the semantic property of applying to all and only objects that are ready. Of course, it is not a task for (descriptive) semantics to answer such question, but a semantic theory must be compatible with an explanation of this sort. Thereby, if we gather evidence that a semantic theory precludes such an explanation, we have evidence that that semantic theory is flawed.

If the premises of incompleteness arguments are true, then it is a fact that people are unable to evaluate sentences like 'Mary is ready' as true or false independently of contextual information. If this is a fact, then people's linguistic practice cannot be under the normative control of clause (A). The reason why people's linguistic practice cannot be so governed is that clause (A) establishes conditions for the application of the adjective 'ready' such that competent speakers are never able to tell whether they are satisfied or not by any object *o*. This is just witnessed by people's inability to evaluate sentences like 'Mary is ready' as true or false independently of contextual information.

The premises of incompleteness arguments, taken at face value, show that the semantic rule expressed by clause (A) is not applicable

because nobody within the linguistic community is able to tell when the conditions for the application of 'ready', as they are captured by clause (A), are satisfied and when they are not. Since rules must be applicable, the conclusion follows that clause (A) does not express any rule, and therefore cannot be a semantic clause, as it cannot play the normative role that is constitutive of semantic rules. Clause (A) does not account for the normative aspect of meaning.

Analogous considerations show that learning of the meaning of the adjective 'ready' cannot amount to learning of the meaning of a word governed by the semantic rule expressed in (A). Presumably we learn the meaning of words such as 'ready' by being exposed to utterances of simple sentences like 'Mary is ready'. If the premises of incompleteness arguments are taken at face value, they show that competent English speakers are never able to track the truth-value of 'Mary is ready' independently of contextual information. If this is true, the premises of incompleteness arguments show that assertions of simple sentences like 'Mary is ready' cannot be expressions of the belief that Mary is ready, i.e. the belief that the conditions for the application of 'ready', as they are captured by clause (A), are satisfied by Mary. If assertions of a simple sentence like 'Mary is ready' are not expressions of the belief that Mary satisfies the application conditions of 'ready', whatever we learn from being exposed to assertions of that sort cannot be the meaning of a word that is governed by the semantic rule expressed by clause (A).

It is important to stress that this argument against Cappelen and Lepore has nothing to do with verificationism. The point is not that competent speakers are unable to evaluate sentences like 'Mary is ready' as true or false because of their epistemic and cognitive limitations. Even if speakers knew everything about Mary, they would not be able to tell whether it is true or false that Mary is ready, unless someone specifies what Mary is said to be ready for. The satisfaction of the application conditions for 'ready', as they are captured by clause (A), is something that is impossible for competent speakers to track. It is like a game whose rules are such that no referee is able to tell whether they are respected or violated by the moves of the players. Clearly such rules could not exert any normative control over the players of the game.

Moreover, in setting up the argument against Cappelen and Lepore one does not need to deny that semantic properties are objective in the sense that they are independent of explicit knowledge and discriminating abilities that competent speakers possess individually or as whole linguistic community. Externalist theories of reference hold that semantic properties are unaffected by explicit and discriminating abilities since they are determined by objective, causal connections to the world. However, externalists do have an account of how words are bestowed with their semantic properties, which basically rests on baptismal events and, above all, multiple groundings. Baptismal events and multiple groundings require dispositions to keep tracks of individuals,

objects, substances, properties and relations in favourable environmental circumstances. Words have the referents they have because, de facto and with the collaboration of the environment, most referential practices are related to those referents. For example, part of the explanation of the fact that the name 'Mary' refers to Mary is that there are (were) people with the disposition to keep track of Mary and the environmental circumstances are (were) favourable (say, Mary does not change the way she looks from one day to the other, or there are not one thousand people looking like her in the same community and people who ground the name 'Mary' onto Mary interact constantly with her). Part of the explanation of the fact that the word 'water' refers to water (the substance whose chemical structure is H_2O), is that there are (were) people with the disposition to keep track of samples of water and the environmental circumstances are (were) favourable. Part of the explanation of the fact that 'blond' refers to the property of being blond, is that there are (were) people with the dispositions to keep track of exemplifications of the property of being blond and the environmental circumstance are (were) favourable. This implies that there are (were) favourable environmental circumstances in which competent speakers are (were) able to point at Mary and say truly 'She is Mary', or to point at a sample of water and say truly 'That is water', or to point at a blond person and say truly 'He/She is blond'. This in turn implies that there are (were) favourable environmental circumstances in which competent speakers are (were) able to truth evaluate sentences like 'That is Mary', 'That is water', 'She is blond', 'Mary is blond'.

Externalist theories of reference keep semantics distinct from theories of linguistic competence. Semantic describes properties of linguistic symbols, theories of linguistic competence describe the abilities of competent speakers to produce and use linguistic symbols. Linguistic competence with referential and inferential abilities is not constitutive of semantic properties. Linguistic symbols are the products of linguistic competence, its outputs (see Devitt and Sterelny 1999: 169). Of course, there is a causal relation between linguistic competence and linguistic symbols. But, as Devitt and Sterelny (1999: 172) point out, there is also a logical relation between linguistic competence and its products: producing linguistic symbols with their semantic properties is what makes it the competence it is. In order for linguistic competence to produce linguistic symbols governed by semantic rules, the conditions for the application of semantically simple words fixed by those semantic rules must be something of which competent speakers are able to keep track in favourable environmental circumstances.

The problem is that if the premises of incompleteness arguments are accepted a true, speakers do not possess the ability to track exemplifications of the property of being ready (*simpliciter*) and do not possess the ability truth evaluate sentences like 'Mary is ready'. Thereby, the externalist account of reference does not work for expressions like

'ready'. And if one loses the account of how the adjective 'ready' got assigned the property of being ready as its semantic content because speakers are unable to track exemplifications of the property of being ready, one loses an account of how clause (A) can exert any normative force over linguistic practices of competent speakers.

Cappelen and Lepore are right that the premises of incompleteness arguments do not entail that certain properties like being ready (*simpliciter*), or being tall (*simpliciter*), or being strong (*simpliciter*), or having enough (*simpliciter*) etc. do not exist. But this is beside the point. The premises of incompleteness arguments show that speakers are never able to track exemplifications of those properties. It follows that minimalist semantic clauses like (A) express semantic rules such that nobody is ever able to tell when they are respected and when they are violated. Such minimalist semantic rules are inapplicable and inapplicable semantic rules cannot exert any normative control over linguistic practice. Semantic minimalism faces a problem with the normative aspect of meaning: if linguistic practice is not under the normative control of semantic rules, there cannot be such things as going right and going wrong with the use of language.

In *Pursuing Meaning* Borg adopts a different strategy against incompleteness arguments. Borg rejects their premises and explains away the intuitions of incompleteness. Borg (2012: 92–102) agrees that speakers have an intuition of incompleteness with respect to sentences like 'Mary is ready', but she argues that intuitions of incompleteness emerge from some overlooked covert and context-insensitive syntactic structure. Borg says that 'ready' is lexically marked as an expression with two argument places. On Borg's view 'ready' always denotes the same relation, the relation of *readiness*, which holds between a subject and the thing for which they are held to be ready. When only one argument place is filled at the surface level, the other is marked by an existentially bound variable in the logical form. The argument role corresponding to the direct object is existentially quantified instead of being assigned a particular value. The suppression of the direct object arguably changes the semantic content of the adjective: it denotes not the original two-place relation, but a property generated by existentially quantifying the object argument-role. Thereby 'ready' makes exactly the same contribution in any context of utterance to any proposition literally expressed. For example, Borg says that in a context where what is salient is the property being ready to join the fire service the sentence 'Mary is ready' literally expresses the minimal proposition that *Mary is ready for something* not that *Mary is ready to join the fire service,* and in a context where the property of being ready to take an exam in logic is salient 'Mary is ready' still literally expresses that *Mary is ready for something*. As Borg (2012: 104) points out, the minimal proposition that *Mary is ready for something* is almost trivially true, because it is true in any possible world where Mary exists. Yet,

Borg warns us not to conflate intuitions about the informativeness of a proposition with intuitions about its semantic completeness.

Borg's explanation of the intuitions of incompleteness is that speakers are aware of the need for the two arguments, which is in tension with the phonetic delivery of only one argument. Speakers are uneasy to truth-evaluate sentences like 'Mary is ready' not because the sentence is semantically incomplete and lacks determinate truth conditions, but because their expectation for the second argument to be expressed is frustrated and the minimal content that is semantically expressed, when the argument role corresponding to the direct object is not filled at the surface level, is barely informative. Borg's response to incompleteness arguments avoids the problem that affects Cappelen and Lepore's version of minimalism. On Borg's view, speakers are able to truth evaluate the minimal content of 'Mary is ready', since that content is the minimal proposition that Mary is ready for something. If 'ready' in sentences like 'Mary is ready' literally means *ready for something*, competent speakers are obviously disposed to track the application conditions for 'ready'.

In a significant respect Borg's solution goes in the same direction as the traditional approach in semantics. As said, on a traditional semantic theory the meaning of context sensitive expressions sets up the parameters that must be loaded with contextual values. Sometimes the parameters are explicitly expressed in the syntax of the sentence as with indexicals, demonstratives, tense markers of verbs. Sometimes, instead, the parameters do not figure at the level of surface syntax. Philosophers and linguists disagree on where the parameters that do not show up at the level of surface syntax are hidden. Some (Stanley 2005a) hold that such parameters are associated with syntactic elements that occur in the logical form. Taylor (2003) advances a different theory. Taylor argues that hidden parameters are represented in the syntactic basement of the lexicon. They are not constituents of sentences but subconstituents of words or phrases. On Taylor's view, the lexical representations of words and phrases specify the parameters that must be loaded with contextual values in order for utterances of sentences to have determinate truth conditions. Taylor's proposal is a way of implementing the view that context sensitive expressions are governed by conventions of saturation and that context sensitivity is always morpho-lexico-syntactically driven. Taylor's view amounts to a denial of the phenomenon of meaning underdetermination and semantic incompleteness and it is a way of treating context sensitivity within the camp of traditional semantics. Thus, when Borg says that 'ready' is lexically marked as an expression with two argument places, she says something that might go in the very same direction as Taylor's. If Taylor's proposal is a way of implementing the traditional view in semantics, so it seems to be Borg's view. Yet, Borg is unwilling to accept this conclusion. Borg refuses to treat 'ready' and all the expressions that

are typically involved in incompleteness arguments as context sensitive expressions.

We will argue that Borg's conception of semantics faces a problem and the utterance-oriented conception of semantics on the traditional approach seems to fare much better with respect to that problem. Borg's version of minimalism maintains that the semantic content of a sentence S is typically different from the contents of speech acts performed by utterances of S. Clapp (2007) raises the following naturalistic challenge to minimalism. If it is a fact that an expression has a certain meaning, this fact must be grounded in facts concerning the linguistic abilities and practices of competent speakers. The difficulty for minimalism is to provide an account of what grounds the fact that an expression has the meaning it has, since minimalism keeps semantic contents apart from speech acts contents. Facts regarding speech acts contents have no bearing on facts regarding semantic contents (Cappelen and Lepore 2005: 211). On the other hand, utterance-oriented semantics has the advantage of relying on regularities of uses in linguistic practices. As Devitt (2007: 52) says, meanings are not 'God given', but as conventions need to be established and sustained by regular uses. On Devitt's view, linguistic rules reveal themselves in the regular uses of certain forms for expressing certain contents. In order to individuate conventions, theoreticians can rely on an inference to the best explanation: they must consider whether the regular use of an expression for performing certain speech acts is best explained by positing a linguistic rule for using that expression. Coming back to the example with the adjective 'ready', Borg owes an explanation of what make it the case that 'ready' literally means 'ready for something' when its second argument place is not lexicalized at the surface level. As Clapp points out, Borg's view that our linguistic knowledge is encapsulated in a dedicated module that represents the biconditional that the sentence 'Mary is ready' is true if and only if Mary is ready for something offers no answer. The problem is simply relocated. The problem now is to explain in virtue of what the language module works the way Borg takes it to work.

One might think that there are other theoretical reasons for favouring Borg's conception of semantics. In the next section we will discuss a recent attempt that Borg made to support the claim that minimal contents play an important theoretical role that contents of other kinds cannot play. We will argue that Borg's argument is inconclusive. In the remainder of this section we will comment on two earlier arguments that Borg provides for proving her version of minimalism superior to Cappelen and Lepore's one and to Bach's radical minimalism.

Borg (2007: 351) argues that her account provides a more credible version of minimalism than Cappelen and Lepore's version. According to Cappelen and Lepore, the sentence 'Mary is ready' expresses the minimal proposition that Mary is ready (*simplicter*). If this is so, then the sentence 'Mary is not ready' expresses the proposition that Mary

is not ready (*simpliciter*). Borg argues that Cappelen and Lepore proposal is unable to explain situations in which both sentences are true together, for instance if Mary is ready to go to the party but not ready to take the logic exam. Borg proposal accommodates this case giving narrow scope to the negation: 'Mary is ready and Mary is not ready' is true if and only if there is something for which Mary is ready and there is something for which Mary is not ready. We will not address the question whether Borg's argument is a good one against Cappelen and Lepore. We point out that it does not raise any difficulty for a traditional utterance-oriented semantics according to which there might be true utterances of 'Mary is ready and Mary is not ready.' Suppose John is talking to Jeff and Mark. Jeff wants to know whether Mary is ready to go to the party and Mark wants to know whether Mary is ready to take the logic exam. John can say 'Well, Jeff, Mary is ready but, Mark, she is not ready' and tell the truth. John can say that having in mind going to the party for the saturation of the first occurrence of 'ready' and taking the logic exam for the saturation of the second occurrence.

Borg (2012: 209) makes an attempt to promote her view against Bach's radical version of minimalism. Borg says that the view that the sentence 'Mary is ready' literally expresses the minimal content that Mary is ready for something copes with the Cancellability Test. She rightly says that readings that make it explicit the presence of an existentially bound variable cannot be cancelled without contradiction. It is not possible to say without contradiction 'Mary is ready, though I do not mean ready for something'. Borg's conclusion is that a reading that cannot be cancelled without contradiction seems to have the right to be the literal content of a sentence. Borg rhetorically wonders why one cannot cancel the existentially bound content and assert the gappy content (the propositional radical) that Bach takes to be the literal content of 'Mary is ready.' In the same vein, Borg says that it is always possible to retract a contextually enriched content. Even in a context in which it is readiness to go to the party that is salient, one can say 'Mary is ready, but I mean to take the logic exam, not to go to the party.'

We want to stress two points in reply to Borg. First, it is true that in Borg's example the speaker retracts the content that Mary is ready to go to the party. But the speaker does so by loading another value for the parameter of 'ready.' This is in line with the metaphysical role that the speaker plays in the determination of what is said. What is said is not determined by what is salient in the context of utterance, or by what the hearer understands, or by what the hearer is expected to understand. It might be very likely that in a context in which going to the party is salient, if the speaker says 'Mary is ready,' the hearer will understand that Mary is ready to go to the party. But this is not determinative of what the speaker semantically expresses. Moreover, it does not follow that Mary is ready for something is the literal meaning of 'Mary is ready' from the premise that such content is not cancel-

lable. On the traditional approach, that Mary is ready for something is a logical consequence of the semantic contents of utterances of the sentence 'Mary is ready.' Clearly, if the semantic content of an utterance of 'Mary is ready' is that Mary is ready to take the logic exam, that semantic content entails that Mary is ready for something, which cannot be cancelled without contradiction. It does not follow that 'ready' is not a context sensitive expression and that 'Mary is ready' literally expresses the minimal content that Mary is ready for something.

Second, it is worth noticing that the speaker cannot retract the content that Mary is ready to go to the party by saying 'Mary is ready, but I mean for something, not to go to the party.' That move would be an open violation of the maxims of the cooperative principle. Indeed, the speaker would make it manifest that she is literally saying something that is almost trivially true, and thereby not informative or relevant. The speaker cannot retract the content that Mary is ready to go to the party by retreating to Borg's minimal content that Mary is ready for something without making it explicit that she is not cooperative. We will come back to this point in the next section.

We have one last comment on minimal contents. Minimalists argue that minimal propositions serve as fall back contents when contextual information helpful for hearers to figure out the speakers' intentional states is inaccessible or insufficient or unreliable. Borg holds that linguistic knowledge is encapsulated in a language module and insulated from non-linguistic information. The linguistic knowledge so encapsulated and insulated guarantees that any competent speaker is able to recover a truth conditional content merely through exposure to the sentence uttered. Yet, semantics on the traditional approach does not need to deny the existence of a layer of truth conditions that are recoverable only on the basis of strict linguistic knowledge. Semantics in the narrow sense is the study of the meanings of simple expressions and their modes of combination. These semantic properties of expressions determine the conditions that must obtain in order for an utterance of a sentence to express a truth. This is the layer of truth conditional content that some philosophers (Perry 2001, Korta and Perry 2011) capture with the notion of token-reflexive content, or utterance-bound content and in model-theoretic or other formal approaches to languages (Kaplan 1989) is represented with semantic compositional clauses that quantify over indexes that represent contextual factors. A competent speaker can know what conditions must obtain for an utterance of a sentence or a sentence at an index to express a truth without having any clue about the speaker's intentional states that determine the values of saturation and, therefore, without grasping the semantic content of the utterance (Korta and Perry's locutionary content). Any other additional layer of truth conditions such as minimal propositions seems to be an arbitrary posit that becomes an idle wheel.[9]

[9] Korta and Perry (2006, 2008) discuss several examples to show that in

3. *Fixing conventions*

In this section we will address a methodological difficulty in Devitt's view and propose a solution to it. As said, Devitt's strategy for defending the traditional approach to semantic is to expand the range of polysemy and indexicality (in the broad sense of linguistically governed context sensitivity) by increasing the number of conventions in language. Devitt's view raises the immediate difficulty of telling what is the evidence for tracking conventions in language. Devitt rejects the recourse to intuitions on truth conditions, judgements on reports on what is said, judgments on contents consciously accessible during online processing of sentences, and judgments on input for rational reconstruction of conversational implicatures. Notoriously, such judgments by laypersons are inconsistent and unreliable because they tend to conflate contents that are semantically expressed with contents that are pragmatically conveyed. On the other hand, the experts' judgments run the risk of being biased by the theories they embrace.

Devitt suggests looking for evidence in the regular and systematic usage of expressions. If speakers regularly and systematically use certain expressions to express certain contents, then theoreticians must consider whether such regularities are best explained by supposing that there are linguistic rules of using those expressions that way. Theoreticians are justified to posit conventional rules if by doing so they obtain the best explanation of speakers' linguistic behaviour.

We believe that Devitt's methodological picture is basically correct but it is too sketchy as it stands and runs the risk of over-generation. Let us consider the following example with 'to cut'. It seems uncontroversial that in many typical contexts, the verb 'to cut' conveys the information that the act of cutting is performed in a canonical way depending on the situation:

Hairdresser context: John cut Marie's hair [with hairdressing scissors]
Cook context: John cut the meat [with a knife].
Fireman context: John cut the car door [with rescue shears].
Woodsman context: John cut the tree [with an axe].
Tailor context: John cut the silk [with tailor's scissors].
Gardener context: John cut the grass [with a lawnmower].

It seems a regularity of use that in specific contexts the verb 'to cut' conveys the information that the act of cutting is performed with a specific tool. Is this information encoded in the meaning of the verb 'to cut'? If this is so, is it encoded in virtue of polysemy or in virtue of a convention of saturation? And if it is a convention of saturation that demands the speaker undertake the semantic burden of having in mind a tool or a way of cutting, how can we tell whether there are

many cases hearers do not need to grasp what speakers semantically say in order to understand what speakers intend to convey. It is enough that they gasp the utterance-bound content.

other parameters that require saturation, for example about the rapidity or the precision and straightness of the cut, or the location where the action of cutting takes place? Devitt sketchy suggestion that if the examination of linguistic usage shows that an expression is regularly used to express certain contents then we have good evidence that such use is conventional is not much help to work out the answers. What kind of data can theoreticians rely on in order to make progress in their semantic theories?

Some philosophers (see Borg 2012: 206) and linguists propose to look for evidence at the syntactic level. Recanati (2004: 102) discusses and rejects the Binding Criterion:

A contextual ingredient in the interpretation of a sentence S results from saturation if it can be 'bound', that is, if it can be made to vary with the values introduced by some operator prefixed to S.

The problem with the Binding Criterion is that it over-generates. As Cappelen and Lepore (2002), Breheny (2004), and Recanati (2004) point out, if the Binding Criterion is employed as a test for detecting parameters that demand saturation, it yields an unacceptable proliferation of parameters. In point of fact, in order to defend the Binding Argument from the charge of over-generation, Stanley (2005b: 235) urges not to interpret it as a criterion for detecting hidden parameters. Stanley says that the Binding Argument must be taken as an inference to the best explanation of bound interpretations: by postulating covert variables one can explain bound interpretations. On Stanley's view, evidence for bound interpretations comes from speakers' intuitions on truth conditions. From Stanley's perspective, then, the Binding Argument does not provide evidence for detecting hidden parameters. Rather, it presupposes evidence for bound interpretations from speakers' intuitions on truth conditions.

Furthermore, Recanati (2004: 110) proposes an alternative explanation of bound interpretations that avoids the presence of covert variable in the logical form of expressions. Recanti rejects the argument from premises 1 and 2 to conclusion 3:

1. In the sentence 'whenever Bob lights a cigarette, it rains', the reference to the location varies according to the value of the variable bound by the quantifier 'whenever Bob lights a cigarette'.

2. There can be no binding without a variable in the logical form.

3. In the logical form of 'it rains' there is a variable for locations, although phonologically not realized.

Recanati argues that this argument is fallacious because of an ambiguity in conclusion 3, where the sentence 'it rains' can be intended either in isolation or as a part of compound phrases. According to Recanati, the sentence 'it rains' contains a covert variable when it occurs as a part of the compound sentence 'whenever Bob lights a cigarette, it

rains', but it does not contain any variable when it occurs as an atomic sentence.

Recanati explanation of bound interpretations exploits expressions that modify predicates. Given an n-place predicate, a modifier can form an n+1 place or an n–1 place predicate. Expressions like 'here' or 'in London' are special modifiers that transform the predicate 'it rains' from a one-place predicate to a two-place predicate but provide also a value for the new argument place. Recanati argues that expressions like 'whenever Bob lights a cigarette' are modifiers like 'here' and 'in London'. They change the number of predicate places and provide a value to the new argument through the value of the variable they bind. Recanati's conclusion is that although binding requires variables in the logical form of compound sentences, there is no need to insert covert variables in sub-sentential expressions or sentences in isolation.

Thus, to appeal to the Binding Criterion amounts to putting the syntactic cart before the semantic horse with the risk of over-generation and fallacy and the appeal to the Binding Argument presupposes a methodology that relies on speakers' intuitions on truth conditions, which Devitt explicitly rejects. If evidence is not to be found at the syntactic level, it must be found elsewhere.

In the previous section, we saw that Devitt puts much weight on the normative aspect of meaning in order to mount an argument against linguistic pragmatism. One might try to analyse the semantic burdens that speakers undertake in utterances of sentences to collect evidence for the structure of semantic contents. This is to say that one might collect evidence by the study of the moves that speakers are allowed or obliged to do for defending or retracting their utterances. Elaborating on Grice (1989), Michaelson (2016: 477) takes into consideration the Cancellability Test:

If *q* is part of the semantic content expressed by a sentence S at a context C, then:
A. One should not be able to consistently utter 'S, but not Q' at C, where
B. 'not Q' is a standard way of denying *q*.

However, with respect to Devitt's attempt to defend the traditional approach to semantics by expanding the range of polysemy, the Cancellability Test has a severe limitation. Consider the sentence 'John and Mary got married and had a child'. Devitt explains the interpretation that John and Mary got married before having a child by polysemy: 'and' is a polysemous word having multiple meanings, one for the truth-functional conjunction and one for the temporally/causally ordered conjunction. Of course, the temporal ordering can be cancelled. One might say 'John and Mary got married and had a child, but not in that order'. Yet, as Michaelson acknowledges, to argue that Devitt's theory is mistaken because it fails the Cancellability Test would be to beg the question against Devitt. It is open to Devitt to claim that the phrase 'but not in that order' does not cancel a pragmatic enrichment but makes

it explicit a disambiguation. The Cancellabiltiy Test does not supply relevant data for deciding whether certain forms of context sensitivity can be explained by polysemy and it has a very limited application for Devitt's purpose of collecting data from the usage of sentences.

The Cancellability Test is based upon the idea that the semantic content of an utterance is something to which the speaker is committed on pain of contradiction or semantic incompetence. Elaborating on this idea, some philosophers like Saul (2012), Michaelson (2016), Borg (2017) have proposed to make use of judgements of lying for tracking semantic contents. The central assumption is that if a speaker utters a sentence S and is not lying, then p is not the semantic content of S provided that the speaker believes the content p to be false and intends to deceive her audience about p. Michaelson (2016: 482) offers the following formulation of the Lying Test:

If *p* is part of the semantic content associated with a sentence P,
as uttered by X to Y, then either:
A. P is a lie, or
B. it is not the case that X believes that p is false, or
C. it is not the case that X intends to deceive Y with respect to p.

Michaelson and Borg[10] employ the Lying Test to argue against the idea that the conjunction 'and' is polysemous. Consider the following example in Borg (2017):[11]

A rich catholic fundamentalist decides to leave her entire fortune to Jack, as long as Jack has lived his life in full compliance with the precepts of Catholicism. The rich fundamentalist asks John for information about Jack's life. John intends to favour his friend Jack wishing him to inherit the huge amount of money and, knowing that Jack had two children before getting married, he says:

Jack got married and had two children.

John intends his speech act to make the rich fundamentalist believe that Jack got married and *then* had two children. John's utterance is misleading and clearly intended to be so. Moreover, John knows that it is false that Jack got married before having two children.

By the application of the Lying Test, Michaelson argues that since John is not lying, believes the temporally ordered content to be false, and intends to deceive the rich fundamentalist about that content, the temporally ordered content is not the semantic content of John's utterance. On Michaelson's view, the Lying Test provides evidence in favour of the unified account of the meaning of 'and' and against the polysemous account.

[10] More precisely, Borg argues against linguistic pragmatism and in defence of minimalism.

[11] Borg's example is a variation of an example in Saul (2012: 37).

We agree that the Lying Test is somehow on the right track for collecting evidence in semantics but disagree on Michaelson's on his conclusion against the polysemous account of the meaning of 'and' (and we disagree with Borg on her use of the Liying Test for defending minimalism). Michaelson says that the polysemous account predicts that John semantically expressed the temporally ordered content that Jack got married before having two children because it is the speaker's prerogative to choose how polysemous expressions should be disambiguate, and John intends for his use to be disambiguated temporally. We argue that Michaelson's argument fails because it conflates the metaphysics of meaning with the epistemology of understanding. Certainly, it is the speaker's prerogative to choose how an expression has to be disambiguated. In the above scenario, if someone charged John of lying, nothing could prevent John from defending himself and claiming that he said that Jack got married and had two children in one order or the other. John's self defence could not be impeached by observing that that is not how the rich fundamentalist interpreted John's utterance or that John knew that that was not how the rich fundamentalist would interpret his utterance. What the hearer does or what the hearer is expected to do is not determinative of semantic content. To say that it is the speaker's prerogative to choose how an expression should be disambiguated is to say that the speaker undertakes the semantic burden of choosing a certain meaning. To the extent that in the depicted scenario John is allowed to choose the truth functional meaning for 'and' and to defend his choice explicitly and in public, there is no reason to force upon his utterance the temporally ordered content, even if John knew that the rich fundamentalist would interpret his utterance that way.

Of course, John's communicative strategy is very tricky, but what makes it tricky is just the fact that in the above scenario John can play with the polysemy of 'and'. Indeed, if we change the scenario and imagine a situation in which John cannot play with the polysemy of 'and', we get evidence in favour of the polysemous account. Suppose that the rich fundamentalist asks John the following direct question and John gives the following answer:

Fundamentalist: Did Jack get married and have two children or did he have two children and get married?

John: Jack got married and had two children

In this case, the intuition that John is lying and not merely misleading his interlocutor is stronger than the intuition that John is not lying. Nobody would accept as legitimate John's defence that he was not saying that *Jack got married and then had two children.* Contrary to the previous scenario, given the formulation of the question asked by the rich fundamentalist, in which it is clear that the conjunction 'and' is used with the temporally ordered content, John cannot respond that he was not saying that *Jack got married and then had two children*, on pain of

making it open that he did not understand the question, and thereby on pain of showing himself linguistically incompetent or non-cooperative.

The view that treats the conjunction 'and' as polysemous offers a straightforward explanation of what happens in the second scenario. The rich fundamentalist uses 'and' with the temporally ordered meaning. Therefore, the retreat to the truth functional meaning would be an unacceptable admission of linguistic incompetence on behalf of John. In this case John cannot play with the polysemy of 'and', given the way in which the rich fundamentalist asks the question. It is not obvious that the unified account of the meaning of 'and' can cope with this case, as it lacks an explanation of the strong intuition that John is lying and not merely misleading his interlocutor.

One interesting aspect of the Lying Test is that it works with a notion of semantic content that is characterised in terms of the semantic burdens that speakers undertake in utterances of sentences. It connects semantic contents to utterances in virtue of the linguistic liability that speakers are held to have for the contents of the speech acts they perform. These semantic burdens can be investigated by studying the moves that speakers are allowed or obliged to make when their utterances are challenged, on pain of linguistic incompetence, irrationality or non-cooperativeness. The analysis of such moves is helpful to work out a solution to the slippery slope argument that threatens the theories that aim to treat context sensitivity as a semantic phenomenon. The slippery slope argument leads to the conclusion that if one starts treating some expressions as context sensitive on the basis of context shifting arguments and incompleteness arguments, one loses a principled way to distinguish context sensitive expressions from context invariant ones and a principled way to select for any context sensitive expression the parameters that demand saturation, because for any expression and after any process of saturation one can always raise further questions about more contextual precisifications.

Our answer to the slippery slope argument is that what matters is not the openness to further questions for more precisifications, but the kind of legitimate answers that speakers are allowed to give. We propose to use more vigorously the No-Idea Test that Recanati discusses in (2010: 84).[12] The basic insight underlying the No-Idea Test is that if an expression demands the saturation of a certain parameter, the speaker is not allowed to reply with 'I have no idea' to a request of precisification. For example, the No-Idea Test provides evidence that the verb 'to arrive' requires saturation for the location of the arrival, as the infelicity of the following dialogue shows:

A. John has arrived.
B. Where has he arrived?
A. I have no idea.#

[12] Recanati (2010: 84) says that the No-Idea Test was originally proposed by Jarmila Panevova.

The reason why the speaker is not allowed to reply with 'I have no idea' is that the speaker cannot avoid undertaking the semantic burden of specifying the location where John arrived on pain of committing her speech act to the content that John arrived in some place or other. This content is in open violation of the maxims of conversation, because it is not relevant and very likely the speaker has no justification for making an assertion with that content. The speaker cannot commit herself to that content, on pain of proving herself non-cooperative.

Likewise, one is not allowed to reply with 'I have no idea' to a request of precisification for those expressions like 'ready', 'tall', quantified nouns phrases, that linguistic pragmatists typically employ to construct counterexamples to semantic theories on the traditional approach. The following dialogues are all infelicitous:

A. John is ready.	A. John is tall.	A. There are no beers.
B. What is he ready for?	B. What is he tall for?	B. Where?
A. I have no idea.#	A. I have no idea.#	A. I have no idea.#

On the contrary, the No-Idea Test shows that the way in which the action of cutting is performed is not part of the semantic content of the verb 'to cut'. The following dialogue looks fine:

A. John cut the cake.
B. How did he manage to cut the cake? There were no cooking utensils in the kitchen!
A. I have no idea.

This is evidence that the verb 'to cut' does not demand saturation for the way of cutting. As Devitt points out, 'to cut' might have a context invariant content along the lines of *to produce linear separation in the material integrity of something by a sharp edge coming in contact with it*. The information about the way in which the action of cutting is performed is pragmatically conveyed, not semantically encoded in the meaning of 'to cut'.

The No-Idea Test provides evidence that 'ready', 'tall', quantified noun phrases pattern with 'to arrive'. Their meaning demands that the speaker undertake the semantic burden of saturating certain parameters. Weather reports are other examples that linguistic pragmatists typically employ to argue against traditional semantic theories. We acknowledge that weather reports are much more controversial cases. On the one hand, the following dialogue might seem infelicitous as much as the previous ones:

A. It is raining.
B. Where is it raining?
A. I have no idea.#

On the other hand, Recanati (2002: 317) has discussed the 'weatherman' scenario for supporting the claim that 'to rain' does not demand saturation for locations: after weeks of total drought, one of the alarm

bells that are connected to rain detectors that have been placed all over the territory rings in the monitoring room. The weatherman on duty in the adjacent room says: 'It is raining'. The following dialogue looks fine (Recanti 2010: 86):

A. (The weatherman) It is raining.
B. Where is it raining?
A. I have no idea. Let us check.

Recanati holds that the truth conditions of the utterance of the weatherman are that it is raining in some place or other. According to Recanati, the possibility of the indefinite reading proves that the felt compulsion to complete truth conditions of weather reports with locations, when such a compulsion is indeed felt, has a pragmatic nature. Recanati (2010) gives a long argument against the possibility of explaining the indefinite reading through a covert existential quantification on the parameter for the location.

As we said, this case is very controversial and we have no space to discuss it at length. We have just a couple of remarks. First, taking for granted that the weatherman is not able to make reference to the location where it is raining (i.e. to entertain a singular proposition about that location), it does not follow that the weatherman does not have in mind that location by description, in such a way that the weatherman is able to denote the location where it is raining (i.e. to entertain a general proposition about that location). Indeed, the weatherman can think of that location as the location where the rain detector that caused the alarm bell to ring has been placed. There is a reading according to which the truth conditions of the weatherman's utterance are that it is raining at the location where the rain detector that caused the alarm bell to ring has been placed. Thus, we put in doubt the claim that the weatherman's example is a genuine case of indefinite reading.

Second, Recanati's argument against the possibility of explaining indefinite readings through a covert quantification rests on a doubtful and idiosyncratic intuition. Recanati argues that there are utterances of 'It is not raining' that cannot be given the indefinite reading that somewhere it is not raining, which is the reading that is predicted by the theory that explains indefinite readings through covert existential quantification over the location parameter. Recanati (2010: 103) discusses a 'reversed weatherman' scenario: after a long period of heavy rain and floods all over the territory detectors for the absence of rain are placed. One day the alarm bell connected to a detector rings and the weatherman on duty says 'It is not raining'.

Recanati's comment is that he finds it rather hard to understand the utterance with the content that somewhere it is not raining (wide scope indefinite reading). Recanati's intuition is that the only available interpretation is that it is not raining anywhere (narrow scope indefinite reading). According to Recanati, the weatherman ought to say 'The rain has stopped', which could be interpreted as meaning that

the rain has stopped somewhere. Thus, Recanati's conclusion is that the theory that explains indefinite readings of weather reports through a covert quantification over the location parameter is unable to explain the unavailability of the wide scope indefinite reading in the reversed weatherman scenario.

We acknowledge that weather reports are very controversial cases and leave the full discussion of them for another paper. We want to stress, however, that Recanati's argument rests entirely on his intuition that the wide scope indefinite reading in the reversed weatherman scenario is not available. We find Recanati's intuition no less controversial than weather reports in general.

References

Apresjan, J. D. 1974. "Regular polysemy." *Linguistics* 14 (2): 5–32.

Asher, N. 2011. *Lexical meaning in Context*. Cambridge: Cambridge University Press.

_____, N. 2015. "Types, Meanings and Coercions in Lexical Semantics." *Lingua* 157: 66–82.

Bach, K. 1995. "Standardization vs. Conventionalization." *Linguistics and Philosophy* 18: 677–686.

Bahramian, A., Nematollahi, N. and Sabry, A. (2017). *Copredication in homotopy type theory*. https://hal.archives-ouvertes.fr/hal-01628150

Borg, E. 2004. *Minimal Semantics*. Oxford: Oxford University Press.

_____, 2007. "Minimalism versus Contextualism." In G. Preyer and G. Peter (eds.). *Context Sensitivity and Minimalism*. Oxford: Oxford University Press: 339–359.

_____, 2012. *Pursuing Meaning*. Oxford: Oxford University Press.

_____, 2017. "Explanatory Roles for Minimal Content." *Nous*. doi: 10.1111/nous.12217: 1–17.

Breheny, R. 2004. "A Lexical Account of Implicit (Bound) Contextual Dependence." In R. Young and Y. Zhou (eds.). *Semantics and Linguistic Theory* (SALT) 13: 55–72.

Cappelen, H. and Lepore, E. 2002. "Indexicality, Binding, Anaphora and A Priori Truth." *Analysis*, 62 (4): 271–81.

_____, 2005. *Insensitive Semantics*. Oxford: Blackwell.

Clapp, L. 2007. "Minimal (Disagreement about) Semantics." In G. Preyer and G. Peter (eds.). *Context Sensitivity and Minimalism*. Oxford: Oxford University Press: 251–277.

Devitt, M. 2007. "Referential Descriptions: A Note on Bach." *European Journal of Analytic Philosophy* 3: 49–53.

_____, M. Draft. *Overlooking Conventions* (incomplete draft of a forthcoming book, discussed at Dubrovnik, September 2018).

Devitt, M., Sterelny, K. 1999. *Language and Reality: An Introduction to the Philosophy of Language*. 2nd edn. Cambridge, MA: MIT Press.

Falkum, I. L. and Vicente, A. 2015. "Polysemy: Current perspectives and approaches." *Lingua* 57: 1–16.

Grice, P. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Grudzińska, J. 2011. "Polysemy: an Argument against the Semantic Account." *Kwartalnik Neofilologiczny* LVIII, 3: 273–282.

Kaplan, D. 1989. "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan*. Oxford: Oxford University Press: 481–563.

Koptjevskaja-Tamm, M. 2008. "Approaching lexical typology." In Vanhove 2008: 3–52.

Korta, K. and Perry, J. 2011. *Critical Pragmatics*. Cambridge: Cambridge University Press

_____, 2006. "Three Demonstrations and a Funeral." *Mind & Language* 21 (2): 166–86.

_____, 2008. "The Pragmatic Circle." *Synthese* 165 (3): 347–57.

Michelson, E. 2016. "The Lying Test." *Mind & Language* 31 (4): 470–499.

Nerlich, B., Todd, Z., Vimala, H., and Clacke, D. D. (eds.). 2003. *Polisemy. Flexible Patterns of Meaning in Mind and Language*. Berlin: De Gruyter.

Nunberg, G. 1979. "The Non-Uniqueness of Semantic Solutions: Polysemy." *Linguistics and Philosophy* 3 (2): 143–184.

Ortega-Andrés, M. and Vicente, A. 2019. "Polysemy and co-predication." *Glossa: A Journal of General Linguistics* 4 (1), 1: 1–23. doi: http://doi.org/10.5334/gjgl.564

Perry, J. 2001. *Reference and Reflexivity*. Stanford: CSLI Publications.

Pustejovsky, J. and Jezek, E. 2008. "Semantic Coercion in Language: Beyond Distributional Analysis." *Rivista di linguistica* 20 (1): 181–214.

Ravin, Y. and Leacock, C. (eds.). 2000. *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press.

Recanati, F. 2004. *Literal Meaning*. New York: Cambridge University Press.

_____, 2010. *Truth Conditional Pragmatics*. Oxford: Oxford University Press

Saul, J. 2012. *Lying, Misleading, and the Role of What is Said*. Oxford: Oxford University Press.

Soames, S. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.

Stanley, J. 2005a. *Language in Context*. Oxford: Oxford University Press.

_____, 2005b. "Semantics in Context." In G. Preyer, and G. Peter (eds.). *Contexts in Philosophy*. Oxford: Oxford University Press: 221–253.

Taylor, K. 2003. *Reference and the Rational Mind*. Stanford, CA: CSLI Publications.

Vanhove, M. (ed.). 2008. *From Polysemy to Semantic Change. Towards a typology of lexical semantic associations*. Amsterdam: John Benjamin.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Overlooking Conventions: The Trouble with Devitt's What-Is-Said

ESTHER ROMERO AND BELÉN SORIA*
University of Granada, Granada, Spain

In his forthcoming book, Overlooking Conventions: The Trouble with Linguistic Pragmatism*, Michael Devitt raises, once again, the debate between minimalism and pragmatism to defend the former. He claims that, by taking some overlooked conventions into account, a semantic notion of what is said is possible. In this paper, we claim that a semantic notion of what is said is not possible, especially if some overlooked compositional conventions are considered. If, as Devitt defends, verbal activity is more linguistically constrained, compositional linguistic rules should be included in his catalogue of overlooked conventions and this entails an important challenge to the minimalist claim that the semantic view of what is said can handle all context relative phenomena. In this paper, we argue that, when conventions concerning compositionality are not overlooked, modulation should be added to the two qualifications (disambiguation and saturation) accepted by Devitt in the constitution of what is said. Thus, what is said is not always literally said and the traditional semantic view of what is said cannot be saved.*

Keywords: Linguistic conventions, semantics, pragmatics, what is said, minimalism, linguistic pragmatism.

1. *Introduction*

In this paper we discuss the account of Michael Devitt's notion of what is said in his latest book *Overlooking Conventions: The Trouble with*

* A version of this article was given at the Mental Phenomena course (Dubrovnik 2018) where Michael Devitt's latest book was discussed. We wish to express our gratitude to the members of that audience for various discussions and insightful comments. Thanks also to John Keating for suggestions on an earlier draft. This research was supported by the Spanish Ministry of Science, Innovation and Universities through the project PGC2018-098236-B-I00.

Linguistic Pragmatism (forthcoming)[1] and in one of his previous publications, "Is there a Place for Truth-Conditional Pragmatics?" (2013). In these works, Devitt gives his particular defence of the "traditional view" and argues that the constitution of what is said is "semantic". This traditional view stems from Paul Grice (1975/89) and has also been defended recently by Emma Borg (2012), Kepa Korta and John Perry (2011), Ernie Lepore and Mathews Stone (2015), among others. Although these authors do not agree on everything, they propose the minimalist thesis according to which a sentential utterance has a proposition as its semantic content. That proposition, a minimal proposition, is a complete truth-conditional content obtained simply by virtue of the lexico-syntactic rules and the context required by ambiguous or context-sensitive expressions.[2]

In some of our previous works we have already given arguments against this minimalist thesis. We have rejected Borg's minimalist position arguing that her defence of minimal propositions against pragmatist objections does not serve to avoid other objections which arise from compositional context-sensitivity (Romero and Soria 2019). We have also challenged Lepore and Stone's semanticist claim that pragmatic reasoning never contributes content to utterances (Romero and Soria 2016). Now we turn to Devitt's defence of minimalism against pragmatism. Taking into account that although there is a certain degree of overlap with other minimalists' arguments, Devitt's rejection of pragmatism focuses on an aspect that deserves specific attention: his claim that pragmatists generally overlook some linguistic conventions.

This claim, however, is not entirely new. There are both semanticists, such as Lepore and Stone (2015), and pragmatists, such as ourselves (Romero and Soria 2016, 2019), claiming that there are overlooked conventions, although they are different and both differ from Devitt's. Lepore and Stone claim that there are conventions related to discourse coherence and other aspects of meaning that are linguistically encoded but are not truth-conditional in nature. Devitt focuses on those linguistic rules that demand slot-filling of regular elements of a certain type, which are claimed to form a part of the truth-conditional content of the sentence uttered. Instead, we defend that there are some

[1] From now on when we refer to Devitt without specifying the date of publication, we are making reference to his proposals from a draft (December 2018) of his forthcoming book. We will only specify "forthcoming" when we quote textually from this version (with pages still unavailable).

[2] Strictly speaking, semantic content of a sentential utterance as a minimal proposition opposes non-propositional conceptions of semantic content such as Sperber and Wilson's (1986/95) or Bach's (2006). According to the latter, given a sentence token, it is not possible to determine what state of affairs should obtain for such a sentence to be true. However, Devitt's proposal not only opposes non-propositional conceptions of semantic content but also what is pragmatically said (the notion defended in linguistic pragmatism) since his notion of semantic content, what the sentential utterance says, is what is said by the speaker in case speaker's meaning includes what is said.

compositional linguistic rules which impose certain type constraints in relation to some core elements and which sometimes demand contextual adjustment (modulation or slot-filling) to get the truth-conditional content expressed by the speaker. The focus is clearly different in these three approaches and although our claims affect the other two, they do in different ways.

In this paper, we focus on the way our position challenges Devitt's and we will do so respecting the theoretical and methodological requirements that Devitt recommends. We claim that if the semantic type is taken into account for regular elements as Devitt defends (e.g. the provision of a location in the semantic frame of 'raining' or the provision of a cause in the semantic frame of 'dying'), the constraints imposed by the semantic type of core elements cannot be ignored (e.g. the provision of a sentient participant in the semantic frame of 'waiting', which cannot be the semantic type of a core element in the frame of 'raining', for example). These type-constraints prevent ill-formed compositions of elements such as 'the man is raining from cancer' or 'the table is waiting in Paris' and allow well-formed compositions of elements in a frame as in 'the man is waiting for his check' or 'the man is dying from cancer'. Evidence for these types of compositional regularities can come, as Devitt argues, from corpora elaborated by linguists. Although Devitt does not name any specific corpus, we suggest that Devitt could use FrameNet[3] to support some of his claims about regularities in the frames of 'raining' or 'dying'. However, he has not considered the compositional constraints for the semantic type of core elements in a frame. A frame is a regular schematic linguistic representation of a situation and "[f]rame elements [FEs] that are essential to the meaning of a frame are called "core" FEs (e.g Speaker in frames connected with communication); expressions of time, place and manner are generally not core FEs." (https://framenet.icsi.berkeley.edu/fndrupal/glossary). For example, in the frame of 'telling' there are several core elements of certain semantic types: sentient speaker, sentient addressee and topic. In the frame of 'waiting', its two core elements are sentient protagonist and expected event. The two core elements in the frame of 'raining' are location and time.[4] If Devitt considered these constraints, he would have to admit these types of compositional conventions and their consequences. If he did not, but wanted to be consistent, he would owe us some principled way to accept the demand of the provision of a location in the frame of 'raining' and reject the demand of a sentient participant in the frame of 'waiting'. De-

[3] FrameNet provides annotated examples with information about how words are used in actual texts. It includes more than 13,000 word senses and more than 200,000 manually annotated sentences linked to more than 1,200 semantic frames. It includes detailed evidence for the combinatorial properties of a core set of the English vocabulary.

[4] Evidence of these two frames can be found respectively at https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Waiting and https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Precipitation.

vitt, however, does not and we think he cannot provide us with such a principled distinction and thus he must accept both. In cases where the semantic types are not provided, they demand determination in context and although the type of contextual adjustment is often slot-filling (as in the provision of a location for a raining event), in certain cases, it must be conceptual modulation as we will see.

Thus, even if we think, as Devitt does, that both pragmatists and semanticists have overlooked some linguistic conventions (Romero and Soria 2013, 2019), unlike him, we think this is a reason to have a pragmatic notion of what is said rather than a semantic one. The reason is simple: some compositional linguistic rules or conventions of the type that Devitt proposes to add sometimes demand modulation to get the truth-conditional content of the sentence uttered. Thus, disambiguation and slot-filling (or saturation) are not the only ways of exploiting linguistic conventions. As we have defended (Romero and Soria 2013, 2019), modulation may be obligatory and without it, not even saturation is possible in certain cases, cases in which slot-filling is dependent on modulation. This is a very serious challenge for the traditional view that Devitt is trying to save since, if it is right, what is said is not always literally said.

This paper is divided in two sections. In the next section, we present Devitt's proposal on what is said and the way in which Devitt articulates it. In the third section, we focus on the challenges to Devitt's semantic notion of what is said. Our disagreement with him leads us to provide the arguments for our defence of a pragmatic notion of what is said. Finally, we present our conclusions.

2. *Devitt's proposal on what is said*

According to Devitt, the study of language is theoretical and empirical and this has two consequences. First, we have to analyse theoretically interesting notions of meaning: a favoured notion of what is said and a notion of what is meant. Second, we need direct evidence from linguistic usage in favour of these notions and not intuitions which are themselves theory-laden and open to question.

We have a theoretical interest in human languages as representational systems constituted by a set of governing rules that people use to communicate the contents of their inner states to each other. These rules are largely conventional: symbols have their meanings by convention. Conventions associated with a linguistic form emerge from the regular use in the community of that form to convey certain parts of messages. The regular use of a linguistic form in utterances with a certain speaker meaning leads to that form having that meaning conventionally in the language of that community.[5] The regular use gives

[5] This theoretical approach to meaning is similar to the strategy initiated by Grice (1957/89, 1968/89)'s or Schiffer (1972)'s.

us evidence of linguistic conventions, of semantic properties, if they provide the best explanation of regularities. We can gather evidence about regularities from linguistics since linguists often acknowledge the role of usage as a source of evidence: in the study of corpora and elicited production.

The regular use of a linguistic form in utterances with a certain speaker meaning somehow leads that form to have that meaning (or part of that meaning) conventionally in the language of that community. For example, the conventions associated with (1)

(1) It's raining

come from the regular use of (1) to communicate messages such as that it is raining in Granada, that it is raining in NY, that it is raining in Dubrovnik, etc. When the speaker believes that it is raining in Granada and she is in Granada, she (in English) produces (1) and this token of (1) means that it is raining in Granada. That meaning is the message the speaker intentionally communicates, her "speaker meaning", when she is being literal and straightforward. Conventions in these cases make reference to what is regularly included, that it is raining [in some place to be determined]. These rules show that a theoretically interesting what is said, a what-is-said that may be the content of a mental state, is "very tainted" in context. Some linguistic rules demand contextual "saturation"; a "slot" should be filled as in example (1). The very frequent provision of a location in the frame of 'raining' can be taken as evidence that it obeys a linguistic rule, it is clearly a linguistic regularity recognized in FrameNet. (1) has its representational properties partly by virtue of the place where it is raining, by virtue of something that is not encoded.

Example (1) is similar to examples that involve words with an indexical or demonstrative element. Their linguistic rules demand saturation in context. For example, the linguistic rule associated to 'that' captures the convention for expressing the demonstrative part of a thought, its encoded meaning, and according to it, a token of the demonstrative 'that' in an utterance of (2)

(2) That is red

"refers to whatever object is linked to it in the appropriate causal-perceptual way" (Devitt forthcoming). So the token of 'that' in (2) has its representational property partly by virtue of something that is not encoded, an apple, for example. The demonstrative in (2) straightforwardly semantically designates the apple (in the situation): in using 'that' the speaker had that apple in mind by virtue of her thought being causally grounded in it. Having the apple in mind in using 'that' simply requires that the part of the thought that causes that use of 'that' refers to the apple in question. What makes an object the referent of 'that' is its causal relation to the part of the thought expressed by (2). The reference of 'that' is determined by a mental state of the speaker. What is

said is often partly constituted by whatever determines the reference of any word with an indexical or demonstrative element.

Sometimes more than one linguistic rule governs a symbol. This multiplicity arises from multiple conventions for the linguistic form. Multiplicity of conventions demands disambiguation and what is said takes one of those meanings. In an utterance of (3)

(3) He went to a bank

The speaker is participating in one of the two conventions for 'bank'. Disambiguation is needed to arrive at the representational properties that are of theoretical interest. This also shows that a theoretically interesting what is said is "very tainted" in context.

The same must be said of (4).

(4) Visiting relatives can be boring

We are interested in which of the two conventions for 'visiting relatives' the speaker is participating in. The explanatory role of a particular linguistic form (simple as 'bank' or complex as 'visiting relatives') depends on which rule has been exploited.

Rules related to saturation and disambiguation are in the speaker and they are not inferential nor, in any interesting sense, pragmatic. They contribute to the theoretically interesting what is said, which, although it is "very tainted" in context, is not pragmatic. The distinction between what is said and what is meant guides, according to Devitt, the semantics-pragmatics debate.

Taking into account examples (2)–(4), what is said departs from the conventional meaning of the sentence when saturation is needed or when disambiguation is involved. Saturation and disambiguation are linguistically demanded. Devitt and pragmatists do not disagree on that, although they disagree in the way disambiguation and saturation is reached. According to Devitt, the intentional act that is necessary for disambiguation and saturation is not an act of communicating a thought as linguistic pragmatists argue but one of expressing a thought.

However, the main point of disagreement comes from their different views on some context relative phenomena such as the utterance of (1) to say that it's raining in Granada. For pragmatists truth-conditional content depends not only on processes linguistically demanded (mandatory) but also on processes non-linguistically demanded (optional) such as the pragmatic enrichment required for (1), a case of unarticulated constituent. For Recanati, in (1) there is no linguistic demand for the provision of a location. The demand is pragmatic through and through and yet it is part of what the proposition explicitly communicated, it is part of what is pragmatically said. For Devitt, on the contrary, every contextual influence in what is said by an utterance is necessarily taken to be linguistically demanded. (1) linguistically demands slot-filling and not a pragmatic enrichment. From his traditional semantic view (or minimalism), the context-relative phenomena that motivate

linguistic pragmatism (or contextualism) can be handled by including previously overlooked conventions without abandoning its main tenet

> that a sentential utterance has its truth-conditional content simply […] in virtue of the conventional rules of the speaker's language. This content is typically thought to be "what is said" by the utterance and its constitution is thought to be a "semantic" matter. (Devitt 2013: 86)

Devitt's recognition of context-relative phenomena can be accounted for in the minimalist tradition by means of just two qualifications. If an expression is ambiguous, its contribution to what is said will depend on which of its meanings the speaker "has in mind". When an utterance contains an indexical, what is said depends on reference fixing. For Devitt, the constitution of what is said is "semantic" since the representational properties provided by the linguistic rules only demand determination in context in cases of disambiguation and slot-filling and there are no purely pragmatic effects on the truth-conditional content said.

The traditional semantic view has been questioned by pragmatists taking into account (1) and other examples such as

(5) The table [in my room] is covered with books
(6) I've had breakfast [this morning]
(7) You are not going to die [from that minor cut]

in which, according to them, pragmatic enrichment (in square brackets) is needed to go from semantic content to what is said. These sentences in context mean what their words mean together with what is marked in brackets, while they say something else literally. (5) says the absurd claim that there is one and only one table and it is covered with books, (6) says that the speaker has had breakfast [sometime in the past], (7) attribute immortality to the addressee. As what is said by means of these sentences does not coincide with what is meant in context, these examples show that "pragmatic" enrichment is needed to get what is said from what is "semantically" determined.

Nevertheless, as we have seen, Devitt claims that example (1) does not require free pragmatic enrichment to get what is said. (1) would be a case of slot-filling and thus the result of contextual determination is a result of a linguistic demand. Evidence that the provision of a location obeys a linguistic rule can be found in its very frequent use, a very clear linguistic regularity. Similarly, there are conventions for expressions included in (5)–(7) that demand saturation in context. The convention associated with the referential use of 'the table' comes from the regular use of (5) to refer to the particular object the speaker has in mind, as also happened in the case of 'that' in example (2). The conventions associated with (6) come from the regular use of (6) to say that the speaker has had breakfast [sometime in the past to be determined]. The past tense of the verb phrase requires determination of a specific past time in context, it must be saturated in context, for example, with this morning. What is semantically said by the utterance of (6) is that

the speaker has had breakfast this morning. To explain Bach's famous example (7), Devitt also claims that there is a convention demanding slot-filling. In (7) there is not an indexical such as tense in (6) but there is a regular provision of the cause of the death in the use of sentences with 'die' as the main verb. To make this claim, however, Devitt should provide us with some sort of evidence. According to FrameNet, in the frame for 'dying' there are at least two frequent elements: a core element (sentient protagonist) and a non-core element (situation or event that led directly to the death).[6] Even if it is considered a non-core element, the situation or event that led directly to the death is a regular sort of information. This type of regularity can be taken as evidence for Devitt to argue that there is a normal disposition of the speaker to include the event that would (or would not) lead to the death in what is said, and thus, in an utterance of (7), what '[from that minor cut]' means can be considered as part of what is said.

In this way, the slot-filling involved in examples such as (1) and (5)–(7) can be considered semantic and, according to Devitt, linguistic pragmatism loses one of its main motivations. Devitt's criticism of pragmatists is that their enlarged what is said is partly "pragmatic" (2013: 96). By adding overlooked conventions that demand slot-filling, he also defends that there is an enlarged what is said in these examples but, since slot-filling is for him a semantic process, his notion of what is said is still semantic.

Linguistic pragmatism, Devitt admits, also takes into account some phenomena that demand enrichments or impoverishments of what is said such as

(8) The *burglar nightmare* was over
(9) a. Max *cut* the grass
    b. Max *cut* the cake
(10) The ATM *swallowed* my credit card

Utterances like (8)–(10), in a context, can convey a more precise or less precise message than the semantic what is said. These messages are achieved by enrichment and impoverishment. The reasons for that, according to Devitt, may be that it is ponderous and boring to communicate the precise message using conventions, as in (8), or that the only available conventions determine a meaning that is vaguer or narrower than the desired message, as in (9) and (10) respectively.

The truth-conditional content expressed by (8) is an imprecise what is said. The imprecise what is said, according to Devitt, would be that whatever the relation between burglar and nightmare denotes, the burglar nightmare is over. What 'burglar nightmare' would thus contribute would be rather imprecise but it will provide the needed constraint: anything that is to count as a burglar nightmare has to be of that imprecise kind. This constraint is a convention that determines a

[6] See at https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Death.

vague truth condition that the speaker enriches in a context to get a more precise message; the speaker conveys the precise proposition she means with the help of the imprecise proposition she expresses.

The truth-conditional content expressed by (9) is also an imprecise what is said. 'Cut' is seen as referring to what is common to cutting grass, cutting cakes, and all other forms of cutting. So, as Devitt (2013: 96, forthcoming) says following Hale and Keyser (1987), it means "something along the lines of *produce linear separation in the material integrity of something by a sharp edge coming in contact with it*". What 'cut' would thus contribute would be rather imprecise but it could provide the needed constraint: anything that is to count as a cutting action has to be of that rather vague kind.

The enrichment in (8) and (9) is pragmatic. A "pragmatic" mechanism needed to get from what is said to a potential message that is an expansion of a semantic what is said; a semantic what is said that is truth-conditional and thus truth-conditional pragmatics is not needed. In these cases, what is said follows from what is meant.

Impoverishment occurs, according to Recanati (2004: 26), in a token of (10). The proposition meant is *less* precise than the proposition said. Devitt follows him in the impoverishment proposal but although (10) may once have been a case of impoverishment, he thinks (10) is now a dead metaphor and thus disambiguation is the strategy involved. For a pragmatist such as Recanati, it is a case of modulation affecting what is said. By contrast for Devitt, if it really were a case of impoverishment, it would be a case of modulation external to what is said.

Devitt's putative solutions for explaining what is said by means of examples (1) and (5)–(7) are of no use to provide an explanation of (8)–(10) and he grants a role for pragmatics in their explanation. Each of their contents is characterized as what is (semantically) said + pragmatic modulation (2018: 47, forthcoming). As this type of content is characterized in part pragmatically, it represents occasional features of linguistic communication.

Linguistic pragmatism also takes into account a metaphorical use of examples such as (11),

(11) The rock, now becoming brittle with age, responds to his students' questions with none of his former subtlety (adapted from Kittay 1987: 71)

or metonymical uses of examples such as (12) and (13)

(12) The beer faucet is waiting for her second 'tapa' (a real utterance of a waitress referring to Belén Soria in a tapas bar in Granada when she was sitting at the counter by the beer faucet)[7]

[7] This is a novel metonymy similar to the now classical example by Geoffrey Nunberg (1979: 149), 'the ham sandwich is waiting for his check'. Devitt considers this example a case of conventional metonymy which should be explained by regular polysemy. Cases like this exemplify "*meta*-conventions, processes for generating lexical conventions, of the following form: wherever a convention is established

(13) There is a lion in the middle of the piazza (taken from Recanati 2010: 5)

that are not included in Devitt's list. These examples hold, according to pragmatists, what is pragmatically said. 'Rock', 'beer faucet', and 'lion' contribute to what is pragmatically said with a modulated meaning in the first case and with extended complex concepts in the other two cases.

Nevertheless, Devitt does not consider that these examples challenge his view of what is said. They are cases in which what the speaker means differs from what is said. They convey contents external to what is said: they are implicatures, non-literal contents. The speaker says something she does not mean as a way of conveying something that she does. Devitt handles (11)–(13) arguing that what is said does not have to be meant. As he says "the fact that *p* is *what is said by* an utterance does not entail that *p* is *meant by* the utterance (does not entail that *p* is the utterance's *message*)" (forthcoming). The metonymical utterance of (13), for example, cannot constrain truth conditions different from its literal ones. A token of (13) says that there is a lion in the middle of the piazza and this semantic content is not included in what the utterance means, that there is a lion statue in the middle of the piazza. What is meant does not coincide with what is said. The utterance has pragmatic properties. This would be a case of implicature and thus it is meant non-literally and indirectly.

In sum, Devitt (2018: 47, forthcoming) has a four-way distinction among the properties of utterances: encoded conventional meaning; what is said (as a result of encoded conventional meaning, disambiguation and reference assignment); what is said + pragmatic modulation; and implicatures. And he considers that two notions of meaning are theoretically well based: what is meant and what is said. The ways in which what is meant goes beyond encoded conventional meaning includes the types of contents shown in Figure 1:

Figure 1. *Devitt's catalogue of utterance contents*

<table>
<tr><td rowspan="3">What is meant</td><td colspan="3">= what is said</td></tr>
<tr><td rowspan="2">≠ what is said</td><td rowspan="2">what is said + pragmatic modulation</td><td>…+ enrichment</td></tr>
<tr><td>…+ impoverishment</td></tr>
<tr><td colspan="2">what is implicated by means of indirect or figurative uses</td></tr>
</table>

that an expression refers to things of type X that expression will also *thereby* refer conventionally to things of related type Y." (forthcoming). However, (12) is not one of the conventional types of regular polysemy and needs a pragmatic explanation.

In general, there are two possibilities for the notion of what is meant. First, what is meant by the uttering of a sentence coincides with what is said by the utterance. Then the utterance has only semantic properties. The speaker is being literal and straightforward as in Devitt's explanation of example (1) above. His what-is-said includes many overlooked conventions that linguistically demand slot-filling as in examples (5)–(7). Second, what is meant by uttering a sentence does not coincide with what is said by the utterance. Then the utterance has pragmatic properties. As what is meant can be constituted in two different ways, there are two types of contents with pragmatic properties: what is said + modulation and implicatures. While implicatures are purely pragmatic properties, what is said + modulation is a type of content with properties that are only in part pragmatic, those related with the result of modulation. Thus, there are different ways in which what is meant can depart from what is said:

> – the proposition meant is a precise proposition with the help of the imprecise proposition said as examples (8)–(9) show
> – the proposition meant is less precise than the proposition said as in (10)
> – the proposition meant is a conversational implicature as in (11)–(13).

3. *Challenges to Devitt's semantic notion of what is said: Overlooking conventions*

Devitt's semantic notion of what is said includes conventions demanding contextual information which are generally overlooked by both semanticists and pragmatists. We are afraid that by adding the kind of overlooked conventions that are involved in utterances of (5)–(7), Devitt should also include much more in what is said since he has to take into account generally overlooked compositional conventions related to the metaphorical utterance of (11) or to the metonymical utterance of (12). Devitt's proposal faces an important challenge with examples of metaphor and metonymy. This challenge arises because all his requirements to save the tradition are really not compatible. It is inconsistent to maintain that compositional conventions should not be overlooked and to reject non-literal contents as part of what is said in examples such as (11). The strategy Devitt follows to defend his overlooked conventions, would lead him, in our opinion, to include in what is said more than just the result of disambiguation and saturation if compositional conventions are taken into account. The properties that an utterance may have as a result of the speaker's exploitation of her language arise not only from encoded conventional meaning together with disambiguation and reference fixing but also from modulation. Thus a sentential utterance has its truth-conditional content not simply by virtue of the (largely) conventional rules of the speaker's language with two impor-

tant qualifications, saturation and disambiguation, since at least one other qualification should be included. Some generally overlooked compositional conventions often demand modulation to get a proposition, both from the point of view of production and interpretation.

Let's suppose, for the sake of argument, that Devitt is right about the meaning-properties of utterances (1)–(7) and his explanation of them. Do we also have to accept that (8)–(10) are adequately explained by what is said + pragmatic modulation? To know if examples such as (8)–(10) must be considered as cases of what is said + pragmatic modulation we need to know first what truth-conditional contents are obtained with (8)–(10) that constitute what is said by these utterances and, second, what expressions within (8)–(10) are undergoing pragmatic modulation. As we are going to show it would become more consistent for Devitt if saturation in context were used to explain how to get the type of what is said by (8) and if disambiguation in context were used to explain how to get the type of what is said expressed by (10). The only case of what is said + pragmatic modulation would be (9) but it is not clear to us why this example is not a case of implicature for him. Let us see the problems for Devitt's treatment of cases (8)–(10).

For us, there is no truth-conditional content obtained with (8) that constitutes an imprecise proposition. The problem for Devitt's proposal of an imprecise proposition expressed by (8) is that without specifying the relation of burglar with nightmare in context, the restrictive modifier cannot constrain the denotation of 'nightmare' and thus 'burglar' is not performing its linguistic task. The speaker cannot have in mind an imprecise proposition expressed by (8) since there is nothing in common between the nightmare the burglar has about something and the nightmare that a person has about the burglar and thus what semantics delivers for 'burglar nightmare' will not be an imprecise part of a proposition. There is no imprecise proposition capable of truth evaluation, something the speaker can think of.

If this is so, the content of (8) is merely a set of propositional constituents that has not admitted semantic composition since some subpropositional context-dependent component of content (the relation between the content of the two nouns) is missing. (8) is similar to (1) and 'burglar nightmare' expresses a constituent of what is said that results from a *semantic* addition demanded by a convention exploited by the speaker, the convention for N+N construction. The speaker of (8) participates in a convention with the use of 'burglar nightmare' as far as what is regularly delivered by the semantics of this N+N construction is the meaning of 'nightmare [in some relation with] burglar [to be determined]'. But this does not determine a vague part of a truth condition that the speaker enriches to convey a more precise message. Thus, 'burglar nightmare' in a token of (8) is not a case of modulation. Neither the meaning of 'nightmare' nor the meaning of 'burglar' undergoes pragmatic enrichment. Devitt's explanation of the content conveyed by

the utterance of (8), what is said + pragmatic modulation, is not plausible. We think it would be more consistent for Devitt to argue that (8) is a context-relative phenomenon more similar to (1) and (5)–(7) than to cases of what is said + pragmatic modulation in the sense that its convention in relation to the N+N construction establish a slot to be filled in context: the relation that nightmare bears to burglar.

(9) is also considered by Devitt as a case in which its truth-conditional content is imprecise and becomes precise by the pragmatic enrichment of 'cut'. Although it is an extension or elaboration of a constituent of the proposition said, Devitt considers it external to what is said. In this way, he can maintain his main point: "the semantic what-is-said that is thus expanded is already truth-conditional and so there is no place here for "truth-conditional pragmatics"." (forthcoming). However, why are utterances of (9a) and (9b) considered cases of what is said + pragmatic modulation rather than implicatures?

Let's look at an utterance of (10). Although he said that this is a case of impoverishment, he also claims that this utterance is a case of conventional metaphor. Thus, Devitt thinks it expresses a truth-conditional content that depends on disambiguation and constitutes what is semantically said by this utterance; it literally says that the ATM swallowed the credit card. In this sense, Devitt does not provide us with a good example of impoverishment.

We could use examples of novel metaphors such as the metaphorical utterance of (11), which, according to Recanati (2004), is a case of impoverishment as well. However, for us, no truth-conditional content is obtained with an utterance of (11) that constitutes what is said and has to be impoverished. According to our conventions, in (11), 'responds' should express a property of animate beings (a sentient speaker).[8] This is similar to Devitt's claim that there are conventions demanding slots to be filled by a location in (1), a time in (6), a cause in (7) and, as we say he should admit, a relation in (8) to get the truth-conditional content. In Devitt's slot-filling proposal, the slots must be filled with entities of a certain semantic type if they are to count as conventions to get the truth-conditional content. Type constraints are part of linguistic regularities as we can see in FrameNet. Taking into account the evidence from this corpus, *location* is the regular type to fill the slot in (1) and *cause* is the regular type to fill the slot in (7) but the cause cannot fill in the slot in (1). When these type constraints affect core-elements they can be taken as linguistic rules (Asher 2011, Romero and Soria 2019). For instance, the verb 'wait' demands a sentient participant as subject of the VP in the active form. If we take into account these compositional linguistic rules, composition of the semantic constituents of (11) is precluded by normal type constraints and a pragmatic adjustment is demanded.

[8] Evidence of this frame can be found at https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Communication_response.

In addition, the pronoun 'his' carries gender and number features which constrain the antecedent of the anaphor and which must be masculine. However, the object referred to by the token of the NP, the rock, is not of the semantic type required by 'responds' and cannot be an acceptable antecedent of 'his'. (11) shows lack of semantic coordination between the meaning of the NP, the rock, and the meaning of 'responds' and thus their composition is not possible. Thus, no resulting truth-conditional meaning can be expected to represent a thought with both of them as constituents. The speaker is not doing, Devitt would say, "what she is normally disposed to do." She is "deliberately assigning another meaning to an expression, as in metaphor or pragmatic modulation" (forthcoming). In his defence of the tradition, however, he rejects that this difference in meaning may affect what is said. By contrast, we claim it does and examples like (11) challenge his defence in a serious way since modulation of the meaning of 'rock' is here necessary for the slot-filling of 'his'.

In (11), 'his' demands a slot-filling through anaphor resolution and anaphor resolution is guided by linguistic rules of agreement. This agreement is possible in (11) only with the modulated meaning the speaker has in mind rather than with the encoded meaning. If the speaker had used 'the rock' to refer to a rock, the speaker would have uttered 'the rock (…) responds to *its* students' rather than 'the rock (…) responds to *his* students' to get the agreement that the rules of language require. However, the speaker of (11) uses 'the rock' to refer to the old professor she has in mind and this partly determines its meaning in the appropriate causal-perceptual way. The old professor behaviour has prompted the speaker to conceive the professor metaphorically as a rock getting brittle with age and the best way to represent the metaphorical thought she has in mind is with the metaphorical utterance of (11). The metaphorical use of 'the rock' is causally grounded in the speaker's metaphorical conceptualization of the professor and by her use of 'his' rather than 'its', she participates in a convention grounded in this metaphorical conceptualization. In cases like this, saturation depends on modulation. If the truth-conditional meaning of an indexical is partly determined by what the speaker has in mind and what she has in mind is a metaphorically conceptualized professor, she produces a metaphorical utterance through a regular mechanism, the metaphorical, which is quite systematic in language use. In (11), the speaker selects 'his' rather than 'its' to represent her metaphorical thought. Thus, modulation cannot be external to what is said on pain of ungrammaticality. (11) is a well-formed metaphorical utterance and it would be ill-formed if taken literally. We do not think the content of an ill-formed literal utterance corresponds with what the speaker has in mind. In (11) what is said is metaphorically said. This can be claimed if we accept that compositional conventions demand modulation to solve the lack of semantic coordination. But by adding this type of conventions we are opposing both pragmatists and Devitt's traditional view. For all of them, the derivation of metaphorical meanings is never lin-

guistically demanded. But if metaphorical modulation is optional, what is the propositional literal content that the speaker has in mind in (11)? A rock which has students and can respond to their questions? We think this is inconsistent for Devitt if compositional rules are recognised conventions. Lack of semantic coordination indicates context-dependence which demands pragmatic adjustment. Conventions tell us that a sentient entity is needed to be able to compose a full content for (11). The speaker is deliberately assigning some (abnormal) meaning to 'rock' so that composition is allowed in a regular way. The resolution of this compositional context-dependence cannot be treated as part of semantics. A pragmatic process is needed to make composition possible since although the speaker participates in the convention when using the word 'rock' she is exploiting it metaphorically to express the metaphorical concept the speaker has in mind, the speaker is also participating in the compositional conventions by her selection of the verb 'respond' and the pronoun 'his' to coordinate semantically with the metaphorical conceptualization the speaker has in mind.

Without the modulated meaning, there is no literal proposition for (11), no impoverishment (or any other type of modulation) of the concept ROCK can be added as something external to what is said. Recanati is right when he includes modulation in what is said and argues for what is pragmatically said. Where Recanati is wrong is in his defence of impoverishment as the result of a pragmatic process that is always optional. Modulation in the utterance of (11) is compositionally and linguistically demanded. If this is so, to handle (11), Devitt's theoretically interesting notion of what is said has to be modified since this utterance does not have a truth-conditional content simply by virtue of the conventional rules of the speaker's language, disambiguation and saturation.

However, Devitt might defend his position by saying that (11) is not a case of impoverishment but of transfer and that transfer is involved in implicature. (11) would not be a challenge for his what-is-said + pragmatic modulation after all. This defence has two problems, though. The first problem is that if transfer is involved in implicature, we do not understand why (9), a case of enrichment, is not also a case of implicature. Transfer together with enrichment and impoverishment are the optional pragmatic processes that characterize the notion of what is pragmatically said. If transfer goes to implicature, enrichment and impoverishment should go too. The reason, we suppose, why enrichment, a type of content external to what is said, is not included in implicature is that modulation affects the meaning of a word and not the meaning of the uttered sentence but this also happens with transfer. The second problem is that implicature and what is said + pragmatic modulation presuppose in Devitt's theory a semantic what is said, but in (11) what is said cannot be obtained without pragmatic modulation.[9]

[9] Devitt does not include transfer as a case of modulation. If transfer is involved in implicature and it is characteristic of metaphor, why does Devitt understand

As whatever process is involved in metaphor, it is compositionally and linguistically demanded to get what is said in cases such as (11), the result of impoverishment or transfer in these cases is not added to what is said simply because there is no literal what is said. In cases such as (11) the result of these pragmatic processes is not an implicature for the same reason. The result must be included in what is said; a proposition said that can be an input for implicatures. To the extent that in (11) there is no semantic proposition said and that what the speaker has in mind is a metaphorical thought, there are reasons to think that in those cases the proposition is a metaphorical proposition and what is said is metaphorically said. The proposition said is non-literal and this is not acceptable for the traditional view that maintains that what is said is always literally said.

Example (12) shares some properties with (11). The type of core elements involved in the frame of waiting are a sentient protagonist and an expected event. The pronoun 'her' carries gender and number features which constrain the antecedent of the anaphor that must be feminine (but no acceptable antecedent is expressed). (12) shows lack of semantic coordination between the meaning of the NP, 'the beer faucet', and the conventional constraints imposed by 'waiting' and 'her'. No explicit NP of the type required (sentient protagonist) is expressed and no feminine acceptable antecedent for 'her' is expressed. Thus composition of their encoded meanings is not available. There are conventions for (12) as complex expressions that compositionally and linguistically demand contextual information. There is compositional context-sensitivity.

What expressions in (12) are undergoing modulation according to pragmatists? If we follow pragmatists such as Recanati, 'beer faucet' has to undergo transfer. Nonliterality is attributable to a specific expression, 'beer faucet', and its meaning is what must be changed. 'Beer faucet' non-literally means 'beer faucet customer'. Conventions tell us that a sentient entity is needed to be able to compose a full content for (12). The speaker is deliberately assigning some (abnormal) meaning to 'beer faucet' so that composition is allowed in a regular way.

We think instead, in a spirit more coherent with Devitt's proposal, that saturation in context can be used to go from a non-propositional semantic content to what is said in (12). The meaning of 'the beer faucet' must work as part of the restrictive modifier of [customer] to say that the customer by the beer faucet is waiting for her second tapa. A new kind of slot-filling appears. 'Beer faucet' means 'beer faucet' and 'the beer faucet' has in (12) its representational property partly by means of the object the speaker has in mind, the customer by the beer faucet. In Devitt's vein, it could be said that this utterance has a truth-conditional content simply by virtue of the conventional rules of

impoverishment, the process involved in novel metaphor according to Recanati, as a case of what is said + modulation? His two proposals on metaphor are not coherent.

the speaker's language, disambiguation and saturation. Nevertheless, what is semantically said by (12) is non-literal. Again we have a case of what is non-literally said.

In order to handle (11)–(12), Devitt's theoretically interesting notion of what is said needs to include either modulation or a new kind of slot-filling. In these cases, the pragmatic adjustment is demanded conventionally due to compositional context-sensitivity (Romero and Soria 2013 and 2019) and it has to be included in what is (pragmatically and non-literally) said. At the end of the day, the properties of these types of utterances are linguistically demanded but inevitably pragmatic.

Additional evidence to show that these properties are inevitably pragmatic is that the metonymical property of the utterance of (12) should be of the same nature as the metonymical property of any metonymy. This means that the metonymical use of (13) should include a non-literal and expanded meaning for 'lion' as a result of the new kind of slot-filling. However, a token of (13) does not conventionally demand this slot-filling and any contextual effect not conventionally (or optionally) demanded by an utterance constitutes without any doubt a pragmatic property. A similar argument could be made if we think of a metaphorical use of (13).

4. *Conclusion*

In this paper we have explained Devitt's semanticist position and how it depends on including more conventions in the constitution of what is said. Although we have also defended that there are *overlooked conventions*, we cannot agree with Devitt's "semantic" notion of what is said since the properties that an utterance has simply as a result of the speaker's exploitation of the conventions of her language sometimes demand contextual modulation and not only disambiguation and saturation.

We have argued that modulation may be demanded by conventions which constrain the compositionality of complex expressions (phrases and sentences) and that Devitt's way of delineating the constitution of what is said by including more conventions leads him further than he would be ready to accept in his defence of traditional truth-conditional semantics for which what is said is always literally said. Some of the overlooked conventions Devitt is trying to highlight have a compositional character and if he attempts to include them in the semantic constitution of what is said without being unsystematic, he should provide a principled way to justify why certain types of exploitation of linguistic conventions are accepted in the determination of what is said (disambiguation and saturation) and others are not (modulation) and why certain types of conventions are accepted in some cases (e.g. the provision of a location in the frame of 'raining') while they are rejected in other cases (e.g. the requirement of a sentient protagonist in the frame of 'waiting'). If there is not a principled way to discard certain

compositional conventions, the wider semantic notion of what is said that he proposes will not be viable.

If the theoretical reason to include the resolution of context-sensitivity in what is said is that, if we didn't, there would be no way to attribute the speaker a thought in order to explain his behaviour, and the overlooked conventions include (as we argue) compositional demands for contextual (non-linguistic) information, Devitt should accept that modulation and certain types of slot-filling (excluded in the traditional semantic approach) must be included in what is said. In certain cases, the semantic role of a core-element can match other constituents only if there is a conceptual adjustment of the encoded meaning as in metaphorical modulation or a metonymical slot-filling. In these cases, what semantics delivers is not something that the speaker can mentally represent as something capable of being true. This is especially evident when saturation is dependent on modulation. If what the speaker has in mind is a metaphorical thought, the encoded meaning undergoes modulation through transfer and the speaker expresses a metaphorical truth-conditional content. If what the speaker has in mind is a complex concept and expresses it by means of a sub-phrasal constituent of a sentence, the speaker is expressing part of a thought metonymically. In both cases, what is said is non-literally said. However, if compositional conventions demand that the result of metaphorical modulation and metonymical complex concepts are included in what is said, we think that even Devitt would find it better to call the resolution of context-sensitivity "pragmatic" rather than "semantic". If we want to give a systematic account of the properties of utterances, we need a pragmatic notion of what is said.

References

Asher, N. 2011. *Lexical Meaning in Context: A Web of Words*. Cambridge: Cambridge University Press.

Bach, K. 2006. "The Excluded Middle: Minimal Semantics without Minimal Propositions." *Philosophy and Phenomenological Research* 73: 435–442.

Borg, E. 2012. *Pursuing Meaning*. Oxford: Oxford University Press.

Devitt, M. 2013. "Is There a Place for Truth-Conditional Pragmatics?" *Teorema: Revista Internacional de Filosofía* 32 (2): 85–102.

_____, 2018. "Sub-Sententials: Pragmatics or Semantics?" In A. Capone, M. Carapezza and F. Lo Piparo (eds.). *Further Advances in Pragmatics and Philosophy*. Cham: Springer: 45–64.

_____, forthcoming. *Overlooking Conventions. The Trouble with Linguistic Pragmatism*.

Grice, H. P. 1957/89. "Meaning." In H. P. Grice 1989: 213–223.

_____, 1968/89. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning." In H. P. Grice 1989: 117–37.

_____, 1975/89. "Logic and Conversation." In H. P. Grice 1989: 22–40.

_____, 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Hale, K, and S. J. Keyser 1987. "A View from the Middle." *Lexicon Project Working Papers* 10, Center for Cognitive Science, MIT, Cambridge, Mass.

Kittay, E. F. 1987. *Metaphor. Its Cognitive Force and Linguistic Structure*. Oxford: Clarendon Press.

Korta, K. and J. Perry 2011. *Critical Pragmatics: An Inquiry into Reference and Communication*. Cambridge: Cambridge University Press.

Lepore, E. and M. Stone 2015. *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Nunberg, G. 1979. "The Non-uniqueness of Semantic Solutions: Polysemy." *Linguistics and Philosophy* 3: 143–84.

Recanati, F. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

_____, 2010. *Truth-conditional Pragmatics*. Oxford: Clarendon Press.

Romero, E. and B. Soria 2013. "Optionality in Truth-Conditional Pragmatics." *Teorema: Revista Internacional de Filosofía* 32 (2): 157–74.

_____, 2016. "Against Lepore and Stone's Sceptic Account of Metaphorical Meaning." *Croatian Journal of Philosophy* 16 (47): 145–72.

_____, 2019. "Semantic Content and Compositional Context-Sensitivity." *Theoria* 34 (1): 51–71.

Schiffer, S. 1972. *Meaning*. Oxford: Clarendon Press.

Sperber, D. and D. Wilson 1986/95. *Relevance: Communication and Cognition*, 2nd ed. Oxford: Blackwell.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Speaker's Reference, Semantic Reference, and the Gricean Project. Some Notes from a Non-Believer

ANDREA BIANCHI
University of Parma, Parma, Italy

In this paper, I focus on the alleged distinction between speaker's reference and semantic reference. I begin by discussing Saul Kripke's notion of speaker's reference and the theoretical roles it is supposed to play, arguing that they do not justify the claim that reference comes in two different sorts and highlighting that Kripke's own definition makes the notion incompatible with the nowadays widely endorsed Gricean project, which aims at explaining semantic reference in terms of speaker's reference. I then examine an alternative account of speaker's reference offered by Michael Devitt within his causal theory and express some doubts about its suitability for explaining proper name semantic reference. From all this, I conclude that there is at least some tension between Kripke's chain of communication picture and the attempt to explain (Griceanly, so to say) semantic properties in terms of speakers' mental states.

Keywords: Speaker's reference, Semantic reference, Kripke's distinction, Speaker's meaning, Grice's program, Devitt's causal theory of proper names.

1. *Introduction*

Does the mind of the speaker play any role in determining the reference of the proper name tokens he or she produces?[1] For most of the philosophers who subscribe to what Keith Donnellan (1970) called "the

[1] I first raised this question in Bianchi 2012, to contrast two ways of being referentialist, and more generally two models of the functioning of language, which I then called the *psychological model* and the *social model*. (I now prefer to call the latter the *linguistic model*, since it is based on highlighting the (semantic) *autonomy* of language from users, which is something that may in principle obtain, *pace* Wittgenstein, even if there is a single user and no social relation at work.)

principle of identifying descriptions," it does.[2] According to them, in fact, in order to refer with a proper name, the speaker must attach (in his or her mind, so to say) a set of identifying descriptions to it, and the referent of the token he or she produces, if there is one, is "that object that uniquely fits a 'sufficient' number of the descriptions in the set" (Donnellan 1970: 339). If the speaker attached a different set of identifying descriptions to the name, the produced token might refer to something else. Therefore, the reference of the token does crucially depend on the mental state of its producer, in this view. Basically, by using a proper name the speaker would be referring to something *because* he or she would be *thinking* of it through a set of identifying descriptions.

However, the principle of identifying descriptions is not very fashionable nowadays, and for quite good reasons. In fact, devastating criticisms to any approach to proper names based on it were offered around 1970 by Donnellan himself as well as, of course, Saul Kripke. As a consequence, various philosophers of language, developing suggestions from the work of both Kripke and Donnellan, began to advocate *historical*, if not altogether *causal*, accounts of proper name reference. Since these accounts highlight the crucial role played in determining reference by worldly historical facts that may be unknown to the speaker (as David Kaplan wrote, "[t]he notion of a historical chain … [offers] an alternative explanation of how a name in local use can be connected with a remote referent, an explanation that does not require that the mechanism of reference is already in the head of the local user in the form of a self-assigned description" (1989: 602–3)), one may be led to believe that the answer to our initial question must be negative: the mind of the speaker does not play any role in determining the reference of the proper name tokens he or she produces. Indeed, Kaplan himself seems to have been at least tempted by this idea, when, in contrasting "the subjectivist views of Frege and Russell" (603) with "the view that we are, for the most part, language *consumers*" (602)—in his terms, *subjectivist semantics* with *consumerist semantics*—he urged us to "see language, and in particular semantics, as more autonomous, more independent of the thought of individual users" (603–4). Some other philosophers followed suit.[3] And I should add that my own attempt to use Kaplan's (1990) notion of repetition to develop Kripke's chain of communication picture into a full-blown theory of proper name reference (Bianchi 2015) also goes in this direction.

[2] Actually, aiming at generality, Donnellan formulated the principle so as to "leave it open … whether the set of identifying descriptions is to be formed from what *each* speaker can supply or from what speakers collectively supply" (339), and the second alternative seems to allow for a negative answer to the question in the text. Indeed, as I noted in Bianchi 2012: 84, the position combining descriptivism and (semantic) anti-subjectivism is not inconsistent. But it is indisputable, I believe, that what drives most of the descriptivists is the idea that the speaker must have *epistemic* control on what he or she refers to.

[3] See especially Wettstein 2004, Hinchliff 2012, and Martí 2015.

Thus, I certainly do not believe that the mind of the speaker plays any role in determining the reference of the proper name tokens he or she produces. But, unfortunately, things are not as clear as it may appear, even if one takes the road opened up by the revolutionary work of Kripke and Donnellan. On the one hand, Donnellan himself seems to have thought otherwise, since he made reference crucially depend on *having in mind*. As a matter of fact, for a long time Donnellan's "historical explanation theory" was obscured by Kripke's chain of communication picture, to which it was wrongly assimilated.[4] As of recently, however, a group of philosophers, related in one way or another with UCLA, where Donnellan taught for many years, have rediscovered, developed, and radicalized his ideas on reference, determining something like a Donnellan *Renaissance* in the field.[5] On the other hand, a number of other philosophers have found a different, subtler, way to find a place for the speaker's mind in the theory of reference, by appealing instead to the distinction between *speaker's reference* and *semantic reference* introduced in the debate by Kripke to argue against Donnellan's account of definite descriptions, and interpreting it so as to endorse what I shall call the *Gricean project*.

I have dealt with Donnellan's and the neo-Donnellanians' account of reference elsewhere and shall not say any more about it here.[6] My aim in this paper is instead to examine and criticize the second approach. In particular, I shall focus on Michael Devitt's version of it. In so doing, I shall continue an ongoing debate with Devitt himself, on his causal theory of proper names and the nature of reference (see Devitt 2015, Bianchi forthcoming, and Devitt forthcoming *a* for the previous stages). I shall proceed as follows. I shall present and discuss Kripke's distinction and the particular interpretation of it that amounts to endorsing the Gricean project in Sections 2 and 3. In Section 4, I shall examine Devitt's causal theory of proper names, paying special attention to its relation to the project. Finally, I shall draw some general conclusions in Section 5.

2. *Kripke's distinction*

As far as I know, the distinction between *speaker's reference* and *semantic reference* makes its first appearance in the literature, quite incidentally, near the beginning of "Naming and Necessity".[7] Before elaborating on "the relation between names and descriptions," Kripke says the following:

[4] See Bianchi and Bonanini 2014 for a detailed reconstruction of Donnellan's historical explanation theory of proper name reference that contrasts it with Kripke's picture.

[5] See in particular Almog 2012 and 2014: chap. 3, Capuano 2012 and 2018, Pepp 2012 and 2019, Almog, Nichols and Pepp 2015, and Wulfemeyer 2017.

[6] See my "Reference and Language," forthcoming.

[7] See however Geach 1962: 31–2 for an earlier hint at the distinction.

> It is a point, made by Donnellan, that under certain circumstances a particular speaker may use a definite description to refer, not to the proper referent … of that description, but to something else which he wants to single out and which he thinks is the proper referent of the description, but which in fact isn't. So, you may say "The man over there with the champagne in his glass is happy", though he actually only has water in his glass. Now, even though there is no champagne in his glass, and there may be another man in the room who does have champagne in his glass, the speaker *intended* to refer, or maybe, in some sense of 'refer', *did* refer, to the man he thought had the champagne in his glass. Nevertheless I'm just going to use the term 'referent of the description' to mean the object uniquely satisfying the conditions in the definite description. (1972: 254 (1980: 25–6))

Kripke is pausing here on a common phenomenon: *sometimes we intend to refer to something to which we do not actually refer*.[8] As in many other cases (think of the intention to help, or kill, or email, someone, for example), not always are our intentions successful—perhaps we do not choose the right means, or the environment does not 'cooperate,' or … That's life, one would say. But then, Kripke makes a surprising move and states, although cautiously (notice the "maybe"), that there is a sense of "refer" according to which even in this case we may say that we referred to what we did not actually refer to, in the first, primary, sense. This is strange. Take the case of emailing and intending to email, and assume, to make it even more similar to the one under discussion, that *a* intends to email *b* but fails and ends up emailing *c* instead. Here, certainly we are not inclined to say that although *a* emailed *c* in the first, primary, sense of "email," *b* was also emailed by *a*, in another sense of "email"—"emailing" does not seem to be ambiguous. Thus, why should we instead take "referring" as ambiguous? I shall come back to this in the next Section. As for now, let me only note that, ironically, it is Kripke himself, and furthermore in the same article where he elaborates on the distinction between speaker's reference and semantic reference, who provides us with reasons for being suspicious:

> It is very much the lazy man's approach in philosophy to posit ambiguities when in trouble. If we face a putative counterexample to our favourite philosophical thesis, it is always open to us to protest that some key term is being used in a special sense, different from its use in the thesis. We may be right, but the ease of the move should counsel a policy of caution: Do not posit an ambiguity unless you are really forced to, unless there are really compelling theoretical or intuitive grounds to suppose that an ambiguity really is present. (1977: 268)

As we have just seen, in "Naming and Necessity" Kripke introduces the distinction between speaker's reference and semantic reference with an example involving a definite description, and definite descriptions are not the topic of this paper. However, in a footnote appended to the above passage, Kripke adds:

[8] Although I am uncomfortable with saying that a definite description *refers* to its *denotatum*, I shall follow Kripke's usage here.

> Donnellan's distinction seems applicable to names as well as descriptions. Two men glimpse someone at a distance and think they recognize him as Jones. 'What is Jones doing?' 'Raking the leaves'. If the distant leaf-raker is actually Smith, then in some sense they are *referring* to Smith, even though they both use 'Jones' *as a name of* Jones.... I speak of the 'referent' of a name to mean the thing named by the name—e.g., Jones, not Smith—even though a speaker may sometimes properly be said to use the name to refer to someone else.... I am tentatively inclined to believe, in opposition to Donnellan, that his remarks about reference have little to do with semantics or truth-conditions, though they may be relevant to a theory of speech-acts. Space limitations do not permit me to explain what I mean by this, much less defend the view, except for a brief remark: Call the referent of a name or description in my sense the 'semantic referent'; for a name, this is the thing named, for a description, the thing uniquely satisfying the description.
> Then the speaker may *refer* to something other than the semantic referent if he has appropriate false beliefs. I think this is what happens in the naming (Smith-Jones) cases and also in the Donnellan 'champagne' case; the one requires no theory that names are ambiguous, and the other requires no modification of Russell's theory of descriptions. (1972: 343 n. 3 (1980: 25n))

After these brief and incidental remarks, there is no more mention of speaker's reference in "Naming and Necessity". In fact, for our purposes here it is important to keep in mind that the chain of communication picture offered by Kripke in his second lecture concerns (proper name) *semantic* reference, not at all speaker's reference.

As is well known, in his 1977 article Kripke develops these remarks and makes the distinction between speaker's reference and semantic reference his main weapon for arguing against the semantic significance of Donnellan's (1966) distinction between attributive and referential uses of definite descriptions. In fact, according to Kripke, Donnellan confused what a speaker refers to, by using a definite description, with what the description he or she uses refers to, on that occasion. Consider Leonard Linsky's famous example. While it is certain that the speaker who utters "Her husband is kind to her" after observing the attitude of a man towards a woman refers to the man, who, however, is not her husband but, let us suppose, her lover, one may (and should, as Kripke then argues on methodological grounds) doubt that the description he uses semantically refers, on that occasion, to that person rather than to nobody (as in Linsky's original case, where the woman is a spinster) or to her husband (as in Kripke's modified version, where she is married to a cruel man). In fact, Kripke goes on, the distinction Donnellan seems to have overlooked applies to other referential terms as well—arguably to all, although Kripke does not mention indexicals and demonstratives. In particular, Kripke discusses the case we have already encountered in the footnote from "Naming and Necessity":

> Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: "What is Jones doing?" "Raking the leaves." "Jones," in the common language of both, is a name of Jones; it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith. (1977: 263)

Here, according to Kripke, one may agree that, by using "Jones," the two people refer to Smith.[9] But it is certainly beyond dispute that what the name they use refers to on that occasion is Jones (in their common language, it is a name *of* him!). All in all, then, it seems as if for every use of a (non-empty) proper name or a (proper) definite description, i.e., for every *token* of them, we need to distinguish two important relations it bears to individuals, *speaker's reference* and *semantic reference*, and in some cases the individual a token is related to by the first relation differs from the individual that very token is related to by the second.[10]

Unlike in "Naming and Necessity", in the 1977 article the distinction between speaker's reference and semantic reference is introduced by means of theoretical considerations, and as part of a "general apparatus." In fact, Kripke's alleged starting point is now Paul Grice's approach to meaning: "[f]irst, let us distinguish, following Grice, between what *the speaker's words meant*, on a given occasion, and what *he meant*, in saying these words, on that occasion" (262). After discussing some examples, he sums up:

> The notion of what words can mean, in the language, is semantical: it is given by the conventions of our language. What they mean, on a given occasion, is determined, on a given occasion, by these conventions, together with the intentions of the speaker and various contextual features. Finally what the speaker meant, on a given occasion, in saying certain words, derives from various further special intentions of the speaker, together with various general principles, applicable to all human languages regardless of their special conventions. (Cf. Grice's "conversational maxims.") (263)

Only at this point does Kripke introduce his distinction. According to him, in fact, speaker's reference and semantic reference "are special cases of [these] Gricean notions" (263). I shall postpone the discussion of this claim to the next Section.

Concerning semantic reference, in the article Kripke does not say much. His characterization of it is the following:

> If a speaker has a designator in his idiolect, certain conventions of his idiolect (given various facts about the world) determine the referent in the idiolect: that I call the *semantic referent* of the designator. (If the designator is ambiguous, or contains indexicals, demonstratives, or the like, we must speak of the semantic referent on a given occasion. The referent will be determined by the conventions of the language plus the speaker's intentions

[9] Note, however, that, as Devitt remarked a long time ago (1981a: 514–5), concerning this case intuitions are much less clear than concerning Linsky's. Perhaps, following Devitt, one should rather say that by that use the two people refer partially to Jones and partially to Smith. I shall not take a stand on this here. (On this issue, see also footnote 13 and Section 4 below.)

[10] As a matter of fact, Kripke's speaker's reference is not a binary relation between a token and an object but a ternary relation between a speaker, a use of a designator and an object. However, from it a binary relation may easily be defined along the following lines: a token of a designator *speaker-refers* to an object if and only if the speaker who produces the former refers to the latter by using the designator on that occasion.

> and various contextual features.) (263)

Coming from Kripke, this appeal to *idiolects* is somewhat surprising, but it is probably due to his willingness to remain neutral about semantic matters when outlining the distinction between semantic reference and speaker's reference—one has to acknowledge the distinction, whatever his or her semantic theory. In fact, in a footnote appended to the passage, Kripke adds: "If the views about proper names I have advocated in 'Naming and Necessity' are correct ... the conventions regarding names in an idiolect usually involve the fact that the idiolect is no mere idiolect, but part of a common language, in which reference may be passed from link to link" (273 n. 20). Indeed, if those views are correct, as I shall assume throughout the paper, semantic reference, at least as far as proper names are concerned, is a *historical* matter. As I have already mentioned, I have tried to develop Kripke's chain of communication picture into a full-blown theory elsewhere, and I shall not say anything more about semantic reference here except for this brief remark: in the above characterization, Kripke explicitly mentions *speaker's intentions* but *only* to deal with ambiguity and indexicality. As a matter of fact, I believe Kripke is wrong in claiming that to deal with these linguistic phenomena we need to appeal to intentions, but I shall not argue in favor of this here. However, what I would like to be noticed is that, except when ambiguity or indexicality is involved, Kripke himself does not seem to think that speaker's intentions play any role in determining semantic reference (unless the notion of convention invoked in his characterization needs to be explained in terms of them, which I do not think is the case if the chain of communication picture is on the right track).[11]

Let us now move on to speaker's reference. Kripke begins with some words of caution, stating that "[s]peaker's reference is a more difficult notion" (263). This is already interesting, given that it contrasts with a certain attitude some philosophers have towards the notion (as if, contrary perhaps to semantic reference, speaker's reference were easy to characterize). Then, he presents the Smith-Jones case we have already encountered and asks how we can account for it. Here is his answer:

> Suppose a speaker takes it that a certain object a fulfills the conditions for being the semantic referent of a designator, "*d*." Then, wishing to say something about a, he uses "*d*" to speak about a; say, he says "$\phi(d)$." Then, he said, of a, on that occasion, that it ϕ'd; in the appropriate Gricean sense ..., he *meant* that a ϕ'd. This is true even if a is not really the semantic referent of "d." If it is not, then *that* a ϕ*'s* is included in what he meant (on that occasion), but not in the meaning of his words (on that occasion). (263–4)

From this, Kripke arrives at his definition of *speaker's reference*:

[11] Actually, this is not completely true, since a few lines after characterizing semantic reference Kripke writes that "[i]n a given idiolect, the semantic referent of a designator (without indexicals) is given by a *general* intention of the speaker to refer to a certain object whenever the designator is used" (264). However, I think that this appeal to general intentions may easily be dispensed with.

> we may tentatively define the speaker's referent of a designator to be that object which the speaker wishes to talk about, on a given occasion, and believes fulfills the conditions for being the semantic referent of the designator. He uses the designator with the intention of making an assertion about the object in question (which may not really be the semantic referent, if the speaker's belief that it fulfills the appropriate semantic conditions is in error). The speaker's referent is the thing the speaker referred to by the designator, though it may not be the referent of the designator, in his idiolect. (264)

So, it seems that, for there to be speaker's reference, there has to be, (1), a speaker's use of a designator to assert something (but, I assume, any other illocutionary act would do as well), backed by, (2), his or her wish to talk about a particular object, and, (3), his or her belief about that particular object that it is the semantic referent of the designator.[12] More precisely, a speaker a refers to an individual b by using a designator c if and only if, (1), a wishes to talk about b, and, (2), a believes of b that it is the semantic referent of c, and, (3), a produces a token of c in the course of accomplishing an illocutionary act.[13]

Is this a good definition? I have some qualms concerning the first clause, because it is not clear to me what *wishing to talk about* consists in exactly. However, for the sake of the argument I shall simply assume that a broadly causal account will work here: what someone wishes to talk about when he or she accomplishes an illocutionary act is whatever object prompts his or her act—though, obviously, much more than this would need to be said. The third clause is trivial. We shall pause on the second clause, which makes a's referring to b by using c depend on a's believing of b that it is the semantic referent of c, in the next Section. However, there is no doubt that it is intelligible. So, if we ignore the qualms concerning *wishing to talk about*, we may conclude that Kripke's notion of speaker's reference is well defined: we know what has to be the case for there to be what he calls "speaker's reference."

[12] As a matter of fact, Kripke is aware that some of the cases discussed by Donnellan (for example, that of "the king" used to refer to someone known to be the usurper) do not involve such a belief. He takes them to be "of a somewhat exceptional kind" and writes: "Largely for the sake of simplicity of exposition, I have excluded such ... from the notion of speaker's reference .... I do not think that the situation would be materially altered if [the notion] were revised so as to admit these cases, in a more refined analysis" (273 n. 22). I shall go along with Kripke's assumption here. Probably, to deal with these cases, the analysis would have to invoke even more sophisticated beliefs.

[13] By the way, let me note that, according to Kripke's definition, we should say that in the Smith-Jones case by using "Jones" the two speakers refer not only to Smith but also to Jones (see footnote 9 above). In fact, they certainly wish to talk about Jones (if not, why would they use "Jones"?) and of course believe of him that he is the semantic referent of "Jones." Actually, Kripke himself seems to acknowledge this (274–5 n. 28).

3. *The Gricean project*

At the beginning of his article, Kripke writes that he believes that the "contrast" between speaker's reference and semantic reference "is of considerable constructive as well as critical importance to the philosophy of language" (255). At the very end of the article, the claim is reiterated in more or less the same words.

Actually, Kripke's distinction has been enormously successful. Nowadays, talk of speaker's reference beyond semantic reference is widespread among philosophers of language. In fact, almost all of them are now convinced that reference comes in two different sorts:[14] there is semantic reference, which contributes to determine the semantic properties of the linguistic expressions we use, and speaker's reference, which contributes to determine other, *pragmatic*, properties of them and which a theory of speech acts should pay attention to.

Notice that this already muddies the waters concerning our initial question, which, if Kripke is right, turns out to be ambiguous: it can concern the determination either of the semantic reference or of the speaker's reference of a proper name token. And, if it concerns the determination of the latter, the answer cannot but be positive: given how speaker's reference depends on the speaker's wishes and beliefs, of course the mind of the speaker does play a substantial role in determining the speaker's reference of the proper name tokens he or she produces. But things can become even worse, since if Kripke's distinction is interpreted so as to endorse the Gricean project, a move we are about to discuss, even semantic reference ends up being 'mind-contaminated,' contrary to what I take to be one of the main lessons of Kripke's chain of communication picture.

But, are we really forced to assume that reference comes in two different sorts?

To begin with, let me note that the fact that the notion of speaker's reference introduced by Kripke is well defined does not settle the issue yet. To see this, consider the following case. Micky wishes to go to Bologna and believes that train 2286 goes there. Hence, she takes that train. Unfortunately, her belief is false: train 2286 goes in the opposite direction, to Milan. This can happen, especially to a person as inattentive as Micky. We know how to describe the situation: Micky intended to go to Bologna but, because of her inattention (and, more specifically, of her false belief about train 2286), she chose the wrong means and ended up going to Milan. But now, suppose that someone introduces the notion of, say, *traveler's going*, defining it in the following way: a traveler a goes to a place b by taking a train c if and only if, (1), a wishes to go to b, and, (2), a believes of b that it is where c goes, and, (3), a takes c for his or her journey. Undoubtedly, the notion is well de-

[14] A notable exception is constituted by the neo-Donnellanians (see the works mentioned in footnote 5 above). Although I strongly disagree with their account of reference, I am sympathetic to their 'unitary' approach to it.

fined: we know what has to be case for there to be what our introducer calls "traveler's going." In particular, according to it we may say that by taking train 2286 Micky *went* to Bologna, although of course according to another, perhaps primary and certainly more standard, sense of "going," by taking that train she *went* to Milan, because Milan is where the train *went*. It is even possible that the notion thus defined helps explain some of the traveler's (e.g., Micky's) actions. But I assume that everyone would regard it as absurd to conclude from this that going (to a place) comes in two different sorts: intending to go somewhere and failing to do so does not amount to going there according to some other sense of "going." By parity of reasoning, we should not be too hasty to conclude that reference comes in two different sorts only because Kripke's notion of speaker's reference is well defined, since the notion of traveler's going is also well defined, and along similar lines.

To establish whether reference really comes in two different sorts, then, we need to go beyond Kripke's definition. The only reasonable strategy, it seems to me, is to consider the theoretical roles the notion so defined is supposed to play, to see, (1), whether it can really play these roles, and, (2), if indeed it can play them, whether the fact that it can justifies the claim that speaker's reference is some sort of reference (in contrast, to repeat, to traveler's going, which no one would take to be any sort of going).

Well, what are the theoretical roles that Kripke's notion is supposed to play? From what Kripke writes at the beginning of his article (see above) we may infer that he takes the notion to have both a *critical* and a *constructive* use. We need, then, to consider the two of them.

Kripke's article is almost entirely devoted to the critical use of the notion of speaker's reference and of the ensuing distinction. As Kripke makes explicit in the last paragraph of it, in fact, the latter can play an important role "as a critical tool to block postulation of unwarranted ambiguities" (271). We have already seen in the preceding Section how this tool basically works. Consider the Smith-Jones example again, and suppose that, impressed by the two speakers' dialogue, some theorists claim that, in both speakers' idiolect, the name "Jones," which they are using, is semantically ambiguous: it habitually refers to Jones, but in the context of the dialogue to Smith.[15] Against them, it can be objected that they are confusing what the speakers are referring to, on that occasion, with what the name the speakers use refers to, on that and other occasions: the claim that "Jones" is semantically ambiguous (in the sense just specified) seems to be unwarranted. Of course, the case of definite descriptions is the one Kripke is mostly interested in. Consider Linsky's example again. It is reasonable to interpret Donnellan (1966) as claiming that the description "her husband," which

[15] By the way, let me note that the neo-Donnellanians tend to make similar claims (see e.g. Almog, Nichols and Pepp 2015: 368–74 and Capuano 2018). Of course, they know well about Kripke's "critical tool," but they are unimpressed by it.

the speaker uses, is semantically ambiguous: it often refers to someone who is the husband of the contextually salient woman, but in the context depicted by Linsky to a man who is not in fact her husband. Against Donnellan, Kripke objects that he confuses what the speaker refers to, on that occasion, with what the description the speaker uses refers to, on that occasion, who is, as for any other use of it, the husband of the contextually salient woman (if there is any): the claim that "her husband" is semantically ambiguous (in the sense just specified), Kripke concludes, is unwarranted.

As I have already made clear, definite descriptions are not the topic of this paper, and this is certainly not the place to evaluate Kripke's argument. Thus, concerning this I limit myself to saying that I believe Kripke's considerations indeed have some bite against Donnellan's views, although by themselves they do not suffice to settle the issue concerning the semantics of definite descriptions (as, I hasten to add, Kripke himself is ready to admit).[16] What is important to notice for our purposes, however, is that even if the argument succeeds, its success does not essentially depend on there being another sort of reference beyond semantic reference. To block postulation of unwarranted ambiguities, in fact, we do not need the critical tool Kripke introduced, although its introduction may have been helpful from a rhetorical point of view. We can get exactly the same results by arguing that in the critical cases, be they the Smith-Jones one or Linsky's, the postulator confuses what the speaker *intended to refer to*, on that occasion, with what the speaker *actually referred to*, on that occasion (which is determined by the semantic properties of the designator the speaker uses). The distinction we need is the simple and commonsensical one between *intending to do something* and *doing something*, as applied to reference. The notion of speaker's reference is, then, an idle wheel here. Worse than that, it can mislead, and has actually misled, people, since it invites one to obliterate the obvious and important difference between successful and unsuccessful intentions, namely that when our intention is successful, we end up doing what we intended; when it isn't, we fail to do what we intended. Let me emphasize the point: *failed reference to something is no reference to it*.

Well, but what about the constructive use of the notion of speaker's reference and of the ensuing distinction? Doesn't it vindicate the claim that reference comes in two different sorts? Unfortunately, concerning it Kripke says almost nothing. In fact, he limits himself to touching on the issue in the very final passage of his article, which is now finally time to quote in its entirety:

> I think that the distinction between semantic reference and speaker's reference will be of importance not only (as in the present paper) as a critical tool to block postulation of unwarranted ambiguities, but also will be of consid-

[16] For some criticisms of Kripke's argument, see for example Devitt 1981a and Devitt 2004.

> erable constructive importance for a theory of language. In particular, I find it plausible that a diachronic account of the evolution of language is likely to suggest that what was originally a mere speaker's reference may, if it becomes habitual in a community, evolve into a semantic reference. And this consideration may be *one* of the factors needed to clear up some puzzles in the theory of reference. (1977: 271)

Thus, Kripke thinks that to explain the evolution of language, and more specifically the establishment of a semantic relation of reference between a designator and an object, we can profitably use the notion of speaker's reference.

The claim seems to me to be open to two interpretations, one moderate, the other radical. The moderate one, which, for reasons that will become clear at the end of this Section, I assume to be the one Kripke had in mind, sees speaker's reference as being involved in the puzzling phenomenon of (semantic) *reference change*, on which Gareth Evans (1973: 11) famously put his finger when arguing against what he called "the Causal Theory of Names." Kripke himself, in fact, mentions Evans' Madagascar case in a footnote appended to the last sentence of the passage just quoted. And Kripke's idea that we can profitably use the notion of speaker's reference to clear up the puzzle has actually been exploited and developed by Devitt (1981b: 150–1; 2015: 121–4), who argues that reference change is explained by "*change in the pattern of groundings*" (2015: 122). Now, I actually have some qualms about Kripke's idea and Devitt's development—I still find the puzzle puzzling (see Bianchi 2015: 104–6)—but even if we concede that the solution works, it does not seem to me that this provides good enough reasons to claim that reference comes in two different sorts. Exactly the same kind of explanation, in fact, can be obtained by appealing to massive reference failure (reference failure that "becomes habitual in a community", to use Kripke's phrase), which somehow determines the establishment of a new semantic relation.[17] Speaker's reference is an idle wheel here as well, in my opinion.

However, as I have said there is a more radical interpretation of Kripke's claim. According to this, the notion of speaker's reference is useful for explaining not only (semantic) reference change, but semantic reference *tout court*. In a nutshell: *there could not be semantic reference if there were not speaker's reference*. This interpretation of Kripke's claim, and more generally of his distinction, amounts to endorsing what I have called the Gricean project, by seeing speaker's reference as *explanatorily basic* with respect to semantic reference.

[17] To avoid misunderstanding, let me make it clear that I am using "reference failure" here to talk not, as is more common, of cases where no reference is in fact made, but of cases where reference is made to something that is not what the speaker intended to refer to. In these cases, the speaker intends to refer to something (e.g., Smith, or the great African island) but *fails* and refers to something else instead (respectively, Jones and a portion of the African mainland).

We have seen in Section 2 that, to introduce his distinction, in his 1977 article Kripke appeals to Grice's work on meaning. In particular, Kripke mentions Grice's distinction "between *what the speaker's words meant,* on a given occasion, and what *he meant*, in saying these words, on that occasion" and claims that the notions of semantic reference and speaker's reference are just "special cases" of Grice's ones. The fact is, however, that one stage in Grice's general program concerning meaning was the explanation of *word* (and sentence) *meaning* in terms of *utterer's meaning*, the other stage being of course that of explaining the latter in terms of *intentions*.[18] As is often the case, the details of Grice's proposal varied over the years, but fortunately we do not need to pause on them for our purposes. On the explanatory priority of utterer's meaning over word and sentence meaning, however, Grice was always crystal clear. In his very first article on the topic, for example, he concludes his criticism of a causal account of meaning, which he attributes to C. L. Stevenson (not to be confused with Devitt's later and quite different causal theory of proper names that we shall examine in the next Section) by saying that "the causal theory ignores the fact that the meaning (in general) of a sign needs to be explained in terms of what users of the sign do (or should) mean by it on particular occasions; and so the latter notion, which is unexplained by the causal theory, is in fact the fundamental one" (1957: 217). And twenty-five years later, in his late revisiting of these issues, he writes:

> It seems plausible to suppose that to say that a sentence (word, expression) means something (to say that "John is a bachelor" means that John is an unmarried male, or whatever it is) is to be somehow understood in terms of what particular users of that sentence (word, expression) mean on particular occasions. The first possible construal of this is rather crude: namely, that usually people do use this sentence, etc., in this way. A construal which seems to me rather better is that it is conventional to use this sentence in this way; and there are many others. (1982: 298)[19]

It is certainly not within the scope of this paper to evaluate Grice's claim concerning the explanatory priority of utterer's meaning over

[18] The first stage, which is the one I am interested in here, is discussed at length in Grice 1968. Summing up that article in a following one devoted to the second stage instead, Grice writes: "Starting with the assumption that the notion of an utterer's occasion-meaning can be explicated, in a certain way, in terms of an utterer's intentions, I argue in support of the thesis that timeless meaning and applied timeless meaning can be explicated in terms of the notion of utterer's occasion-meaning (together with other notions), and so ultimately in terms of the notion of intention" (1969: 150).

[19] Interestingly, in the immediately following paragraph Grice adds: "I do not think that [sentence (word, expression)] meaning is essentially connected with convention. What it is essentially connected with is some way of fixing what sentences mean: convention is indeed one of these ways, but it is not the only one. I can invent a language, call it Deutero-Esperanto, which nobody ever speaks. That makes me the authority, and I can lay down what is proper" (298–9). Thus, *contra* Devitt (see the next Section), Grice believes that there can be word meaning (i.e, a word can have *semantic* properties) even in the absence of conventions.

word or sentence meaning, even less his entire program concerning meaning, although what I am saying may have some bearing on it.[20] My focus here is rather what I have called the Gricean project, the related claim that speaker's reference is explanatorily prior to semantic reference, a claim that we have seen emerge from a radical interpretation of Kripke's passage about the "constructive" use of his distinction. This, in fact, has an obvious impact on the issues we are interested in. First, if the claim were true, semantic reference would depend on a more basic relation, which it would be difficult not to consider as a form of reference, hence we would be almost forced to finally acknowledge that reference comes in two different sorts. Second, since, as we have seen, speaker's reference in turn depends on wishes and beliefs, we would have to give a positive answer to our initial question (and, more generally, adopt a psychological model of the functioning of language): the mind of the speaker would play a role in determining both the speaker's reference (directly, so to say) and the semantic reference (more indirectly, via the explanatory dependence of semantic reference on speaker's reference) of the proper name tokens he or she produces.

Now, many philosophers of language who would describe themselves as having a broadly speaking Kripkean approach to reference do indeed endorse, either explicitly or implicitly, the Gricean project. In the next Section, I shall discuss Devitt's case, whose causal theory constitutes a detailed account of speaker's reference, semantic reference, and the explanatory dependence of the latter on the former. To give only one further example, in a recent article Mark Sainsbury defended the claim that "[a]lthough reference is often transmitted causally, what determines semantic reference is conventionalized speaker-reference" (2015: 195), in the following way:

> The "semantic reference" of a name, as used in a community, is its conventionalized, stabilized or normalized speaker-reference in the community. "London" refers to London among many speakers who live in England (and elsewhere) because it's a conventional or stabilized or normal fact about these speakers that they use the specific name "London" … only if they intend thereby to refer to London. The notion of semantic reference is a theoretical one, and one that needs to be constructed to suit theoretical purposes. … [W]e need a conception of semantic reference that will supervene on use and help explain features of usage (for example, agreement, disagreement, correction). Basing semantic reference on speaker-reference is the most straightforward, and perhaps the only, way to achieve this. Speaker-reference can be theoretically described without any theoretical commitment to semantic reference, so the supervenience relation has a reductive character. Much work has been done, and much remains to be done, to sort out what the supervenience relation should be based on. Here I give a trio of possibilities (convention, stability, normalization); a determinate thesis would need to choose from among them, and also clarify the preferred option. (209)

[20] For some early criticism of Grice's claim, see Black 1973, Biro 1979, and Yu 1979. For a defence of it, Suppes 1986.

But, let us finally ask, is the Gricean project really something that should be pursued? More specifically, can Kripke's speaker's reference be used to explain semantic reference? On the face of it, the answer to the latter question should be a round "No." As we saw in the preceding Section, in fact, according to Kripke's definition a speaker cannot refer to *b* by using a designator *c* if he or she does not believe of *b* that it is the semantic referent of *c*. But, in order to believe of something that it is the semantic referent of something else, of course the speaker needs to have the concept of semantic reference. Since it is scarcely imaginable that one has this concept without there being semantic reference, we must then conclude that speaker's reference presupposes semantic reference: the second clause in Kripke's definition rules out the possibility of explaining the latter in terms of the former (and this, let me add, renders Kripke's distinction much less Gricean than he himself alleged it was).[21] In a nutshell: according to Kripke's definition *there could not be speaker's reference if there were not semantic reference.*

4. *Devitt's causal theory of proper names*

We have reached the conclusion that Kripke's definition of speaker's reference rules out the interpretation of his distinction amounting to endorsing what I have called the Gricean project, which is the most promising, if not the only, way to use the distinction to claim that reference comes in two different sorts, and that as a consequence our initial question should be given a positive answer. However, there is still an option that we have to discuss. Those who for whatever reasons (for example, because they sympathize with Grice's general approach to language) believe that something like the Gricean project must be on the right track, might insist that Kripke was onto something important when he introduced the notion of speaker's reference, but that his definition of it was inadequate. They might even support the latter claim by voicing some independent doubts about the second clause of Kripke's definition, noting that it over-intellectualizes the speech act of referring. According to the definition, in fact, in order to refer to something one needs to have fairly sophisticated semantic beliefs. It is quite implausible that children have such beliefs, but it is no less

[21] I first noted that Kripke made speaker's reference "parasitic" on semantic reference in Bianchi 2011: 277. See also Bianchi and Bonanini 2014: 182, and Bianchi forthcoming. Peter Hanks has recently made exactly the same point. As he writes, Kripke "defines the notion of speaker reference partly in terms of the notion of semantic reference" (2019: 14). Therefore, "[i]f Kripke is right, and the concept of semantic reference figures crucially in the definition of speaker reference, then it cannot be that speaker reference is somehow prior to semantic reference" (*ibid.*). Much earlier, Rod Bertolet noted some tension between Grice's framework and Kripke's distinction ("There is … no easy assimilation of the example Kripke discusses to Grice's distinction between what a speaker's words mean and what he means by them or in saying them" (1981: 72)), but the reasons he offered are quite different.

implausible that they are not able to refer.[22] Moreover, this seems to be in stark contrast with the picture of reference Kripke himself offered in "Naming and Necessity"—referring is easy: to succeed in it we do not need to know, or even believe, anything about what we are referring to, but only to be connected with it by means of an appropriate chain of communication—and with the assumption he implicitly makes in the article in which he gives his definition that every time one uses a designator to assert something he or she is referring (even though in the large majority of cases the speaker's referent coincides with the semantic referent of the designator).[23]

Although to my knowledge nobody has ever explicitly stated the option just outlined, I believe that, upon reflection, quite a lot of philosophers would be ready to subscribe to it. In the passage quoted in the preceding Section, for example, Sainsbury writes that "[s]peaker-reference can be theoretically described without any theoretical commitment to semantic reference," which is not something that anyone accepting Kripke's definition could say. In this Section, I shall focus on Devitt's causal theory of proper names, which may be taken as a way of articulating the option within a rich, naturalistic, framework. As we shall see, without discussing Kripke's, Devitt offers a different definition of speaker's reference (in his terms, *speaker-designation*), which does not appeal to (beliefs about) semantic reference.[24] Before starting my examination, however, I would like to highlight something that more or less follows from what I have said so far but could be missed by someone who approaches Devitt's theory without paying due attention to the details of Kripke's distinction. Devitt, like anyone else who pursues the Gricean project, puts the notion Kripke introduced to a novel use, a use that was not amongst those Kripke was thinking of. Because of this, to make the notion acceptable he *cannot* simply appeal to intuitions concerning cases such as the Smith-Jones one. In fact, these intuitions *at most* justify the introduction of a notion defined as Kripke did, where a belief about semantic reference plays a crucial role, and, as we saw, a notion so defined cannot play the explanatory role Devitt wished it to play. Thus, Devitt needs to vindicate *his* distinction between speaker's reference and semantic reference in a different way.

Devitt's causal theory of proper names makes its first appearance in print in "Singular Terms," an article that draws from his PhD dissertation and is published in 1974, after "Naming and Necessity", the avowed source of inspiration, but before the article where Kripke elaborates on the distinction between speaker's reference and semantic refer-

[22] This objection is hinted at in Devitt 1981a: 513.

[23] Perhaps, Kripke might reply by noting that there is a sense according to which, for any designator "*a*" we have in our lexicon, we may be said to believe that *a* is the semantic referent of "*a*." Even if this were true, however, it would not allow him to account for cases involving children where the speaker's referent seemingly diverges from the semantic referent.

[24] For yet another definition, more Gricean in that, unlike Devitt's, it is couched in terms of intentions (it appeals to the notion of intending to direct someone's attention to something), see Bertolet 1987.

ence. Interestingly, in "Singular Terms" Devitt does not draw any such distinction. On the contrary, he offers a unitary account of reference (in his terms, *designation*). As I have argued elsewhere, this account is more Donnellanian than Kripkean, in that it explains proper name reference in terms of *having in mind* ("We can say roughly … that a name token designates an object if and only if the speaker had the object in mind (meant the object) in uttering the token" (1974: 189)), where having an object in mind is explained not in terms of having (identifying) knowledge, as done by those philosophers who adopt the principle of identifying descriptions, but causally ("one has an object in mind in virtue of a causal connection between one's state of mind and the object" (188; see also Devitt 1976: 409–10). Note, incidentally, that this implies a straight positive answer to our initial question: as in Donnellan's historical explanation theory and in the neo-Donnellanians' accounts, according to the first formulation of Devitt's causal theory the mind of the speaker *directly* determines the reference of the proper name tokens he or she produces.[25]

Only in his book *Designation* does Devitt introduce into his framework a distinction similar, but, importantly, not identical, to Kripke's. After a first outline of his causal theory of designation, which resembles the one proposed in the 1974 article, Devitt embarks on a defence of the language of thought hypothesis, and relates it to "a Gricean distinction between speaker meaning and conventional meaning" (1981b: 80):

> Consider an utterance. In my view, *what the speaker means* by the token he utters is determined by the meaning of the thought that causally underlies his utterance. On the other hand, the *conventional meaning* of the token in a community is determined by what a member of that community using a token of that physical type would commonly mean and be taken to mean. What he would commonly mean and be taken to mean depends in some way on what people have commonly meant by words of that physical type and by sentences of that structure. (80)

Like Kripke, then, Devitt starts from Grice's approach to meaning. Contrary to Kripke, however, he explicitly subscribes to Grice's claim that word and sentence meaning should be explained in terms of speaker meaning (although not to the further one that speaker meaning should be explained in terms of intentions): "I explain conventional meaning in terms of speaker meaning and speaker meaning in terms of thought meaning" (80).[26] Note, also, that in Devitt's hands, Grice's sentence (word, expression) meaning has become *conventional* meaning: the explanandum is now a *conventionally determined* property of linguistic tokens (compare footnote 19 above).

[25] I elaborate on these issues in Bianchi forthcoming. See Devitt forthcoming *a* for some discussion.

[26] See also Devitt 1981a: 519: "We seem to need notions of speaker meaning that enable us to explain conventional meaning. It seems that conventional meaning must be built up in some way from common speaker meanings." For a recent general defence of this approach, see Devitt forthcoming *b*: chap. 3.

Unfortunately, I cannot examine Devitt's Gricean account of meaning in its full generality here and I shall have to limit myself to discussing and criticizing his causal theory of proper name reference, which rests against that background. I believe that if what I shall say about it is on the right track, something should be readjusted in the background as well—the relationships between mind and language are not that simple!—but I shall not argue in favor of this here.

In a recent article—his latest revisiting of his causal theory of proper names—Devitt offers the following definitions ("biconditionals") for (proper name) speaker's reference and semantic reference:

> *Speaker-Designation*: A designational name token speaker-designates an object if and only if all the designating-chains underlying the token are grounded in the object. (2015: 125)
> *Conventional-Designation*: A designational name token conventionally-designates an object if and only if the speaker, in producing the token, is participating in a convention of speaker-designating that object, and no other object, with name tokens of that type. (126)

Let us begin by noting that the definition of semantic reference appeals explicitly to speaker's reference. For a proper name token to refer, in fact, the speaker who produced it must be participating in a (pre-existing, I assume) convention of *speaker's referring* to something by using that name. Thus, there could not be semantic reference (conventional-designation) if there were not speaker's reference (speaker-designation): Devitt is clearly pursuing the Gricean project.

Given this, the first thing that we have to check is whether, unlike Kripke's, Devitt's notion of speaker's reference can indeed be used to define semantic reference in a non-circular way. The fact that the notion is defined in terms of "designating-chains" could lead one to believe that it cannot be so used, since the word "designating" in "designating-chain" might induce the suspicion that designating-chains involve (past) semantic reference (conventional-designation). Here, however, appearances are misleading.

Designating-chains are introduced by Devitt in *Designation* in the following way:

> "underlying" a name token is a "causal chain" "accessible to" the person who produced the token. That chain, like the ability that partly constitutes it, is "grounded in" the object the name designates…. I shall call such a causal chain a … "designating-chain." (1981b: 29)

They are thus characterized: "D[esignating]-chains consist of three different kinds of link: groundings which link the chain to an object, abilities to designate, and communication situations in which abilities are passed on or reinforced (reference borrowings)" (1981b: 64; 2015: 110). What is important to note for our present purposes is that designating-chains underlying a proper name token do not necessarily originate in a baptism or something like that, and do not require what Devitt calls "reference borrowings." For example, in the Smith-Jones case, there is, according to Devitt, a designating-chain underlying the "Jones" tokens produced by the speakers in their colloquy originating in their perception of Smith, although there is another one originating in Jones'

baptism.[27] Hence, some designating-chains do involve (past) semantic reference, but some do not, and this suffices to avoid circularity, as Devitt himself notes in *Designation*:

> Conventions are explained in terms of speaker meanings. Speaker meanings are explained in terms of thought meanings. Thought meanings are partly explained in terms of conventions. We seemed to have a circle. What we really have is more like a spiral, a spiral that starts from crude thought meanings. (1981b: 85)

Thus, we may conclude that unlike Kripke's, Devitt's notion of speaker's reference can indeed be used to explain semantic reference. The remaining, crucial, question is obviously whether the resulting explanation is a good one. To answer, we need to better examine the two definitions Devitt offers.

Devitt's account of speaker's reference is very similar to his "Singular Terms" account of reference *tout court*, hence to Donnellan's historical explanation theory and to the neo-Donnellanian accounts, as Devitt himself recognizes, although with some reservations concerning Donnellan (see Devitt forthcoming *a*). Basically, it is an account of the *state of mind* leading to the production of a proper name token, or, as Devitt also likes to say, of the *thought* the speaker is *expressing* by the token, as the following comment to an example clearly shows: "The token [speaker-]designated that person in virtue of being immediately caused by a thought that is grounded in that person by a designating-chain" (2015: 111). In fact, Devitt's causal theory of speaker's reference bears one of the extreme consequences of Donnellan's historical explanation theory: once one has a thought about an individual, he or she can express the former and (speaker-)refer to the latter by whatever name he or she wants.[28] The token he or she then produces (speaker)-refers to the individual the thought is about, no matter how that individual was baptized and what any preceding tokens of the same name referred to:

> A person can, of course, speaker-designate an object by a name without there being any convention of so doing. All that is required is that a token of the name have underlying it a designating-chain grounded in the object. So I could now speaker-designate Aristotle with any old name simply on the strength of the link to Aristotle that is constitutive of my ability to designate him by 'Aristotle.' (2015: 120)

The main difference between Devitt's view in *Designation* (and later articles) and his preceding (as well as Donnellan's and the neo-Donnellanians') view is that he does not claim any more that the state of mind leading to the production of a proper name token, or the thought the speaker expresses by the token, determines what the token *semantically* refers to. For a proper name token to semantically refer to something, in fact, the speaker producing it must be participating in a *convention* of speaker-referring to it with tokens of that type, as Devitt's

[27] This is why, according to Devitt (see footnote 9 above), those tokens *partially* speaker-refer to Smith and *partially* speaker-refer to Jones.

[28] For more elaboration on this, see Bianchi forthcoming.

definition of semantic reference (*Conventional-Designation*) states. No convention, no semantic reference.

What about Devitt's account of speaker's reference? I must confess I do not have much against it, except that I do not consider it an account of … *reference*, of any sort. As I said, it is an account of the *state of mind* leading to the production of a proper name token, or of the *thought* the speaker is *expressing* by the token. More specifically, it is a *causal* account of that state of mind's, or of that thought's, *aboutness*. Now, that aboutness is to be accounted for in causal terms is something I wholeheartedly agree with. I am also quite comfortable with the so-called *representational theory of mind*, and with the *language of thought hypothesis*, which provide the theoretical background to Devitt's causal account.[29] One minor perplexity I have concerns Devitt's apparent identification of the (complex) state of mind leading to the production of a proper name token with the thought the speaker expresses with the token. A consequence of this is Devitt's idea that there can be *partial* speaker's reference.[30] Consider the Smith-Jones case once again. As we have already seen, Devitt claims that "[b]ecause there are d-chains to both Jones and Smith, … neither was the speaker's referent but each was his *partial* referent" (1981a: 515). While I agree that the (complex) state of mind leading to the speaker's production of that token of "Jones" *concerned* both Jones and Smith, I find it more natural to say that the thought he expressed on that occasion was only *about* Jones, although it was brought about by a number of other thoughts of his, some of which were (fully) about Jones and some of which were (fully) about Smith. But this is perhaps only a verbal disagreement, and in any case it does not bear directly on the issues I am interested in here.

My main point, as I suggested, is simply that Devitt's speaker's reference does not seem to have much to do with reference. Devitt's is an account of the state of mind leading to the production of a proper name token, and as such can help explain language use. For example, it can help explain why, in the Smith-Jones case, the two speakers use the name "Jones" when they see Smith in the distance raking the leaves (note, however, that the explanation also needs to appeal to the fact that Jones is the semantic referent of "Jones"). But how does all this relate to proper name reference, if we assume, as Devitt does, that Kripke's chain of communication picture is on the right track?

Of course, we already know Devitt's answer. He is pursuing the Gricean project, hence he aims at explaining proper name semantic reference in terms of what he calls "speaker's reference." The specific

[29] For my endorsement of (a peculiar version of) the representational theory of mind and the language of thought hypothesis, see Bianchi 2005 and 2007. The version is peculiar in that it takes the language of thought to be the language we speak (cf. Field 1978 and 2001). In Designation, Devitt himself came very close to embracing it (1981b: 75–9). My endorsement of it partly explains my resistance to the idea that thought aboutness explains semantic reference (it is the other way around!) However, my criticisms below of Devitt's account of proper name semantic reference (and, more generally, of the Gricean project) are independent of this.

[30] For an early criticism of Devitt's idea of partial reference, see McKinsey 1976.

form of this explanation is indicated in Devitt's definition of semantic reference (*Conventional-Designation*): a proper name token semantically refers to an object if and only if the speaker, in producing the token, is participating in a convention of speaker-referring to that object, and no other object, with name tokens of that type. Thus, according to Devitt *there could not be (proper name) semantic reference if there were not conventions of speaker-referring, in which the producers of proper name tokens participate*. But, is this really what we should say about proper name semantic reference, if we assume that Kripke's chain of communication picture is on the right track?

Consider the famous passage where Kripke introduces his picture in the second lecture of "Naming and Necessity":

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman was a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can't identify him uniquely. He doesn't know what a Feynman diagram is, he doesn't know what the Feynman theory of pair production and annihilation is. Not only that: he'd have trouble distinguishing between Gell-Mann and Feynman. So he doesn't have to know these things, but, instead, a chain of communication going back to Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link, not by a ceremony that he makes in private in his study: 'By "Feynman" I shall mean the man who did such and such and such and such'. (1972: 298–9 (1980: 91–2))

I take this to be a picture of how proper names *semantically* work (as we saw in Section 2, in "Naming and Necessity" Kripke also introduces his distinction, but he was certainly not aiming at providing a picture of speaker's reference in the second lecture). But note that in the picture, no mention is made either of speaker's reference or of conventions, even less, of course, of conventions of speaker-referring. Much more simply, a name come to be introduced by someone for something, after which it is spread around through use. And even if one wished to see in this spread the establishment of a convention, he or she should acknowledge that according to Kripke's picture proper name tokens already (semantically!) refer before the convention gets established. The fact is that semantic reference, at least as far as proper names are concerned, is basically a *historical* relation. Kripke himself summarizes the point in the following way:

> In general our reference depends not just on what we think ourselves, but on other people in the community, the history of how the name reached one, and things like that. It is by following such a history that one gets to the reference. (1972: 301 (1980: 95))

Let me note, by the way, that one almost immediate consequence of Kripke's picture is that the reference of a proper name is not determined or *fixed* anew every time a token of it is produced. On the contrary, any token of it, except for the first, *inherits* its reference from preceding ones, to which it is historically connected. Again, no speaker's reference, and no participation in a convention of speaker-referring, seem to be involved in this.

Now, as Kripke himself admits, his characterization is "far less specific than a real set of necessary and sufficient conditions for reference would be" (1972: 300 (1980: 93)). To develop his picture into a definition of proper name semantic reference, many details need to be filled in, and many problems settled.[31] Thus, Devitt might sensibly argue that it is when we try to fill in the details and settle the problems that we realize that we have to appeal to speaker's reference and conventions of speaker-referring. But is it really so?

Consider the introduction of a name—in Devitt's terms, *reference fixing*—first. Devitt might argue that it requires what he calls "speaker-designation": to introduce a name for something, one must speaker-refer to it with the name. Which means, roughly, that the introduction must be "immediately caused by a thought that is grounded in [it] by a designating-chain"—the state of mind leading to the production of the 'introductory' token must be about the individual that gets named. Now, there is no doubt that standard name introductions involve a lot of mental goings-on in the introducer(s)'s minds, and I have no difficulty in conceding that very often the individual that gets named is the one speaker-referred to, in Devitt's sense. But is it always so? Reference fixing is a complex phenomenon, with various factors often playing a role.[32] I take it to be possible for a name to be introduced for something that is not speaker-referred to by the introducer(s), or for a name to be introduced without any speaker's reference being made, or even without any mental goings-on taking place—couldn't some sort of sophisticated machine mechanically and more or less randomly assign names? What is important in Kripke's picture of proper name semantic reference, I would like to say, is *that* a name is introduced—a relation between a name and an individual is established—not *how* the name is introduced—*how* the relation is established.

Consider next the spread of a name after its introduction—in Devitt's terms, reference borrowing. Devitt might argue that it requires conventions of speaker-referring: to semantically refer to something with a token of an already introduced name for that something, one must participate in a convention of speaker-referring to it with name tokens of that type. Now, I have nothing against talking of conventions in this case, provided only that one admits 'infra-personal' conventions—

[31] Notable among the latter is, of course, the one raised by Evans with the Madagascar case, which we mentioned in Section 3.

[32] See Martí 2015: 86–89 for some converging considerations.

conventions that do not involve other people.[33] But are they really conventions concerning speaker's reference (or, as Devitt also likes to say, conventions regarding the expression of thoughts)? And for a name token to semantically refer to something, is it really necessary that it be produced by someone who is participating in such a convention? According to Devitt, "[*p*]*articipating in a convention* [of such a kind] concerns the process of a speaker using the name because she has a disposition, dependent on the dispositions of others, to use it to express thoughts grounded in a certain object" (2015: 126). But, nowadays many semantically referring name tokens are literally produced by copying machines. Do these machines really have any disposition to express thoughts? Again, what is important in Kripke's picture is that most proper name tokens semantically refer in virtue of a certain historical connection they have with other tokens of the same name. It may be difficult to say exactly what this historical connection amounts to (for my attempt, which uses Kaplan's (1990) notion of *repetition*, see Bianchi 2015), but appealing to conventions of speaker-referring seems to me a false step.

Of course, much more than this needs to be said about both reference fixing and reference borrowing, but even these scattered considerations seem to me to cast a dark shadow on Devitt's explanation of proper name semantic reference in terms of speaker's reference. Those who believe that Kripke's chain of communication picture is on the right track, as Devitt and I certainly do, should rather abandon the Gricean project.

5. *Conclusion*

Let me recapitulate. In this paper, I have critically examined the distinction between speaker's reference and semantic reference, a distinction that was introduced in the philosophical debate by Kripke in the Seventies and is now taken for granted by most philosophers of language. I first focused on Kripke's definition of speaker's reference, and used the example of the structurally similar definition of traveler's going to argue that it does not justify the claim that reference comes in two different sorts. Then, I briefly considered the theoretical roles the notion of speaker's reference is supposed to play. According to Kripke, in fact, the notion has both a critical and a constructive use. From the critical point of view, it can serve as a "tool to block postulation of unwarranted ambiguities"; from the constructive one, it can help explain the puzzling phenomenon of reference change. But a quick look at how the notion would play these theoretical roles reinforced my doubts about the claim that reference comes in two different sorts. Finally, I took a closer look at another major role many philosophers assign to speaker's reference, that of contributing to the explanation of semantic reference. Interpreting Kripke's distinction in this way amounts to endorsing what I have called, for ob-

[33] For the reasons of this proviso, see footnote 1, footnote 19, and especially Martí 2015: 89–91.

vious reasons, the Gricean project. The Gricean project is in fact the most promising, if not the only, way to use Kripke's distinction to claim that reference comes in two different sorts. Concerning this, however, I first noted that Kripke's definition of speaker's reference makes the notion incompatible with the project: since speaker's reference is defined in terms of semantic reference, it cannot be used to explain it. Then, I examined Devitt's causal theory of proper names, which offers a detailed account of both speaker's and semantic reference. Devitt explicitly pursues the Gricean project: unlike Kripke, he defines speaker's reference without appealing to semantic reference, and then explains the latter in terms of the former. However, I argued that there is at least some tension between this explanation and Kripke's chain of communication picture, a picture Devitt's causal theory was meant to develop.

My tentative conclusion is that those philosophers who believe Kripke's chain of communication picture is on the right track, as many do nowadays, should abandon the Gricean project, and with it the claim that reference comes in two different sorts. And perhaps, even the claim that the mind of the speaker plays a role in determining the reference of the proper name tokens he or she produces. If we stop distinguishing between speaker's reference and semantic reference, we may hope to make some progress in our understanding of reference.[34]

References

Almog, J. 2012. "Referential Uses and the Foundations of Direct Reference." In J. Almog and P. Leonardi (eds.). *Having in Mind: The Philosophy of Keith Donnellan*. Oxford: Oxford University Press.

_____, 2014. *Referential Mechanics: Direct Reference and the Foundations of Semantics*. Oxford: Oxford University Press.

Almog, J., Nichols, P., and Pepp, J. 2015. "A Unified Treatment of (Pro-) Nominals in Ordinary English." In A. Bianchi (ed.). *On Reference*. Oxford: Oxford University Press.

Bertolet, R. 1981. "Kripke's Speaker's Reference." *Analysis* 41: 70–72.

_____, 1987. "Speaker Reference." *Philosophical Studies* 52: 199–226.

Bianchi, A. 2005. "Words as Concepts." In J. J. Acero and P. Leonardi (eds.). *Facets of Concepts*. Padova: Il Poligrafo.

_____, 2007. "Speaking and Thinking (Or: A More Kaplanian Way to a Unified Account of Language and Thought)." In M. Beaney, C. Penco, and M. Vignolo (eds.). *Explaining the Mental: Naturalist and Non-Naturalist Approaches to Mental Acts and Processes*. Newcastle: Cambridge Scholar Publishing.

[34] Parts of this article are taken from "Speaker's Reference and Semantic Reference: A Theoretically Useful Distinction?," an unpublished paper that I presented at the *First Parma Workshop on Semantics and Pragmatics* (September 2011) and the *Workshop on Reference and Frege Puzzles* (Umeå, November 2012). I am grateful to all those who intervened on those occasions. I would also like to thank Joseph Almog, Antonio Capuano, Michael Devitt, Dunja Jutronić, Paolo Leonardi, and Stephen P. Schwartz for their comments on an earlier draft (and Dunja for her patience as well!).

_____, 2011. "Reference and Descriptions." In J.-O Östman and J. Verschueren (eds.). *Handbook of Pragmatics Online, 2011 Installment*. Amsterdam: John Benjamins. Reprinted in M. Sbisà, J.-O Östman, and J. Verschueren (eds.). *Philosophical Perspectives for Pragmatics*. Amsterdam: John Benjamins, 2011 (page numbers given relate to this volume).

_____, 2012. "Two Ways of Being a (Direct) Referentialist." In J. Almog and P. Leonardi (eds.). *Having in Mind: The Philosophy of Keith Donnellan*. Oxford: Oxford University Press.

_____, 2015. "Repetition and Reference." In A. Bianchi (ed.). *On Reference*. Oxford: Oxford University Press.

_____, forthcoming. "Reference and Causal Chains." In A. Bianchi (ed.). *Language and Reality From a Naturalistic Perspective: Themes from Michael Devitt*. Cham: Springer.

Bianchi, A. and Bonanini, A. 2014. "Is There Room for Reference Borrowing in Donnellan's Historical Explanation Theory?" *Linguistics and Philosophy* 37: 175–203.

Biro, J. I. 1979. "Intentionalism in the Theory of Meaning." *Monist* 62: 238–258.

Black, M. 1973. "Meaning and Intention: An Examination of Grice's Views." *New Literary History* 4: 257–279.

Capuano, A. 2012. "From Having in Mind to Direct Reference." In W. P. Kabasenche, M. O'Rourke, and M. H. Slater (eds.). *Reference and Referring*. Cambridge: MIT Press.

_____, 2018. "In Defense of Donnellan on Proper Names." *Erkenntnis*, https://doi.org/10.1007/s10670-018-0077-6.

Devitt, M. 1974. "Singular Terms." *Journal of Philosophy* 71: 183–205.

_____, 1976. "Semantics and the Ambiguity of Proper Names." *Monist* 59: 404–423.

_____, 1981a. "Donnellan's Distinction." *Midwest Studies in Philosophy* 6: 511–524.

_____, 1981b. *Designation*. New York: Columbia University Press.

_____, 2004. "The Case for Referential Descriptions." In M. Reimer and A. Bezuidenhout (eds.). *Descriptions and Beyond*. Oxford: Clarendon Press.

_____, M. 2015. "Should Proper Names Still Seem So Problematic?" In A. Bianchi (ed.). *On Reference*. Oxford: Oxford University Press.

_____, Forthcoming a. "Stirring the Possum: Responses to the Bianchi Papers." In A. Bianchi, ed., *Language and Reality From a Naturalistic Perspective: Themes from Michael Devitt*. Cham: Springer.

_____, Forthcoming b. *Overlooking Conventions: The Trouble with Linguistic Pragmatism*. Cham: Springer.

Donnellan, K. S. 1966. "Reference and Definite Descriptions." *Philosophical Review* 5: 281–304.

_____, 1970. "Proper Names and Identifying Descriptions." *Synthese* 21: 335–358.

Evans, G. 1973. "The Causal Theory of Names." *Aristotelian Society Supplementary Volume 47*: 187–208. Reprinted in G. Evans, *Collected Papers*. Oxford: Clarendon Press, 1985 (page numbers given relate to this volume).

Field, H. 1978. "Mental Representation." *Erkenntnis* 13: 9–61. Reprinted in H. Field, *Truth and the Absence of Fact*. Oxford: Clarendon Press, 2001.

_____, 2001. "Postscript: Mental Representation." In H. Field, *Truth and the Absence of Fact*. Oxford: Clarendon Press.

Geach, P. T. 1962. *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Ithaca: Cornell University Press, third edition 1980.

Grice, H. P. 1957. "Meaning." *Philosophical Review* 66: 377–388. Reprinted in P. Grice, *Studies in the Way of Words*. Cambridge: Harvard University Press, 1989 (page numbers given relate to this volume).

_____, 1968. "Utterer's Meaning, Sentence-Meaning, and Word-Meaning." *Foundations of Language* 4: 225–242.

_____, 1969. "Utterer's Meaning and Intention." *Philosophical Review* 78: 147–177.

_____, 1982. "Meaning Revisited." In N. V. Smith (eds.). *Mutual Knowledge*. New York: Academic Press. Reprinted in P. Grice, *Studies in the Way of Words*. Cambridge: Harvard University Press, 1989 (page numbers given relate to this volume).

Hanks, P. 2019. "Reference as a Speech Act." In J. Gundel and B. Abbott (eds.). *The Oxford Handbook of Reference*. Oxford: Oxford University Press.

Hinchliff, M. 2012. "Has the Theory of Reference Rested on a Mistake?" In W. P. Kabasenche, M. O'Rourke, and M. H. Slater (eds.). *Reference and Referring*. Cambridge: MIT Press.

Kaplan, D. 1989. "Afterthoughts." In J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan*. Oxford: Oxford University Press.

_____, 1990. "Words." *Aristotelian Society Supplementary Volume* 64: 93–119.

Kripke, S. 1972. "Naming and Necessity." In D. Davidson and G. Harman (eds.). *Semantics of Natural Language*. Dordrecht: Reidel.

_____, 1977. "Speaker's Reference and Semantic Reference." *Midwest Studies in Philosophy* 2: 255–276.

_____, 1980. *Naming and Necessity*. Reprint with a new preface of Kripke 1972. Oxford: Blackwell.

McKinsey, M. 1976. "Divided Reference in Causal Theories of Names." *Philosophical Studies* 30: 235–242.

Martí, G. 2015. "Reference without Cognition." In A. Bianchi (ed.). *On Reference*. Oxford: Oxford University Press.

Pepp, J. 2012. "Locating Semantic Reference." UCLA Ph.D. dissertation.

_____, 2019. "What Determines the Reference of Names? What Determines the Objects of Thought." *Erkenntnis* 84: 741–759.

Sainsbury, M. 2015. "The Same Name." *Erkenntnis* 80: 195–214.

Suppes, P. 1986. "The Primacy of Utterer's Meaning." In R. E. Grandy and R. Wagner (eds.). *Philosophical Grounds of Rationality: Intentions, Categories, Ends*. Oxford: Clarendon Press.

Wettstein, H. 2004. *The Magic Prism: An Essay in the Philosophy of Language*. Oxford: Oxford University Press.

Wulfemeyer, J. 2017. "Reference-Shifting on a Causal-Historical Account." *Southwest Philosophy Review* 33: 133–142.

Yu, P. 1979. "On the Gricean Program About Meaning." *Linguistics and Philosophy* 3: 273–288.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

The Qua *Problem and the Proposed Solutions*

DUNJA JUTRONIĆ
University of Maribor, Maribor, Slovenia
University of Split, Split, Croatia

One basic idea of the causal theory of reference is reference grounding. The name is introduced ostensively at a formal or informal dubbing. The question is: By virtue of what is the grounding term grounded in the object qua-horse and not in the other natural kind whose member it is? In virtue of what does it refer to all horses and only horses? The problem is usually called the qua problem. What the qua problem suggests is that the causal historical theory in the final analysis depends on some kind of unexplained intentionality. This is a great problem since the whole project is an attempt to explain intentionality naturalistically. In this paper, I have two aims: (i) to discuss the most important attempts at solving the qua problem; and (ii) to evaluate the solutions. (i) I focus on the following attempts for the solution of the qua problem: Sterelny (1983), Richard Miller's (1992), mentioning briefly more recent attempts by Ori Simchen (2012) and Paul Douglas (2018). I also concentrate on the attempts in mind and brain sciences as presented by Penelope Maddy (1983) and more recently by Dan Ryder (2004). (ii) In evaluating the solutions, I argue that when a metaphysical question "what is to name" is replaced/or identified with the question about the mechanism of reference, namely "in virtues of what does a word attach to a particular object", then the final answer will/should be given by neurosemantics. The most promising attempt is Neander's (2017), based on the teleological causal explanation of preconceptual content to which the conceptual can be developed, as Devitt and Sterelny suggested in their work (1999).

Keywords: *Qua* problem, reference grounding, mechanisms of reference, intentionality, neuroscience, neurosemantics.

1. *Introduction: the causal theory of reference and the* qua *problem*

According to the representatives of the description theory (Frege 1893, Russell 1905) reference is determined by a description or descriptions that the speaker can give for the person or the thing. According to causal theorists, Keith Donellan (1972), Saul Kripke (1980), Hilary Putnam (1975) and further elaborated by Michael Devitt (1981), reference is not determined by descriptions but by a causal chain that links the speaker to the person or a thing. Here I concentrate on the theory elaborated by Devitt and Sterelny (1999). The first attempt was given in Devitt (1981).[1]

One basic idea of the causal theory of reference is reference grounding. The name is introduced ostensively at a formal or informal dubbing. The other basic idea of the causal theory is reference borrowing. Hearers can gain the ability to use the name in conversation by the fact that they are told what the term is by others who have also learned about it from somebody else. The chain goes back to the grounder.

The *qua* problem is the problem arising in reference grounding. The problem is the problem of discovering *in virtue of what* a term is grounded in the cause of a perceptual experience *qua*-one-kind and not *qua*-another (Devitt and Sterelny 1999: 79–82).

Devitt favorite example is the cat named 'Nana'. The use of that name was grounded in virtue of perceptual contact with that particular cat. That is, the name refers to that cat in virtue of a grounder/baptizer having had perceptual contact with her. However, the contact is not with that entire particular cat, some contact with Nana could be perhaps as she peers around a corner. The question that the *qua* problem poses is why 'Nana' refers to the whole individual and not an individual time-slice or an undetached part of her. The same problem arises in case of a natural kind term such as 'horse'. The term can be gounded in a couple of horses or even one horse, but horses are not only horses, they are vertebrates, they are mammals. They are members of different/many natural kinds. By virtue of what is the grounding term grounded in the object *qua*-horse and not in the other natural kind whose member it is? In virtue of what does it refer to all horses and only horses? Why does the term applied in such groundings not project to other members or these other natural kinds? The problem is even worse. What limits such kinds to only natural kinds? Object of 'horse' could be grounded as a pet, wooden toy, etc. Why do we not gound them as members of such kinds? The term '*qua* problem' has been coined by Kim Sterelny (1983).

[1] A short power point presentation of this paper was given at the International conference: *Devitt's 80th. Many Faces of Philosophy* held in Maribor (May 9–10, 2018). A much shorter version of this article will appear in Borster and Todorović (ed.) forthcoming, celebrating Devitt's 80th birthday.

The solution that Devitt and Sterelny (1987/1999) explore is that the baptizer/grounder needs to have some idea—some mental content—about the thing that she/he is naming. For example, you need to have an idea that you are naming a whole individual, despite the limitations of your causal contact with it. Having such an idea/mental content allows the descriptive element to enter the causal chain, so Devitt and Sterelny consider compromise with descriptivists to solve the *qua* problem. "It seems that the grounder must, at some level, 'think of' the cause of his experience under some general categorical term like 'animal' or 'material object'" (1999: 90–93). The supposition is that the individual or kind actually named must be the individual or kind the speaker intends to name, so that facts about the speaker's beliefs and concepts enter into the determination of reference. Only indefinite descriptions are required along with some causal historical contact. What the *qua* problem suggests is that the causal historical theory in the final analysis depends on some kind of unexplained intentionality. This is a great problem since the whole project is actually an attempt to explain intentionality naturalisticaly.

Thus Devitt and Sterelny (1999) say they are torn between two explanations of reference. The interest in the final explanation takes them away from the descriptive theories towards causal theories. But the historical-causal theory of reference has a deep problem, the *qua* problem which, as Devitt and Sterelny say, does not seem to have the resources to solve. Later in 2002 Devitt says: "I have struggled mightily with this problem (1981a: 61–4; Devitt and Sterelny 1999: 79–80), but I now wonder whether this was a mistake: perhaps the problem is more for psychology than philosophy" (2002: 115, note 15).

The question is then: Can the *qua* problem be solved and is it a philosophical problem?

I proceed as follows: In section 2. I focus on the following attempts for the solution of the *qua* problem: Sterelny (1983); Richard Miller (1992); and two recent ones by Simchen (2012) and Douglas (2018). In section 3. I concentrate more on the attempts to the solution of the *qua* problem in the sciences, as presented by Penelope Maddy (1983) and Dan Ryder (2004). In section 4. I look more closely into Devitt and Sterelny (1999) and Karen Neander (2017) proposal and suggestions. Section 5. is a reflection on the mechanisms of reference and section 6. is the Conclusion.

2. *The* qua *problem and (possible) pure causal solutions*

There has been a number of attempts at solving the *qua* problem. I will look into, to what I consider, the most important ones. And chronologically I start with Sterelny (1983).

2.1 *Kim Sterelny (1983)*

Sterelny's solution to the *qua* problem from 1983 adds two additional requirements on the grounder. First, the grounding requires not just contact with the sample of a general kind but the "assignment of causal powers to the kinds" (1983: 116). The grounder must have in mind a set of causal powers of the sample, possibly the observable ones. These causal powers are grounded in some structure which is common to a certain kind (e. g. cathood for cats). So, for example, if the grounder has in mind something like 'mouse catcher', or 'coachroach-eater', she will be able to ground the term 'cat' in the sample of the kind cat.

The second requirement is the possession or acquisition of recognitional capacities of a general category, i.e., the grounder of the name must have acquired a reliable recognitional capacity for the kind referred to. "One can ground a term on a kind only if one has the ability to discriminate, reasonably reliably, members of the kind" (1983: 116). Thus, the speaker will ground the term only if he has in mind the causal symptoms of kindhood and if he has the ability to discriminate those symptoms. Talking about the recognitional capacities to discriminate general categories, Sterelny says that they are not psychological states individuated internally but that they are constituted by the way an individual is embedded in his physical and probably social environment (1983: 117). The individual simply identifies. He has a learned perceptual capacity similar to an ability to recognize shapes. In that sense, it is only knowledge-*how*.

Miller (1992) who himself tries to offer a better solution to Devitt and Sterenly's solution from 1987 rightly notices that Sterelny's solution to the *qua* problem has the following weak point. What is problematic is the second requirement, i.e., the requirement that the grounder has a reliable ability to discriminate members of the kind. Miller says: "The reliable ability to discriminate kangaroos will not serve to pick out kangaroos *qua* kangaroos because our hypothetical grounder of the term also discriminates speedy herbivores, hopping marsupials, tourist attractions, and food sources. Since the speaker has the ability to discriminate all these classes, reference to these classes is not ruled out by the restriction as it stands" (1992: 428). Miller does not mention the first requirement, 'the assignment of causal powers to the kinds', i.e., that the grounder must have in mind a set of causal powers of the sample. In my view, it is rather mysterious how the grounder has the causal powers in mind when the causal powers can be multiple: 'cathood', 'animalhood'. How does the grounder decide? That is the problem that *qua* problem poses, so it cannot be the requirement or the solution to the problem.

2.2 *Richard Miller (1992)*

Miller offers, what he believes is, a purely causal theory of grounding. He argues that Devitt and Sterelny's (1987) descriptive-causal theory

of grounding doesn't work as a theory of reference, but that a purely causal account does. Miller points out that although the problem was first recognized more ten years ago[2] it remains unsolved and largely neglected. He also stresses that the difficulty seems crucial also to causal theories of perception and mental representation. He focuses on reference but says that "causal theories of perception and mental representation unavoidably hover in the background" (1992: 425).[3]

Miller sets himself a task of showing that although Devitt and Sterelny (1987), tentatively explored a compromise with descriptive theory in order to solve the *qua* problem for reference grounding, he thinks that compromise is "unwise" because "no hybrid theory can solve the *qua* problem" (1992: 427). Miller's suggestion rests on Sterelny's reliabilist solution which should be modified to bring out the fact that the sample upon which the term is grounded causes the reliable ability to discriminate the kind in virtue of its membership in the kind itself. The more precise formulation of his solution is the following: *The speaker S can use his perceptual contact with x to ground 'N' on the kind Q if x qua Q causes S to acquire a reliable ability to discriminate Qs.*

This 'tightening up', as Miller puts it, of the causal relation, solves the *qua* problem. Individuals—in Miller's case individual kangaroos—have the causal powers that they do in virtue of the classes to which they belong. Miller stresses that the 'x qua Q causes S' locution needs to be explained. The *qua* problem arises because individuals can correctly be said to belong to many classes. His solution depends on the fact that individuals have causal powers in virtue of their belonging to certain classes. There is no need to look outside the causal powers of things for a solution to the *qua* problem because the *qua is built into the causal powers themselves* (italics mine). The particular stands for whichever class shares the causal nature which brought about the acquisition of the ability to use the name (1992: 429). In other words, to stress once again: The *qua* is built into the causal powers.

One may surely wonder how is the *qua* built into the causal powers themselves? And this is exactly what Miller asks: In virtue of what was the grounding in the natural kind to which the individual belongs and not in any of the other kinds to which it also belongs? His answer is that what the grounder gained was a disposition to think 'kangaroo' when confronted with kangaroos and not a disposition to think 'kangaroo' when confronted with marsupials, tourist attractions or food sources. In virtue of what was the grounding in an individual and not its time-slice? What the grounder gained was a disposition to think 'George' when confronted with George and not a disposition to think 'George' when confronted with the time-slice of George. The individual

[2] And now more than a quarter of a century ago.

[3] He also points out that: "Philosophers who complain that CTR is too sketchy to be worthy of serious consideration ought to examine the detailed and systematic development of the theory in Michael Devitt, *Designation* (New York: Columbia University Press, 1981: 61–63)" (1992: 425, footnote 2).

or kind referred to reliably causes the speaker to think 'N' (1992: 430). However, there is no need for the individual acquiring the name to be aware of the properties which caused her to acquire it. In fact, she will often not be conscious of them at all. Miller concludes: Even children with vocabularies of less than one hundred words do it with ease. It is a brute fact that people learn to react one way to 'dogness' and another way to 'catness' without the need for descriptions. The underlying natures of dogs and cats are causes and our ability to use 'dog' and 'cat' are effects. The descriptions we come up with are mere epiphenomena.[4]

There are a couple of problems with Miller's pure causal suggestion.

1. "The speaker S can use his perceptual contact with x to ground 'N' on the kind Q if x qua Q causes S to acquire a reliable ability to discriminate Qs." This seems to run immediately into the ignorance and error problem, i.e., grounders need not have this ability to discriminate in order to refer. The speaker can refer even when ignorant of what the person or kind really is.

2. Miller says: "There is no need to look outside the causal powers of things for a solution to the *qua* problem because the *qua* is built into the causal powers themselves" (1992: 429). Apart from this claim about something (some stuff?) being built in the causal powers and even if we grant that *qua* is somehow built into the causal powers the grounder still has to think about which causal power is in question. Miller says: "If this had been a marsupial but had not been a kangaroo, it would have caused the speaker to acquire the ability to discriminate marsupials" and "if this had not been a marsupial and had been a kangaroo, it would have caused the speaker to acquire the ability to discriminate kangaroos" (1992: 429). But the grounder is confronted *at the same time* with marsupial and kangaroo. How is the possible fact that *qua* is built into the causal power going to help the grounder? Causal power is built into kangaroos and causal power is built into marsupial. How does the grounder know? Obviously, he has to "think" of one or another. There does not seem to be a straightforward direct or pure causal link.

3. Miller says: "The truth of these referential hypotheses depends on the truth of the counterfactuals: "If this had been a marsupial but had not been a kangaroo, it would have caused the speaker to acquire the ability to discriminate marsupials" and "if this had not been a marsupial and had been a kangaroo, it would have caused the speaker to acquire the ability to discriminate kangaroos" (1992: 429). The counterfactual suggestion has the same problem as stated above in 2. The grounder, again is confronted with both marsupials and kangaroos and the counterfactuals cannot determine which disposition (to think 'kangaroo' or 'marsupial') is going to gain priority in reference fixing. How

[4] He adds: "This ability to react to underlying natures without knowing what they are will probably seem mysterious to descriptivists, but it ought not. Such an ability is obviously present in mammals" (1992: 431).

does the grounder acquires "a recognitional capacity which fits George like a glove fits a hand" (1992: 430–31), is unanswered.

4. Miller mentions the fact that even children with vocabularies of less than one hundred words react to underlying properties. He says that it is a brute fact that people learn to react one way to 'dogness' and another way to 'catness' without the need for descriptions. But the innate ability to react to underlying properties is not a good argument for the *qua* problem since this problem needs the answer *in virtue of what* we react and not which ability makes us react. In sum, Miller's solution is not the solution to the *qua* problem seen as a pure causal mechanism.

There are two more recent attempts which try to solve the *qua* problem by pure causal mechanism, i.e., avoiding intentional element(s) in the grounding and I try to show that they also fail.

2.3 *Ori Simchen (2012)*

Ori Simchen in his article "Necessity and Reference" (2012) takes up a question: Is it possible for a name that in fact names a given individual to have named a different individual? Simchen focuses on the relation between referring tokens (utterances or inscriptions) of proper names and the referents of those tokens. He argues that the relation is a necessary one: a referring token could not have failed to refer to the thing to which it actually refers. It is plausible that a name refers to something only because its referring tokens refer to that thing. Building on this view, Simchen argues that referential intentions necessarily specify the things they actually do, so no referring token of a proper name could have failed to refer to its actual referent. Simchen tries to show how this approach solves the *qua* problem. He says: "We note that the present approach contains a ready response to a version of what Michael Devitt has termed "the *qua* problem" as applied to referential intentions" (2012: 217–218).

In a rather intricate argument Simchen claims that in employing a name referentially, the primary referential intention is a specific attitude even if it is accompanied by a secondary generic attitude in the form of a descriptive intention to refer. There should be difference between primary referential (cognitive) attitude and secondary referential intention and Simchen states that the primary referential intentions are nondescriptive, they are specific cognitive attitudes rather than generic ones (2012: 220). These cognitive attitudes seem to be a matter of necessity. Simchen says: "We conclude that a given token of a referring term refers to what it refers to as a matter of necessity" (2012: 222). On the other hand, referential intentions are different from primary cognitive attitudes which are supposedly nonintentional although it is not clear how. Jessica Pepp in her overview of the collection, when presenting Simchen's article, does not even mention the nonintentional cognitive attitudes which seem to be crucial for the solution of the *qua* problem as seen by Simchen. All she says is that "Simchen's argument

for the necessity of the relation between tokens and referents relies on the view that speakers refer to things in virtue of their intentions to do so" (2012: 18). If the view is that speakers refer to things in virtue of their referential intentions, then one cannot see how this could be the solution of the *qua* problem. Furthermore, if referring term is a matter of cognitive necessity how does this answer the question that the *qua* problem poses, namely, *in virtue of what this cognitive necessity creates a particular referential links*? It is doubtful that Simchen proposal is the solution to the *qua* problem.[5]

2.4 *Samuel Paul Douglas (2018)*

Samuel Paul Douglas in his article *The Qua-problem and meaning scepticism* (2018) offers another solution to the *qua* problem. The article is not primarily concerned with the *qua* problem but considers solutions given to meaning scepticism and tries to see why Kripke (1982) did not consider a causal-theoretic approach to meaning scepticism. I shall mention meaning skepticism problem only in passing, concentrating on Douglas's offered solution to the *qua* problem.

While Kripke (1982) considered a range of solutions to the sceptical paradox, a causal or causal-hybrid type of solution was not among them. It has been argued by Kusch (2006) that this is due to the *qua* problem. Kusch argues that the absence of a possible causal solution was justified since the attempt of solving the *qua* problem leaves the causal response still open to the sceptical challenge. This is because the *qua* problem includes the requirement that the baptizers have some idea of what it is that they are naming and this introduces an intentional element that the sceptic can potentially exploit.

The core question that sceptic asks is the same as the question asked for *qua* problem: What fact makes it the case that a speaker means, or refers to, one thing rather than another. As we saw, the solution that was proposed to the *qua* problem by Devitt is to introduce a descriptive element into the act of baptizing.[6] In other words to repeat, speakers would need to have some idea—some mental content—about of what kind of thing they are dubbing or baptizing. Before offering his own so-

[5] Andrea Sauchelli (2013), in discussing Ami Thomason on existence question mentions that Thomason bases her solution on the solution that Devitt and Sterelny gave. Thomason's introduces something that she calls the *conditions of applications* which are supposed to solve the *qua* problem. Namely, for example, the name 'Hokusai' is grounded and refers successfully because, in the grounding process, the agents responsible for the naming of Hokusai implicitly intended to apply the name to an entity *qua* human being. But like in Simchen's case, Devitt and Sterelny's claim is that, by introducing intentions, the *qua* problem is created. Namely, exactly what is implicitly intended is left unspecified and this underspecification is actually the core of the *qua* problem.

[6] Douglas misquotes Devitt and Sterelny (1999) as Devitt (1991) which is actually Devitt's book on *Realism and Truth* (Oxford: Blackwell) where there is no mention of the *qua* problem.

lution, Douglas mentions the approaches of Sterelny (1983) and Miller (1992). Douglas argues that that the *qua* problem can be overcome in a way that resists sceptical attack by making use of the notion of *assertability conditions*. His approach requires two key premises. The first is that there are conditions under which some assertions made by speakers will be accepted by their linguistic peers, or they will be not accepted, and that these conditions constrain the behaviour of speakers (2018: 75). Let us mention right away that this premise is relevant for reference borrowing thus not for reference fixing and reference borrowing is not a problem in question. The other key premise is that these assertability conditions supervene upon the same causal chain of events that ground reference under Devitt and Sterelny's (1999) account. This is the premise that is relevant to reference grounding. In Douglas's words: "The reference of a term supervenes upon the causal chain of events that connects our use of that term with its referent (this is the point of a causal theory of reference). At any given time, assertability conditions *must* supervene upon that very same causal chain of events" (2018: 76). Douglas offers an example: Consider a hypothetical situation where an individual the linguistic community known as Sam was baptized with a different name—Bob—and the causal chain of events proceeded from there as previously described. If this were the case, the assertability conditions would necessarily be different to those we experience now. If the chain starts with "Bob," and no alternative grounding events occur to change which name refers to that individual, then the assertability conditions are always going to push speakers towards saying "Bob" and not "Sam," because only Bob features in the causal history of that individual in the relevant sense.

First thing to notice is that this example is again more relevant to reference borrowing, (what the speakers in the community are going to do) rather than reference fixing. Douglas says: If the chain starts with "Bob". But the relevant question is not if it starts but *how* it starts, how "only Bob features in the causal history of that individual in the relevant sense." As Devitt puts it, to paraphrase, *in virtue of what has the grounder grounded the term in Bob*. In virtue of what the individual was named Bob and not Sam? Or in Douglas's own words: "… speakers would need to have some idea—some mental content—about what kind of thing they are dubbing or "baptizing." Douglas's solution does not give an answer to the question that is asked. Further on Douglas says: "If the past use of a word has no influence on the present use of a word, or its influence is indeterminate in nature, trying to make sense of language becomes fraught with difficulty" (2018: 76). But past uses are not going to give us an answer to the question how the *first use* of the term was determined. If it was determined by what was in the mind of the baptizer, then the intentional element that the *qua* problem points to is still a pending danger.

Douglas thinks that he has solved the *qua* problem and he says:

"Finally, this principle (assertability condition) needs to be applied to solving the *qua*-problem. *This solution lies in the fact that all the words a baptizer might think in the process of baptizing, are themselves constrained by their causal history* and the resulting conditions under which certain meanings of them can be asserted." (2018: 76, italics mine). What is puzzling is the following: How can *all the words a baptizer might think in the process of baptizing*…be interpreted? Isn't it the case that the *qua* problem in order to be solved by pure causal links has to eliminate the fact that baptizers "think" in the process of baptizing?

3. Qua *problem and the brain sciences*

The first attempt at solving the *qua* problem by adverting to the functioning of the brain and giving the neural explanation was made by Penelope Maddy (1984) in her article 'How the Causal Theorist Follows a Rule'. She is engaged in considering Wittgenstein's views on rule following but almost her whole concern is focused on the *qua* problem. In this section I also discuss a more recent relevant attempt by Dan Ryder (2004).

3.1 *Penelope Maddy (1984)*

Maddy says that her goal is "to suggest that the causal theorist has the beginning to a reply to Wittgenstein's sceptical conclusion" (1984: 464). Maddy is arguing that there is a way to solve sceptical problem without appeal to descriptions or intentional states. I present Maddy's arguments and I want to show that Maddy's suggestion has a lot going for it. She mentions the role of reference borrowing and the historical chain that goes back to the initial baptizing but she rightly concentrates on the moment when the word's reference is fixed where inevitably the *qua* problem looms large. Talking about the natural kind gold and the *qua* problem she points out that "the causal theorist would agree that the gesture of pointing is not enough to pick out the metal as opposed to its shape or color" (1984: 464). Her answer to the *qua* problem relates straightfowardly to the neurological theory. How does the baptizer, for example, perceive and name something as a triangle rather than the apexes of the triangle? Here is the quote of the relevant suggestion in full:

> The evidence suggests that our ability to perceive develops over time by the growth of neural structures called 'cell assemblies.' Repeated viewing of a triangular figure first produces an assembly that responds selectively to apexes, then assemblies for base angles, and finally an integrated assembly that responds to triangles. This large assembly incorporates the others, though they can still function independently. Without these assemblies, the pattern of stimulation from causal contact with a triangle is a short-lived and chaotic buzz; with them, that same pattern of stimulation produces a much longer, more organized reverberation. The development of the triangle assembly is what allows us to see the triangle as a unit, as similar to

> other triangles, to remember it, and so on. In other words, given only the original pattern of stimulation from the triangle, we could only be said to "see" it in the sense in which one "sees" a hidden figure in a complex drawing before one notices it. With the cell assembly, we can be said to perceive the triangle as such. (1984: 465)

Maddy discusses a number of objections that someone (including sceptic) can raise to her suggestion that the answer to the *qua* problem lies in neurology. I mention the most relevant ones for the present discussion.

1. One demand is that the analysis of psychological notions be conceptual and not scientific. The argument is the following: "We could have the psychological properties we do, that we could perceive and refer, with very different bodies, and perhaps even with no bodies at all. If so, even if cell assemblies and such do give a causal account of the mechanisms by which we actually happen to perceive and refer, this sort of account cannot tell us what perceiving and referring actually are" (1984: 467). Maddy gives, in my opinion, a very good answer to this objection and that is the following: "But if the point at issue is whether or not our reference is determinate, all that is needed is an account of *how this is possible*, then there is no reason such an account need be conceptual rather than scientific" (1984: 468, italics mine).

2. In her answer to the argument that there could be no reference without a community of referrers, she rightly says that this fact does not establish stronger conclusion that the practice of this community is the mechanism that determines reference (1984: 468). She says: "though conceptual analysis may reveal that referring is a practice employed by a linguistic community, the referents of particular expressions in that community's language might still depend on mechanisms peculiar to that community and the world it inhabits: no reference without community, but community reference determined by community-specific mechanisms and circumstances" (1984: 468). What is important to notice here is that Maddy puts great stress on specific mechanisms by which reference is determined. In her case these are neurological mechanisms that with learning experience actually come to be "wired in" procedure that the baptizer simply obeys. She argues that reference is not indeterminate since it is determined by various neural and causal facts: "cell assemblies and causal account of the mechanisms."[7]

3. Kusch (2006) criticizes Maddy's solution. He says that "at first sight it seems as if Maddy is able to solve the *qua* problem in a way that avoids descriptions and other intentional items. In her theory, the work of fixing the level and scope at which the baptizing occurs is done

[7] In answering the sceptic (i.e. henchman) she says: "Thus there is a fact of the matter about which of us in the object level debate -me or the henchman- is right. I may not be able to convince the henchman that I am the one who's right, I may not even be absolutely certain at the meta-level about which of us is right, *but there is a fact about which one of us is right and one of us is wrong*. This is what Wittgenstein (i.e. sceptic) denied" (1984: 469, italics mine).

by non-intentional items such as the stimulation of cell assemblies in the brain" (2006: 135). But Kusch thinks that Maddy's solution fails to specify the nature of the relation between brain events and mental states. His main criticism is that Maddy does not tell us how these brain events relate to mental states. He considers two options that Maddy could go and finds them both unpalatable. One option is that our mental states are reduced to events in the brain. The other is to go down the road of eliminativism. Perhaps Maddy would prefer to be an eliminativist about intentional states but he finds this an extreme view. He concludes therefore that Maddy's proposal for improving the causal theory of reference fails and that Kripke (1984) was right not to discuss the causal theory of reference since it is unworkable as an answer to the sceptical argument.

We cannot go into the discussion of Kusch's suggestion about reductionism or eliminativism but Maddy surely does not go for eliminativism but for reductionism. Maddy does not neglect the question about the relation of the physical and psychological and one of her answers to possible objections that there are no type-type correlations (or identities) between psychological and physical states she says: "It isn't necessary that your cell assembly for triangles be physically similar to mine; all that is needed is for the patterns of neural stimulation triangles produce in me to belong to a single physical type. This much is assumed by the fairly well supported scientific theory of cell assemblies" (1984: 466–467).

In sum, Maddy is combining the causal theory of reference with neuroscience. Her goal is to suggest that there is a way of solving the *qua* problem without appeal to descriptions or to the intentional states, suggesting that the answer lies in neurology. What the baptizer has named will be answered by his brain state, his cell assemblies. Going back to our example of naming the cat "Nana" depending on whether the baptizer is focusing on the cat, or the color of the cat, or the cat as an animal, the brain of the baptizer will be in different states. The perception is linked to different cell assemblies in the brain. And thus, there will be a fact of the matter as to whether the baptizer "meant" the sample for his baptismal act to be the cat, or color or an animal.

3.2 *Dan Ryder (2004)*

Before going back to more philosophical suggestions for solving the *qua* problem, I want to look into a much more recent attempt similar to Maddy's, i.e., the attempt which relies again on neuroscience and computational theory. Dan Ryder in his 2004 article under the title "SINBAD Neurosemantics: A Theory of Mental Representation" presents an account of mental representation based upon the 'SINBAD' theory of the cerebral cortex. He says: "The 'neurosemantic' theory that I present is derived from the SINBAD model of the cortex… 'SINBAD' stands for 'Set of INteracting BAckpropagating Dendrites'; it is a computational

theory of cortical plasticity based on functional considerations as well as anatomical and physiological evidence. If the theory is correct, networks in the cortex have a powerful tendency to structure themselves isomorphically with regularities in their environment" (2004: 212). We cannot go into details of SINBAD but here is the main outline of the idea in Ryder's own words:

> Here then, in brief summary, is how SINBAD networks operate. The multiple dendrites on a SINBAD cell must find functions of their inputs that are correlated. Assuming these correlations are not accidental, the cell will tune to their source. In tuning to a source of correlation, a cell will provide other cells with a useful input, i.e. an input that helps their dendrites to find correlated functions. Thus, these further cells, in turn, tune to sources of correlation, and the process repeats. The end result of this complex multiple participant balancing act is that a SINBAD network comes to be dynamically isomorphic to the environment from which it receives inputs. (2004: 222–223)

Ryder concludes that once one understands the underlying SINBAD mechanism, it is relatively simple to understand, in basic outline, the theory of mental representation that emerges from it. SINBAD cells have the purpose, job, or teleofunction of yielding reliable 'predictions', by participating in internal dynamic structures that are isomorphic to the environment. Dan Ryder's SINBAD theory of content appeals to developmental and learning history but focuses primarily on changes at the neural level. Each neuron in the brain receives incoming signals through branch-like structures called 'dendrites'. He says: "Since the cerebral cortex is the seat of the mind, this gives us some reason to believe that SINBAD representation realizes mental representation in us, and other creatures with a cerebral cortex" (2004: 232).[8]

What is important for our discussion is that Ryder claims to show how SINBAD neurosemantics can provide accounts of the *qua* problem.[9] Suppose multiple encounters with horses cause a SINBAD cell to acquire matching dendritic functions—is it a horse stand-in or an animal stand-in? Here is the explanation or solution of the *qua* problem.

> There will normally be a fact of the matter which kind explains how a cell has acquired its predictive abilities. *The kind horse and the kind animal are sources of different sets of multiple correlations that have different underlying (evolutionary) explanations—that is why they are distinct kinds.* Horses tend to neigh, are usually domesticated, have a particular shape, particular eating habits, hooves, manes, etc. Animals are characterized by a more abstract set of correlated features with a more ancient evolutionary explanation for their coherence: the capacity for spontaneous motion, a range of sizes, a disjunction of typical methods of locomotion, a range of typical colours, and so on. When a cell's representational content is determinate, its dendritic correlations will be explained by (a part of) one of those sets of

[8] More on Ryder in Rupert (2008). Rupert sees Ryder's approach as compatible to his causal developmental theory.

[9] He also has comments related to misrepresentation, equivocal representation, twin cases, and Frege cases.

> correlated properties rather than the other. *It will either be the properties whose correlation typifies horses, or the properties whose correlation typifies animals that will have historically guided the cell to equilibrium by causing synaptic activity.* (2014: 233–234, italics mine)

There is a noticed similarity with Maddy's suggestion in stressing the workings of the brain in the attempt to explain the grounding mechanisms. More on this point in the conclusion.

4. *Back to the philosophical solutions*

4.1 *Kim Sterelny (1990)*

The unsolved *qua* problem prompted Sterelny at another attempt (1990: 124–137). Sterelny suggests that the *qua* problem might be solved by adding the teleological element to the basic descriptive-causal solution that Devitt and he proposed in 1987. Sterelny believes that Kripkian story may not be right story for primitive content, but rather plays a role in the explanations of more cognitively sophisticated structure whose content presupposes a conceptual backdrop. The proposal is to add teleological elements to the causal story. Sterelny says that "there will be an important teleological element in our total theory of mental representation, though any attempt to extend the teleological story to the human propositional attitude faces the most appalling difficulties" (1990: 138).[10] Since his proposal is incorporated in the proposal of Devitt and Sterelny (1999) I discuss it in the next section.[11] I revert to Sterelny (1990) in section 4.3 for more examples.

4.2 *Devitt and Sterelny's (1999) proposal*

As we saw, the *qua* problem does not only concern kinds but the *qua* problem also concerns part-whole ambiguity. What has the grounder named: rabbit, parts of rabbits? Or in the vivid example by Sterelny: "Why is my concept of Mick Jagger a name for Jagger, rather than Jagger's voice? Or Jagger's lips?" (1990: 116). There must be something in the mental state of the grounder which determines that the term has been grounded *via* perceptual experience as something as a whole object and a member of a particular kind.[12] Devitt and Sterelny rightly say that it is neither useful nor sufficient to say that it is the grounder's intentions that makes is so. In virtue of what did the grounder intend the whole object? "It seems that the grounder must, at some level, 'think of' the cause of his experience under some general categorial item like 'animal' or 'material object'. It is because he does so that the

[10] See also Devitt and Sterelny (1999: 101).

[11] For more on Sterelny (1990) see Jutronić (2000).

[12] See the most recent exchange on this issue between Reimer and Devitt in Bianchi (forthcoming).

grounding is in Nana 'the cat' and not in the temporal and spatial part of her" (1987: 65).[13] I here review their argumentation.

There will be no grounding if the sample of the perceptual experience does not correspond to the general categorial term which is used in conceptualization.[14] Thus concessions must be made. Causal theory of reference cannot be 'pure' causal. It has to be 'descriptive-causal' because the term is consciously or unconsciously tied with the description in grounding. Descriptive element has entered the designational chain. What is it that determines the nature of the sample? Is it the grounder's mental state? But Devitt and Sterelny admit that it is very hard to say what exactly determines this relevant nature.

The further claim is that this modification caused by the *qua* problem is only a modification of the causal theory of reference grounding while reference borrowing stays unchanged. Borrowers do not have to associate the right categorial term. Putnam's examples with 'elm' and 'beech' is harnessed to their support (Putnam 1975: 226–227). The example with whales also. What people centrally associated with whales was the description 'fish' and this is incorrect, but people nevertheless referred to whales.

Devitt and Sterelny offer what they call a hybrid theory.[15] It consists of: 1. Description theory of reference fixing and 2. Pure causal for reference borrowing. The move is from pure causal theory but the extent of the move should not be exaggerated because:

a. The associated general categorial term does not identify the object.
b. Modification is only in the grounding theory.

The reference borrowing remains unchanged and pure-causal: borrowers do not have to associate the correct categorial term.[16] They borrow their reference from others and are unlikely to have true beliefs about the underlying nature of the relevant kind but are also unlikely to have beliefs sufficient to identify its members. The causal theory lightens the epistemic burden. Thus, the borrower need not have any true beliefs, let alone knowledge, about the sense. The sense is largely external to the mind and beyond the ken of the ordinary speaker. What about other kinds terms? Devitt and Sterelny claim that we cannot borrow reference for other kinds terms. For example, for the term 'pencil' we need the description theory for reference fixing, the 'experts' who fix the reference must associate the appropriate description with the term even the rest of us need not. This then can be combined with a causal theory of reference borrowing explaining how the rest of us depend on

[13] This is the only quote I use from the first edition of *Language and Reality*. For more on the *qua* problem in the 2nd edition (1999: 79–81; 90–93; 98–99).

[14] See more about this in the discussion between Reimer and Devitt in Bianchi (forthcoming).

[15] In section 5.3. and 5.5. (pages 96 –101) of their 1999 book.

[16] For discussion on reference borrowing between Devitt and Jutronić see: Devitt (2006; 2008) and Jutronić (2006; 2008).

the experts. But the causal theory for borrowers here is supplemented with descriptions. A person could not use the term 'pencil' to refer to pencils if he was completely mistaken about them. Their discussion of hybrid theory is quite dense, with very few examples, so here is my attempt at a possible graphic presentation of their view.

PROPER NAMES (Blanka)

	descriptive	descriptive-causal	causal
grounding		+	(**qua-problem**)
borrowing			+

NATURAL KIND TERMS (gold)

	descriptive	descriptive-causal	causal
grounding		+	(**qua-problem**)
borrowing			+

OTHER KIND TERMS (pencil)

	descriptive	descriptive-causal	causal
grounding		+	
borrowing	**cannot be borrowed without description**		

What is important for the present discussion is the fact that there is a *qua* problem for proper names and natural kinds terms, arguably not for other kind terms. At the end of that section D&S say: "The *qua*-problem for our historical-causal theory gives ample motivation for us to look elsewhere for an explanation of how reference is ultimately fixed" (1999: 101).

The *qua* problem is discussed then in greater details in section 7 on 'Thought and meaning'. The assumption is that our cognitive capacities are closely correlated with our linguistic capacities. More specifically, the structure of mentalese is closely related to (public) language (1999: 145). The reference fixing of a linguistic word depends on the reference fixing of the mental word that it expresses, so a theory of the one carries over to the other (1999: 156).[17] In this section, they look into and consider pure-causal proposals of indicator and teleological theories. These theories have been developed as theories of the relationship between thought and the world. Devitt and Sterelny think that these theories are best construed as simply theories of ultimate reference fixing to which other theories could be added (1999: 157). They go into presenting criticism for the indicator theories (1999: 161) and suggest to go totally teleological, explaining representation by biological function alone since biological function is explained in terms of the history

[17] The same was assumed by Miller when he said: "that causal theories of perception and mental representation unavoidably hover in the background" (1992: 425). In other words, our ability to refer to things in language, and to create words that refer to things, depends on the prior ability to think and mentally refer. For this reason, discussions of mental reference and reference in language often go hand in hand.

of selection. Their original proposal is the following and I quote it in full:

> We are attracted by a less ambitious use of teleology to explain meaning. Instead of taking about biological functions to determine the contents of *thought* we take them to determine the contents of more or more basic representational states, *perceptions*. Perceiving a rabbit is a matter of being in a representational state with biological function of representing a rabbit. An interesting thing about this idea is that it does not *replace* the historical-causal theory of reference fixing, it *supplements* it. That theory…suffered from the *qua*-problem: In virtue of what is a particular grounding of 'rabbit' a grounding in rabbits rather than mammals, vertebrates or whatever? *The present idea offers a teleological answer: the grounding is in rabbits because it involves a perceptual state that has the function of representing rabbits. The teleological theory of perception becomes an essential part of the theory of groundings*…It incorporates teleology into the historical-causal theory of reference fixing. (1999: 162, italics mine)

This is a very important, promising and fruitful suggestion but since there are a very few examples given in their proposal, it is helpful to go back to the ideas elaborated a bit more in Sterelny 1990,[18] and also to see more details about the mechanisms of how this is possibly going to work as a solution of the *qua* problem in Neander (2017).

4.3 *Kim Sterelny (1990)*

Sterelny stresses that *qua* problem is the key unsolved problem for Kripkian causal theories, and this suggests that Kripkian story, to repeat, may not be a right story of primitive content, but rather plays a role in the explanations of more cognitively sophisticated structure whose content presupposes a conceptual backdrop. If the *qua* problem cannot be solved for non-basic concepts, could it be solved at least for the basic concepts?[19] Sensory concepts are likely candidates for basic concepts but they also pose the *qua* problem. Does my concept RED (when first acquired) name a color or a shade of that color, or even an intensity level of light? "Concepts for which the *qua* problem does not arise look decidedly thin on the ground" concludes Sterelny (1990: 118).[20] Nevertheless, it seems very attractive to add teleological ele-

[18] Devitt and Sterelny 1999 proposed solution of the *qua* problem actually relies much on Sterelny 1990 which was also presented in Devitt and Sterelny 1987. Sterelny's chapter 6 'Explaining Content' discusses different theories of content, concluding with the teleological view of perception (1990: 111–137).

[19] Sterelny talks here of concepts while our discussion is about terms. Nevertheless, one theory should be good for both. As Sterelny says: "Kripkian causal theories were originally developed as a semantic theory of language, but if they work at all they should work for the language of thought. The essential idea is that the content of a concept is determined by causal links between the individual acquiring that concept and its reference" (1990: 114). See also footnote 18.

[20] Stanford and Kitcher also express their doubts: "We should at least mention Devitt and Sterelny's interesting suggestion …that there may be primitive terms (categorials or simple demonstratives, say) which can be directly grounded in a

ments to the causal story since the appeal to the biological function of an internal representation is naturalistic, and it gives a more discriminatory machinery.[21]

In Sterelny's view from 1990, the semantic base consists of concepts that are formed from modular input systems. To go back to the example of color. The structure produced is not a shade of color or a particular intensity of light, although it is caused by some particular shade or some particular intensity. For the biological function of our color vision receptors is the representation of a stable and useful fact about our environment, namely the color of surfaces. For color vision, like many other modular processes, is serviced by constancy mechanisms. Perceptual processing works to keep track of invariances in the world, not the varying stimulations from it. Teleology then solves the *qua* problem, since the base-concepts are modular concepts. Above the base, the story stays much the same, but not quite the same. For example, Sterelny says, that Eric has the concept of tigers partly in virtue of his contact with tiger specimen and partly in virtue of his descriptive knowledge of tigers. Both are required for possession of the concept. Causal contact without any descriptive knowledge is not sufficient and with the descriptive element comes the *qua* problem. Sterelny thinks that with the introduction of the teleological dimension, the descriptive elements of 'tiger' possession in the modular system are not beliefs or intentional states, but the *Gestalt* of tigers. Unless the modularity hypothesis is completely wrong, there will be some course-grained purely perceptual representation of tigers. That representation, of course, has nothing like enough information in it to select the necessary and sufficient conditions of being a tiger. Some tigers will not fit. Something could fit it without being a tiger. The causal link with actual tigers is still necessary for possession of a tiger concept. The teleological dimension added then gives enough cognitive background for the rest of the machinery Devitt and he posited. Other descriptive-causal concepts, and fully defined concepts, can be acquired on these foundations.[22]

manner that avoids the *qua* problem. If so, perhaps the descriptions needed for reference-grounding will themselves reduce to primitive terms whose reference can be grounded without any descriptive component. *To our knowledge, however, noone has been able to make good on this suggestion, and we shall not pursue it here"* (2000: 127 note 6, italics mine).

[21] Sterelny says that "there will be an important teleological element in our total theory of mental representation, though any attempt to extend the teleological story to the human propositional attitude faces the most appalling difficulties" (1990: 138; see also Devitt and Sterelny 1999: 101).

[22] Sterelny states a possible objection to his proposal, i.e., that it has much in common with the traditional philosophical program called concept empiricism. The program took sensory concepts to be fundamental and given by our innate perceptual equipment. He then dissociates his view from concept empiricism: the properties modules represent are not sensory properties (our experience of the world) but objective features of the world that**:** a) were biologically important to our ancestors; b. are reasonably reliably detectable by an encapsulated special purpose mechanism.

Devitt and Sterelny's (1999) suggestion is, in my opinion, the most promising direction for the solution of the *qua* problem. Teleosemantics of *perceptual* content is where to look for the solution. Teleosemantics will yield a perceptual content that can be the basis for explaining in virtue of what the grounder had tiger and not mammal, or part of tiger in mind when he grounded the term 'tiger'. Recently their suggestion seems even more plausible with the fine elaboration of the teleosemantic explanation of the preconceptual/nonconceptual level of sensory perceptual representations found in Neander (2017).

4.4 *Karen Neader (2017)*

All I want to do in this section is to state some of the most important questions and aims that Neader (2017) makes in her new book *A Mark of the Mental.*[23]

The main questions are: Do the mental representations with original intentionality derive it from nonintentional nature and, if so, how? If intentionality is not a fundamental feature of the universe, what is it more fundamentally? What is its ontological grounding? On which nonintentional facts and properties of the world does it depend, constitutively? (2017: 9). Some of the main aims are: to encourage optimism with regard to the naturalization project and also to encourage those who support teleosemantics to look into a causal-informational version of it (2017: 3).

Neander defends a theory of mental content that blends elements of a teleosemantic approach with elements from a causal theory of reference and a version of a (similarity-based) state-space semantics (2017: 22). She has long developed and defended an *etiological theory* whose gist is that the (or a) function of an item (if it has one) is what it was selected to do (2017: 39). The only thing that all teleosemantic theories have in common is the claim that semantic norms, at their most fundamental, supervene somehow on functional norms, among other things.[24]

The guiding intuition for sensory-perceptual representations is that their contents are not what causes them to be produced but what is "supposed" to cause them, in the teleonomic sense. Their contents are what the systems that produce them have the function to detect by producing them. Her argument says that sensory-perceptual representation refers to what is *supposed* to cause it. (italics mine)

What concerns us here most is her argumentation for content determination. A content-determinacy challenge asks of a given representation to explain why it counts as having the content it has rather than some other content (2017: 150). Why does RED have the content *there's*

In short, he thinks that we need *conceptual foundationalism* without definitions in which we give a teleological account of the content of base concepts.

[23] See her helpful interview on the web, February 15th, 2018.

[24] We should, Neander argues, return to something much like Stampe's (1977) starting proposal. His idea was that appealing to functions is a promising way to improve a causal theory of reference.

red (and not, say, *there's color* or *there's a fire truck*)? In the parlance of *qua* problem: Why the grounder names the tiger and not the mammal, or part of the tiger?

Neander proposes what she calls Simple starter theory (2017: 149). Simple starter theory is based on causal theory (CT) which says: A sensory-perceptual representation, R, which is an (R-type) event in a sensory-perceptual system (S), has the content *there's C* if and only if S has the function to produce R-type events in response to C-type events (in virtue of their C-ness). The simple causal version of teleosemantics entails that, for example, the frog's perceptual representation can have the content *there's something small, dark, and moving*, and not *there's a fly* or *there's frog food*. It tries to solve the question of how content is determined.

How does it do it? Here the question of mechanisms come into view. *A sensory-perceptual system has sensory receptors, which are cells or other units adapted for transducing energy from the environment into a medium that a cognitive system uses for information processing. Thus, importantly, if there are two dispositions they call for two different mechanisms* (2017: 169, italics mine).[25]

Neander mentions Sterelny (1990) and his question: Why does a sensory-perceptual representation (R) refer to C and not to Q when Q is a proximal (intermediate) link in a C-to-R causal chain? (2017: 222). And her answer is that, R refers to C rather than the more proximal Q if the system responsible for producing Rs was adapted for responding to Qs (qua Qs) by producing Rs as a means of responding to Cs (qua Cs) by producing Rs, but it was not adapted for responding to Cs as a means to responding to Qs. (so it is not the shade of color red but color red in the example given by Sterelny 1990). In sum, the simple causal-informational version of teleosemantics, CT, says that a sensory-perceptual representation refers to the environmental feature it is the function of the system to detect by producing the representation. But Neander warns us, its scope is restricted to nonconceptual sensory-perceptual representations.

If the causal-informational version of teleosemantics offered by Neader delivers sufficiently determinate contents for nonconceptual sensory-perceptual representations, then Devitt and Sterelny's suggestion to look for the answer of the *qua* problem in this direction is a promising line that might lead us from preconceptual to conceptual. Rather than introducing a descriptive element into that content there is hope (and now more than hope) that teleosemantics will yield a perceptual content that can be the basis for explaining in virtue of what the grounder had, for example, Mick Jagger and not his lips, in mind when he grounded 'Jagger'.

How we can get from nonconceptual to conceptual content? Neander says: "What is left is the ramping-up problem, which is the problem of understanding how to get from a theory of content for nonconceptual

[25] Note the similarity with Maddy and Ryder.

representations to a theory of the referential power of sophisticated human thought" (2017: 26). Neander gives us hints since she (I think rightly) believes that the distinction between conceptual and nonconceptual representations is not that sharp. One of the suggestions is that the mind, for example, abstracts or subtracts from the specific features of specific triangles to form an abstract idea of triangularity (2017: 206).[26] Or by averaging the shapes of category members. "We could likely produce recognizable results by averaging the shapes of diverse cows, diverse cats, diverse carrots, diverse cars, and so on. These categories are counted as 'basic' categories in part for this reason, and they are apparently learned more easily than other categories" (2017: 210).

5. *What are the mechanisms of reference?*

How much should a philosopher worry about mechanisms, in this particular case, mechanism(s) of reference? Where shall we look for an answer. Turning to the mechanism we are admitting, in Devitt's words, that we cannot find the answer within philosophy but the answer might be given by psychology or psycholinguistics? Looking into mechanism of reference we seem to be leaving the philosophical ground. However, if the psychological mechanisms point to the solution of the *qua* problem can we say that we have the solution which is in a way indirectly solution to the metaphysical, i.e., philosophical question. If a philosopher who is a naturalist closely relates his answers to science, then scientific answers are very relevant to his philosophical questions and solutions.

It is worth looking at bit more into the relation between metaphysical (philosophical) questions, semantic dispositions and mechanisms behind them. At which point can we say that the *qua* problem stops being a philosophical problem? One thing to notice is that when you look up the entry on reference in *Stanford Encyclopedia* online, all the talk is about mechanisms. Here are just a few passages (italics are mine):

> The central issues, the central questions, concerning reference are four: (i) What is the *mechanism* of reference? In other words, in virtue of what does a word (of the referring sort) attach to a particular object/individual?
> Assuming that at least certain sorts of terms do in fact refer, the central question regarding linguistic reference becomes: how do such terms refer? What, in other words, is the '*mechanism*' of reference?
> This suggests that names are semantically different from descriptions, which in turn suggests that *the mechanism* by which a name refers cannot be identified with some definite description. (Michaelson and Reimer 2019)

Wettstein (2004) says that the phrase 'the mechanism of reference' originates with McGinn (1981). McGinn says: "Reference is what relates words to the world of objects on whose condition the truth of sentences hinges. It is natural to wonder what sorts of relations underlie the reference relation, to wonder, that is, *what constitutes the mechanism of reference*" (1981: 157, italics mine). McGinn seems to closely

[26] Maddy's idea about triangles quoted in section 3a is a similar idea.

relate answering the question about the mechanism of reference to answering one of the foundational philosophical question.

In the book *Reference and referring* Pepp stresses: "…reference is often thought of as the bond between language and the world, or between language and the aspects of the world that language is used to talk about. Referring is often thought of as the activity in virtue of which that bond holds. A distinctive question about reference and referring concerns what makes this bond hold, *or what the nature* of this activity is: what is the *mechanism* by which language is tied to the particular things that are its subjects? *I will call this the 'mechanism question' about reference*" (2006: 1, italics mine). Here again, let us notice, the question: *what is the nature of referential relation (bond) is identified with the question what the mechanism of this bond is*?

How is metaphysical question *what is to name* related *in virtue of what* question? Descriptive theories of reference, to my knowledge, have not been referred to as mechanism of reference. They were replaced by the causal historical theory of reference. The relation between name and referent was reduced to a causal chain. Kripke called it a "better picture" but we can see this label as a metaphorical expression for a new kind of mechanism of reference. Devitt would surely agree that Kripke was giving an important and crucial philosophical contribution to the theory of reference but can still insist/claim that the explanation of the mechanism of reference cannot be given by philosophers. Miller in his article trying to support the pure causal theory of reference grounding says: "… I also trust it will not seem like handwaving for the philosopher to say that a detailed account of the actual causal mechanism of perceptual constancy is a job for the experimental psychologist" (1992: 431). Miller, like Devitt, is actually saying it is not the philosopher's task to give an account of the mechanisms. Or is it an even stronger claim that philosopher is in no position to give such an account?

More generally, beyond the *qua* problem and its solution, one can ask where does a philosophical question stop being philosophical? Neander (2017) distinguishes why questions and how questions and says that why questions ask about the origin, presence and persistance of something while how questions ask about how systems operate. She finds this distinction in Mayr (1961: 1502) who drew a distinction between two main branches of biology that he called 'evolutionary' and 'functional.' The evolutionary biologist is concerned with why-questions, whereas the functional biologist "is vitally concerned with the operation and interaction of structural elements, from molecules up to organs and whole individuals." (2017: 48). Now Neander says that those whom Mayr calls 'functional biologists' are those whom she here calls 'physiologists and neurophysiologists.' She does not say where her own teleosemantic theory, or teleosemantic theories in general, belong. Do they answer why questions or how questions? Where does Neander see herself, as a philosopher or a scientist or something in between?

Obviously as a philosopher but a great deal of her discussion is the discussion of *how* questions since one of her main goals is to solve the *mechanisms* of nonconceptual/preconceptual content.

As we saw Maddy (1984) is giving the answers to the *qua* problem in neurological terminology. "Neural connections between perceptual assemblies for samples and perceptual assemblies for word types are "wired in" (Maddy 1984: 474) and she points out: "But if the point at issue is whether or not our reference is determinate, all that is needed is an account of how this is possible, and there is no reason such an account need be conceptual rather than scientific (1984: 468).

More recently Ryder (2004) is talking about learning mechanism as brain mechanism in which each dendrite is adjusting so as to bring that dendrite's contribution closer to that of each of the other dendrites that contribute to the firing of the cell in question in order to yield reliable 'predictions'. The answer is given by science. It is a core assumption in cognitive science that cognitive processes involve *formal* operations on structured representations. That is to say that these operations are conceived as causally sensitive to the physical, chemical, or neurophysiological properties of the representational vehicles rather than their semantic properties.

Going back to Neander, she says that her book is 'ambitious' because it tries to make genuine progress in relation to one of the most difficult problems in philosophy of mind—that of understanding the fundamental nature of intentionality (2017: 243). As was pointed out, her argument relies on claims concerning explanatory concepts and practices in the mind and brain sciences.[27] She says: "Informational teleosemantics is supported by the explanations of cognition that the mind and brain sciences currently provide" (2017: 74). Teleosemantics is based on what "the mainstream branches of the sciences devoted to explaining cognitive capacities ascribe normal-proper functions to cognitive mechanisms and assume that these include functions to process information. It makes excellent sense to try to understand how far these information-processing functions can take us in understanding the nature of mental content" (2017: 96). Her philosophical argumentation is based on scientific theories and she believes the two cannot/should not fall apart. "Whichever approach is adopted, the science and the philosophy cannot be divorced if the content ascriptions a philosophical theory of content generates are to be relevant to explaining cognition" (2017: 96).

A naturalistic theory of intentionality is one that explains intentionality using the resources available from the natural sciences. From the standpoint of philosophers that are naturalists, semantic naturalism is committed to the idea that the relevant kind of theory of intentionality ought to be reductive and construed in terms of some natural science.

[27] See specially section (4) on the *Methodological Argument for Informational Teleosemantics.*

The vocabulary is that of the natural sciences, and in biosemantics this means that it is the vocabulary of biology. Given that functionalism is commonly based on a physicalistic ontology, the mental states that are supposed to be a part of "causal pushes and pulls inside the head" are proclaimed to be physical states, more specifically, neural states. Thus, mental concepts apply to neural states of the brain.

If we accept the above, then it is plausible to talk about levels of explanations of particular referential bonds, starting maybe from common sense, through different kinds of philosophical causal theories on one hand, and neurological "hard-science" explanation on the other. Where philosophy stops and science begins is not easy to say. They are, from naturalistic point of view, continuous. Different philosophers draw different lines between the two. For example, Lycan says: "Remember also that the principles of psychosemantics itself are philosophy, not science. And they remain unsettled to say the least" (in his 2006 talk). He would probably not agree with Devitt when he says "these (referential) mechanism seem to me to be psychological matters, not philosophical ones". On the other hand, if the ultimate answers are expected to be given by science, in this case brain sciences, and this is probably what Devitt had in mind when he decided to stop worrying about *in virtue of what* question.

6. *Conclusion*

1. There is no pure causal theory of grounding, in spite of the discussed attempts to show that reference grounding is a causal process. It is clear that descriptions play a role in fixing or grounding the reference. And the given attempts to solve the *qua* problem in purely causal term fail. Stanford and Kitcher in examing what they call "Simple Real Essence Theory" (SRT) say: "As Michael Devitt and Kim Sterelny point out, a theory like SRT is too simple…because it is *utterly mysterious*, *how without something more than our causal relation to the sample*, we can pick out one, rather than another, of the many kinds the sample instantiate" (2000: 100–1, italics mine).

2. As far back as in 1981 Devitt argued that a causal-historical theory can be naturalized if it is articulated in terms of causal relations of the right kind, although it will then still be incomplete. In other words, it will lack a solution to the *qua* problem. I found Devitt and Sterelny's suggestion to incorporate teleology into the historical-causal theory of reference fixing (1999: 162) a very promising idea. The idea is straightforward: If mental states or semantic properties as not fundamental, any appeal to them in an analysis of the reference relation must eventually be accounted for in other terms and teleosemantics seems to ground them.

3. If naturalism is an approach to philosophy that involves using sci-

ence, ultimately physics, as our guide to the fundamental ontology of the universe the solution the *qua* problem is found by those who do neuroscience and brain neurology. We saw an earlier attempt by Maddy (1983) and more recent one by Ryder (2004) who also looks for an answer in brain sciences. Since the cerebral cortex is the seat of the mind, Ryder argues that SINBAD representation realizes mental representation with a cerebral cortex. Usher (2004) states that the merit of the SINBAD model is to provide an explicit *mechanism* showing how the cortex may come to develop detectors responding to correlated properties and therefore corresponding to the sources of these correlations. Such and similar attempts offer hope of naturalistic explanation of reference, i.e., in bringing semantic relations within the scope of physicalist view of the world. The real explanatory work is done by science but the work is far from been done as Stanford and Kitcher, discussing the natural kind terms, point out and say: "sadly, the course of reference fixing in actual scientific cases is even more complex than (our) analysis shows" (2000: 114, italics mine).

4. Neander argues that the naturalistic theories on which most work has been done of late are the teleosemantic theories. According to such an analysis (Neander 1991) items of a type have the function of doing what that type of item was selected for doing. Teleosemantics will yield a perceptual content that can be the basis for explaining in virtue of what the grounder had Mick not his lips in mind when he grounded 'Jagger'. Neander's detailed analysis of preconceptual content gives great hope that conceptual can be developed from this more basic content and gives us explanation how the *qua* problem can be solved. But as Neander reminds us "we should also keep in mind that serious work on naturalistic theories of content has only been going on for decades rather than centuries and that, on a philosophical timescale, that is quite a short time (in *Stanford Encyclopedia*)". Or as Devitt said in Maribor: "Rome was not built in a day."

References

Borstner, B. and Todorović, T. (eds.). Forthcoming. *The Many Faces of the Philosophy of Michael Devitt*. De Gruyter Publishers.

Bianchi. A. (ed.) 2015. *On Reference*. Oxford: Oxford University Press.

_____, Forthcoming. *Language and Reality from a Naturalistic Perspective: Themes from Michael Devitt*. Cham: Springer.

Davidson, D. and Harman, G. (eds.) 1972. *Semantics of Natural Language*. Dordrecht: Reidel.

Devitt, M. 1981. *Designation*. New York: Columbia University Press.

_____, 2006. "Responses to the Rijeka Papers." *Croatian Journal of Philosophy* 6: 97–112.

_____, 2008. "Reference Borrowing: A Response to Dunja Jutronić." *Croatian Journal of Philosophy* 8: 361–6.

_____, 2015. "Should Proper Names Still Seem So Problematic?" In A. Bian-

chi (ed.) 2015. *On Reference*. Oxford: Oxford University Press: 108–143.

_____, Forthcoming. "Stirring the Possum: Responses to the Bianchi Papers." In: Bianchi Andrea. (ed.) forthcoming.

_____, Forthcoming. "Reply to Reimer." In A. Bianchi (ed.). *Language and Reality from a Naturalistic Perspective: Themes from Michael Devitt*. Cham: Springer.

Devitt, M. and Sterelny, K. (1999). *Language and Reality*. 2nd edition. Oxford: Basil Blackwell. The first edition appeared in 1987.

Devitt, M. and Hanley, R. (eds.) 2006. *The Blackwell Guide to the Philosophy of Language*. Oxford: Blackwell Publishing.

Donnellan, K. 1972. "Proper Names and Identifyling Descriptions." In D. Davidson and G. Harman (eds.). *Semantics of Natural Language*. Dordrecht: Reidel: 256–79.

Douglas, S. P. 2018. "The *Qua*-Problem and Meaning Scepticism." *Linguistic and Philosophical Investigations* 17: 71–78.

Frege, G. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. Jena: Verlag Hermann Pohle, Band I.

French, A. P. et al. (eds.). 1977. *Midwest Studies in Philosophy: Studies in the Philosophy of Language. vol. 2*. Minneapolis: University of Minnesota Press.

Jutronić. D. 2000. "Knowledge and Reference, or the Qua-Problem Revisited." *Synthesis Philosophica* 15 (1–2): 189–209.

_____, 2003. "Some Problems with Reference Borrowing." In B. Berčić and N. Smokrović (eds.). *Proceedings of Rijeka Conference: Knowledge, Existence and Action*, Rijeka: Hrvatsko društvo za analitičku filozofiju and Filozofski fakultet Rijeka: 12–16.

_____, 2006. "Is Reference Borrowing a Causal Process?" *Croatian Journal of Philosophy* 6: 41–9.

_____, 2008. "Reference Borrowing and the Role of Descriptions." *Croatian Journal of Philosophy* 8: 349–60.

_____, Forthcoming. "Attempts at Solving the *Qua*-Problem." In B. Borstner and T. Todorović (eds.) *The Many Faces of the Philosophy of Michael Devitt*. De Gruyter Publishers.

Kabasenche P. et al. (eds.). 2012. *Reference and Referring*. Cambridge: The MIT Press.

Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Lycan, W. 2006. "Consumer Semantics to the Rescue." Paper presented in a symposium in honor of Distinguished Woman Philosopher Award recipient Ruth Garrett Millikan, Society of Women Philosophers (December 2006).

Maddy, P. (1984). "How the Causal Theorist Follows a Rule." *Midwest Studies in Philosophy* 9 (1): 457–477.

Mayr, E. 1961. "Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist." *Science* 134 (3489): 1501–1506.

McGinn, C. 1981. "The Mechanism of Reference." *Synthese* 49:157–86.

Michaelson, E. and Reimer, M. 2019. "Reference." *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition). E. N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/spr2019/entries/reference/>.

Miller, R. 1992. "A Purely Causal Solution to One of the Qua Problems." *Australasian Journal of Philosophy* 70 (4): 425–434.

Neander, K. 2012. "Teleological Theories of Mental Content." *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition). E. N. Zalta (ed.). URL = <https://plato.stanford.edu/archives/spr2018/entries/content-teleological/>.

_____, 2017. *A Mark of the Mental*. Oxford: The MIT Press.

Pepp, J. 2012. "Reference and Referring." In P. Kabasenche et al. (eds.). 2012: 1–33.

Putnam, H. 1975. *Mind, Language and Reality: Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.

Reimer, M. Forthcoming. "The qua-problem for names (dismissed)." In A. Bianchi (ed.) forthcoming.

Rupert, R D. 2008. "Causal Theories of Mental Content." *Philosophy Compass* 3: 353–80.

Russell, B. 1905. "On Denoting." *Mind* 14: 479–493. Reprinted in *Logic and Knowledge*. Marsh, R. C. 1956. (ed.). London: George Allen and Unwin.

Ryder, D. 2004. "SINBAD Neurosemantics: A Theory of Mental Representation." *Mind and Language* 19: 211–40.

Sauchelli, A. 2013. "Ontology, Reference, and the Qua Problem: Amie Thomasson on Existence." *Axiomathes* 23 (3): 543–550.

Simchen, O. 2012. "Necessity in Reference." In W. P. Kabasenche et al. (eds.). 2012: 209–235.

Stampe, D. 1977. "Toward a causal theory of linguistic representation." In P. A. French et al. (eds.). 1977: 81–102.

Stanford, P. K. and Kitcher, P. 2000. "Refining the causal theory of reference for natural kind terms." *Philosophical Studies* 97: 99–129.

Sterelny, K. 1983. "Natural Kinds Terms." *Pacific Philosophical Quarterly* 64: 100–125.

_____, 1990. *The Representational Theory of Mind*. Oxford: Blackwell.

Usher, M. 2004. "Comment on Ryder's SINBAD Neurosemantics: Is Teleofunction Isomorphism the Way to Understand Representations?" *Mind and Language* 19 (2): 241–248.

Wettstein, H. 2004. *The Magic Prism*: *An Essay in the Philosophy of Language*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Expressions and their Articulations and Applications

UNA STOJNIĆ
Princeton University, Princeton, USA
ERNIE LEPORE
Rutgers University, New Brunswick, USA

The discussion that follows rehearses some familiar arguments and replies from the Kripke/Putnam/Burge critique of the traditional Frege/Russell/Wittgenstein views on names and predicates. Its main contributions are, first, to introduce a novel way of individuating tokens of the same expression, (what we call "articulations") second, to then revise standard views on deference, (as this notion is understood to pertain to securing access to meaning for potentially ignorant, and confused agents in the externalist tradition going back to Putnam and Burge) and lastly, to emphasize the often conflated distinction between disambiguation and meaning fixing. Our line on deference is that it is not, and should not be conceived as, an intentional mental act, but rather indicates an historical chain of antecedent tokenings of the same expression.

Keywords: De facto deference, articulation, network, Kripke, Putnam, Burge, Dummett, Evans.

Introduction

The discussion that follows is largely extracted from two chapters of a book we are currently writing. Other than rehearsing some familiar arguments and replies from the Kripke/Putnam/Burge critique of the traditional Frege/Russell/Wittgenstein views on names and predicates, its main contributions are, first, to introduce a novel way of individuating tokens of the same expression, (what we call "articulations") second, to then revise standard views on deference, (as this notion is understood to pertain to securing access to meaning for potentially ignorant, and confused agents in the externalist tradition going back to Putnam and Burge) and lastly, to emphasize the often conflated distinction between disambiguation and meaning fixing. Our line on deference is that it

is *not*, and should not be conceived as, an intentional mental act, but rather indicates an historical chain of antecedent tokenings of the same expression. What any of these claims and distinctions amount to should be clear in what follows.

1. *Names, articulations and naming*

Suppose you pick up a name in a casual conversation, say, simply by hearing a group of interlocutors using it. Your interlocutors may have been using the name for a particular individual for some time, but for you it is novel. Should you opt to use the name in order to try to name the same individual as your interlocutors, it might be that whatever success you achieve with your use of the name piggybacks on whatever success your interlocutors had with their uses of the same name. That is to say, your success seems predicated on your deferring to the speakers who exposed you to the name.

Here is a simple illustration of how easy it is to pick up a name:

> A says: Napoleon was a famous military leader.
> B asks: Was Napoleon born in the 15th century?
> A replies: No! He was not!

B's success in naming Napoleon is predicated on deference to A. B has never been exposed to Napoleon's name before.

It might turn out that your interlocutors' own success also relies on deference to whomever they picked up the name from, and so on and so on through a network of users extending all the way back to an introduction of the name, where, we might presume, a connection between the name and whichever individual it names was somehow first forged. Put differently, by virtue of your deference to whomever first exposed you to the name, you thereby enter into a *network* of users, all tied together by deference to individuals who first exposed them to the name—a network that stretches all the way back to the name's introduction.

Of course, everything we've said so far about the establishment of, and successful inclusion in, a network of interlocutors leaves completely open how exactly (or even whether) an individual came to be the bearer of that name in the first place, that is, everything we've said so far leaves open the philosophical question of what, if anything, is "the semantic glue to stick our words onto their referents" (Lewis 1984: 221). That is obviously an interesting and important philosophical question, and it is one that has occupied the dogged attention of generations of meta-semanticists, but we don't know its answer and, for our present purposes, we don't have to. And so, it will *not* be our focus here. Instead, ours will be on the network itself, and what its existence suggests about the constitution of successful uses of a name, and in general, of language. This investigation requires answers to (at least) two questions:

1. What must a speaker know or do in order to successfully token/articulate a particular name on an occasion of use?
2. What must a speaker know or do in order to successfully apply that use of that name to a particular individual?

We note that, while we begin our discussion focusing on proper names, we are ultimately interested in questions as they apply to expressions generally: what does it take for a speaker to successfully token any expression, and what does it take for this token to have a successful application? We will turn to these questions in sections 3 and 4 below.

With respect to both questions (1) and (2), it should be obvious (though it is not clear that early proponents of deference acknowledged or focused on *both* features of linguistic usage) that successful tokening of a name and naming can occur even in the face of widespread error about, and ignorance of, not only what the name names but also the name itself. For example, a proper name may, and indeed, surely often is, likely to admit of many different *sorts of articulations*, both statically and dynamically. After all, it can be written, typed, spoken or signed, *inter alii*. And in any one of these media, there invariably is a high degree of flexibility for how it can be tokened; e.g., in how it can be spelled or pronounced.[1] Further, it may change its canonical spelling or pronunciation across time or place. And, of course, at any given time, it might even misspelled or mispronounced according to whatever standards are in place—and yet it might still be tokened (Hawthorne and Lepore 2011). What, we may ask, can possibly hold all these tokenings together as articulations of the *same* proper name?

This question has received very little careful attention in the literature. Perhaps, many contributors thought its answer was obvious. For example, there is very likely, in normal circumstances, a fact of the matter about which expression (that is, in the cases under discussion, which proper name) the speaker *intends* to use. Many may have thought this intention, by itself, can determine which name is being tokened.[2] But, of course, this depends on the intention. Suppose the speaker intends to use, with a particular articulation, that name the speaker picked up in a conversation or in a reading. Then, can we conclude that the speaker is using *that very name*? This view has the advantage that, regardless of how much off the mark, or however idiosyncratic, the speaker's tokening may (turn out to) be (perhaps, some

[1] We will return below to the question of how much tolerance is permissible before a loss of identity.

[2] See, for instance, Kaplan (1991; 2011). Kaplan (2011) qualifies the intention view somewhat: there is a certain standard of performance an utterance has to satisfy in order to count as a (even bad) performance, rather than non-performance. For instance, simply grunting might not qualify as uttering a word. However, provided such a standard is satisfied, intention suffices to determine the identity. See also Hawthorne and Lepore (2011) for a critical discussion. Hawthorne and Lepore also advocate for a standard that separates performances (even bad ones) from non-performances.

would have it, as we shall discuss below, provided that it meets some contextual standard for counting as a performance), or even how confused the speaker's concomitant beliefs may be about which expression is being tokened, the speaker can still succeed in using a particular name.

But we have to be careful here. We do not mean to suggest that each time someone speaks, they have to *explicitly* form an intention to use the name they picked up from A (where A is the individual who introduced the speaker to the name). Rather, our view is that, somehow or other (in ways that perhaps even psycholinguists don't fully understand yet), a speaker *selects* a name from her (mental) lexicon.[3] Of course, there is a fact of the matter about who introduced this speaker to the name in her lexicon that she is selecting; the name selected is identical to the name in the mouth—or more precisely, representation—of the agent who first introduced the speaker to it. In selecting the term, the speaker is, in a sense, deferring to the agent who introduced her to the term. But notice, in this way, deference should not be understood as the term is typically used, namely, as an active intentional mental act, but it is rather *de facto*—in effect, something that largely passively happens to a speaker. Therefore, someone might be mistaken, in the sense in which we are using the term, about to whom they are deferring in virtue of their being mistaken about who introduced them to the name that they selected. Nevertheless, no matter what, there is still an *historical* fact about who introduced the name into the speaker's (mental) lexicon—who they got the name from. So, in order to perform any utterance, a speaker has only to choose a particular linguistic form, one which features a representation of a certain name; given this, there is a fact of the matter about which name figures in the speaker's representation of an utterance: it is whichever name the speaker selected from her lexicon, which is that name that featured in the representation of whoever introduced the lexical item into the speaker's mental lexicon in the first place (and so forth back to the initial introduction of the name). "Deference" in this sense is *not* an intentional act by the speaker to token whichever expression the individual who introduced the speaker to the expression tokened. Rather, the speaker intends to token some particular expression in his (mental) lexicon (but there is a fact of the matter about which tokened expression introduced that lexical item into the speaker's (mental) lexicon in the first place).

A different sort of worry arises when there is a departure from an accepted conventional norm for the articulation of some expression. The greater the degree of departure, the more likely it is that confusion will ensue. The further off a spelling or pronunciation is from some accepted standard, the less likely the hearer will be able to recognize which name is being articulated; and then, there is also the worry that, because *different* names can share a single articulation, a hearer might

[3] For a further, more detailed, development of this point, see Stojnić, ms.

mistake which name the speaker is tokening. How many individuals do we all know whose name is typeset as "John"? *But* there is an historical chain of tokenings ending with the name's introduction[4] that determines which name is actually being tokened on any given occasion of use. The individuation is not a matter of any particular articulatory shape (*contra* Davidson (1979: 90); cf. Hawthorne and Lepore (2011)); the bond between a name and its sundry tokenings is secured through a community wide network of deference to others about which particular name is being tokened—despite whatever wide-spread error and ignorance surrounds any given usage.[5]

The *epistemological* worries concerning how we decide which name an articulation is of, we believe, have boundless (defeasible) solutions. For example, if we are talking about the butcher shop, and not the produce stand, then, most likely, the "John" we mean (that is, the name we articulated with "John") is the butcher's name and not the gardener's, even if both names are articulated with the same pronunciation and with the same spelling. This is much like how we go about "disambiguating" uses of "bank"; if the speaker is walking along the river when uttering "bank", we are likely to resolve one way, but if the speaker is talking about depositing money when uttering "bank", we resolve to a different expression—the same articulation, but different words.

There probably is no end to how many strategies we might employ in going about making these sorts of decisions, even though there is a fact of the matter about which decision is correct, and the potential for error always exists. This means that a speaker can mispronounce or misarticulate a name, while still tokening it, but at the same time the audience can be mistaken in "disambiguating" a name: they might be misled by the evidence available, taking speakers to have tokened one name, when in fact they were tokening another. Such epistemological considerations belong to the theory of disambiguation, not the theory of meaning (determination), in as much as they delineate the set of cues language users use to recognize a particular form as the one that has

[4] In the literature, there is invariably talk of a speaker *intending* to use a name with the same reference as the person's uses of the name from whom the speaker learned the name. (See, e.g., Kripke (1980), *inter alia*.) We chose to switch over to *de facto* deference talk instead, since (a) requiring the speaker to have such explicit intentions, we believe, is requiring too much, and (b) we believe, as already noted above, that *de facto* deference talk can be cashed out independently of intention talk (especially, if intentions are understood as beefy propositional attitudes). For more on this point, see Stojnić, ms; Stojnić and Lepore, ms.

[5] Recall, again, that deference isn't here understood as an "intention to defer", as a plan to token a certain symbol. Even if someone doesn't *intend* to defer to X, who introduced them to the term, the tokened symbol will mean whatever it meant in the mouth of whom it was acquired from (and so on)—the *symbol* will be *de facto* deferentially individuated. So, deference is *de facto*, not deference by plan or intention. The speaker simply has to select the expression from her lexicon; the individuation, and meaning, of the expression is determined by *de facto* deference to whomever introduced the speaker to the term.

been uttered, and not to determine how to interpret its meaning (cf. Stojnić, Stone, and Lepore, 2013; 2017).

In short, the take-home message so far is that it is all too easy, even at the stage of name identification, to conflate epistemology and metaphysics, disambiguation and meaning determination, and so, vigilance is required in respecting relevant distinctions—in this case, the distinction between an expression and its articulations. To repeat, though many individuals have names typeset as "John," on any given occasion of use there is a fact of the matter about which one of these different names is being tokened by an instance of this shared articulation.[6] And this fact is *fixed* by a speaker's tokening a particular expression, which is individuated by *de facto* deference. In this regard, we reject customary talk of numerous individuals bearing the same name, as in: "Proper names typically have more than one bearer. Thus, a contemporary token of 'Aristotle' might designate the famous philosopher or it might designate the late shipping magnate Onassis" (Devitt 2015: 110). We think not. There are (at least) two names "Aristotle". "Aristotle" is ambiguous, if you like.

This is not to deny that the audience may face hurdles, perhaps, for all intents and purposes, insurmountable ones—ones that inspire requests for elaboration and assistance—in identifying which name is being tokened. Nor it is even to deny that speakers might be confused in all sorts of ways about which name they are tokening. (For instance, they might erroneously believe that the name they are tokening is identical to the name they learned from A, when, in fact, it isn't.) This doesn't prevent them from either tokening the name, or applying it.

But, no matter how muddy the epistemology becomes, the metaphysics remains clear. The name being articulated, on any given occasion of use, is determined by *de facto* deference of the speaker to the name acquired first from some other speaker. And so, the answer to our first question (which one of us defends further in detail elsewhere (Stojnić, ms)) about what speakers must know or do in order to successfully token a particular name on an occasion of use is that they needn't *know* anything; rather, they must *do* something—namely, token a particular expression (select that expression) in their mental lexicon, which, in turn, defers (*de facto*) to the tokening that first exposed the speaker to the name; and so on and so on back to its neologism.

We are now ready to turn to our second question about what speakers must know or do in order to successfully *apply* a name, and, not surprisingly, we find that many of the warnings we had to heed about misarticulating a name have their echoes in a speaker's ignorance of, and errors in, applying that name. So, suppose that the speaker believes a name picks out a butcher, when, in fact, it picks out a gardener.

[6] This particular specification of our view assumes that names are not predicates (or generic names, in the sense of Kaplan (1990)). But even if it turns out that they are, what we have to say about predicates below suffices to establish our point about *de facto* deference all over again.

These sorts of error can persist, and may even be pervasive, and yet present no obstacles to successful naming. (If you believe, falsely, that "John" picks out a butcher, you will have (successfully) said something false when you say "John's a butcher".) How is that possible?

Once initially determined (however that is achieved[7]), successful naming can obtain even in the face of confusion and widespread error, again, both about the name's identity and its meaning. A use of the proper name "Aristotle" names whomever it names, regardless of any mistaken beliefs or other misinformation interlocutors carry into a conversation where this name is being used. This is because the network of *de facto* deferential speakers, "stretching back from our uses to the first uses of the name to designate Aristotle" (Kripke 1980: 25), secures this same naming for current users of the same name (where the fact that the same name is being tokened is itself secured through a network of *de facto* deference, as explained earlier).[8] Once it is settled that a speaker is using the name "Aristotle", and that this name names a particular individual, then the speaker's use also names that individual, regardless of how confused or ignorant the speaker is about which name is being used and whom or what it names.

All that matters for achieving these results is that someone exposed the speaker to that name, and it names some individual (through their own network of deference to whomever they picked up the word from); more precisely, and keeping in mind our answer to the first question, all that matters is that the speaker *is tokening* the name "Aristotle". So long as the speaker selects the name "Aristotle" from her mental lexicon in forming the utterance (where the identity of the expression is determined through *de facto* deference), and thereby, tokens the name "Aristotle", then they name whomever the name "Aristotle" names (if anyone) in the network of *de facto* deference that the speaker is participating in.[9] In this regard, the application of the name is fixed once the name itself is created.

[7] Again, we do not care if it's as a matter of a causal covariance, or a Fregean sense, or however a name's meaning is established. We are interested only in what is required of the speaker to count as a user of an expression, and not what is required of an expression to have meaning.

[8] Again, we are sidestepping important philosophical issues, because we *can*, given our purposes and aims about what, if anything, must be in the head of the neologizer of the name. Our interest is in the other members of the network, so to speak, and what, if anything, they must know or be connected to in order to successfully token, and successfully apply a name. We care at present only about how meaning can be exploited by a novice once created.

[9] In this regard, we are disagreeing with Kripke that a speaker when he uses a name "must ... intend when he learns it to use it with the same reference" (1980: 96). If we are right, the intention to use a term (assuming the use is not one whereby the expression is introduced and its meaning fixed) with a particular reference is relevant for the identity of the expression uttered, as well as the meaning that expression has on the occasion of use.

We are ready to move on to how we intend to expand the network model to other sorts of expressions.

2. *Challenges*

In Part 1, we tendered answers to two central questions about names and naming:

1. What does it take to successfully token/articulate a name?
2. What does it take to successfully use/apply a name, to name something?

We have spent some time defending a particular answer to our first question. According to us, speakers needn't know *anything* in order to token an expression. They need to *do* something: they need to select an item in their (mental) lexicon in forming their utterance. Expressions in the mental lexicon, in turn, are grounded by *de facto* deference, and so, are individuated by virtue of a causal/historical/social network of deference.

With respect to our second question about what speakers must know or do in order to successfully *apply* a name, we noted that many of the warnings about misarticulating a name have echoes in a speaker's ignorance and errors in applying that name. Our knowledge might be dramatically incomplete (as well as erroneous). Even if all we know about Feynman is that he's a physicist, we can still use "Feynman" to refer to a particular physicist, namely, Feynman. Indeed, even if whatever minimal information about Feynman we have is incorrect (e.g., we think he's a novelist), we can still use the name "Feynman" to say things about Feynman (cf. Kripke (1980)).

Likewise, even if it is commonly assumed that "Godel" picks out the man who first proved the incompleteness of arithmetic, "...it is perfectly intelligible to suppose that it might be discovered that Godel was not the first to prove incompleteness..." (Kripke 1980). But must there be some other description whereby we pick out Godel? And, if not, isn't the use of the name by someone so ignorant or misinformed a mere case of parroting? That is, if someone doesn't know anything about Godel at all, can she still *really* use 'Godel' to refer to Godel (cf., Dummett 1991: Ch. 4)? And if so, does that use count as a successful use?

According to us, what matters is whether someone exposed the speaker to a name of an individual. If so, then for all subsequent uses, the speaker *de facto* defers to the exposer with respect to the name. The speaker need not know that she does so. More precisely: so long as the speaker is tokening the name "Aristotle", the name names whomever it names in the network of deference the speaker is participating in (if it names anyone). The speaker *de facto* defers, because the name is grounded via the network.

Not everyone agrees. Dummett, for example, has replies to both the argument from error and the argument from ignorance. On our view,

if a child is introduced to the name "Newton" with the description "the man who discovered that there is a force pulling things to earth," then, even though this gives the child a false belief about Newton, the child can still reference Newton with her uses of "Newton". In this respect, therefore, even the description used to fix the referent needn't be true of individual named. Here Dummett balks, labeling the view—the 'heroic' course, namely, the view that "…someone who had no more than heard the name "Newton" without any means of fixing its referent, without knowing anything at all about its bearer, would nevertheless *understand* it and be capable of using it with the reference commonly attached to it" (Dummett 1973: 137, emphasis our own).

Dummett is equally skeptical about the limitlessness of ignorance. He writes, "...there are certainly cases in which a proper name is used without its user attaching to it anything that Frege would consider a sense. If, when I come home, one of my children says to me, "Mr. Cunningham telephoned and asked if you would ring him back", the child may no more know the sense or the reference of the name "Mr. Cunningham", which, let us suppose, he has never heard before, than does a piece of paper on which such a message is written; the child is acting merely as a recording apparatus..." (Dummett 1973: 138).

Dummett insists upon replacing the network model with a cluster/division of labor proposal, according to which, "…what makes it possible to entertain the possibility that Godel might be discovered not to have proved, or not to have been the first to prove, the incompleteness of arithmetic is the fact that there exist other generally accepted ways of determining the reference of the name "Godel". This is always the case with any name about whose bearer a good deal is known by at least some who use the name; and it is never the case with a name about whose bearer practically nothing is known save that it satisfies the description which fixes the reference of the name" (Dummett 1973: 139).[10]

[10] Note that in regards to individual speaker's ignorance, Dummett responds that "one of the ways in which it is essential to language that it is a common instrument of communication is that there is no sharp line between the case in which a speaker makes a fully conscious employment of the sense canonically attached to a word and that in which he acts as a recording apparatus. We are able to exploit the fact that a word has a generally recognized sense, which may be discovered by standard means, even when we have only a partial knowledge of that sense; and we do […]" "[it] is not possible that none of those who use a name have any criterion for identifying the bearer of the name, that all of the use it with only partial criterion in mind, but with the intention of referring to the commonly agreed referent" (Dummett 1973: 139–40). We caution, again, that is important to separate meaning determination—the metaphysical question we can set aside—from successful tokening and application of a term. Provided the meaning is fixed—in whichever way—there is no pre-requisite, on our view, on successful tokening or application that *any* speaker has even partial knowledge of the *meaning* or that they have an intention to refer to a particular referent, or defer to a particular community. We will return below to the claim, often repeated in the literature, that there has to be *someone* in the community who possesses the relevant linguistic knowledge.

In this passage lies the seeds of Dummett's dismissal of the network model; he further writes, "Kripke expressly wishes to allow that the association with a name of a description which in fact does not apply to the person or thing for which the name was originally introduced does not deprive that name of reference to that person or thing: it merely reveals a false belief about the referent of the name. There is therefore no room in Kripke's account for a shift of reference in the course of a chain of communication: the existence of such a chain, accompanied all the time by the required intention to preserve reference, must be taken as guaranteeing that reference is in fact preserved. Intuitively, however, there is no such guarantee: it is perfectly possible that, in the course of the chain, the reference has been unwittingly transferred. Once this is conceded, the account crumbles away altogether. We are left with this: that a name refers to an object if there exists a chain of communication, stretching back to the introduction of the name as standing for that object, at each stage of which there was a successful intention to preserve its reference. This proposition is indisputably true; but hardly illuminating" (Dumett 1973: 151).

Dummmett's view is obviously in sharp contrast with our own. And there is much in what he says in these quoted passages above that we take issue with; for example, his insistence on understanding, and his worries about shifting reference, as presenting insuperable problems for the network model. Elsewhere, we take on these challenges (Stojnić and Lepore, ms; Stojnić, ms). But, for now, it's best we proceed with our own positive view, according to which, to repeat, much like successfully tokening a name, successfully using a name doesn't require speakers to know (much of) anything.[11]

To successfully use a name, a speaker need only token it, i.e., select it from his (mental) lexicon. Its meaning (referent) is, in turn, grounded

[11] It is perhaps worth pausing for a moment on the shifting reference problem. Since on our account the meaning of an expression (if any) is transferred through a network of deference, what do we say of cases of apparent shifts in meaning, as might be with, e.g., 'Madagascar' (cf. Evans 1973)? If meaning is deferentially transferred through a network, then aren't we bound, via Marco Polo's mistake, to refer to a part of mainland African content with our uses of 'Madagascar'? While we have no space to defend this view here, we maintain that the alleged shifts in meaning are best understood as novel acts of neologizing, whereby a new expression is introduced and a novel meaning for it might be grounded (Stojnić, ms). Such (re-)baptsims can occur either transparently to agents involved, or tacitly (just as any other introduction of a novel word can be a conscious effort on the part of the speaker—as when the speaker says 'I'll name you 'Alice"—or can happen without interlocutors realizing they are introducing a novel word—as might be with some instances of zero derivation, e.g. by uttering "He houdinied his way out of the cell", without either the speaker or the audience realizing "to houdini" is not already a word). Notice that, how meaning gets fixed (if at all) in the re-baptism case, is the same metaphysical question of how meaning is fixed in the baptism case that we do not purport to answer here. What is important for our purposes is that the chain of deference only takes one to the (nearest) baptism event. That a homonymous expression might have been previously introduced with a different meaning is simply irrelevant.

by *de facto* deference to the tokening of the name by the member of the network from whom the speaker acquired it (and so on). With that said, we are now ready to transition from proper name expressions to predicative ones.

3. *Predicates and their tokenings/articulations and predications*

As in the case of names, we have two central questions:

1. What does it take to successfully token/articulate a predicate?
2. What does it take to successfully use/apply it (e.g., to ascribe a property)?

In brief, our main pitch is that the shift from names to predicates is seamless since, mostly, what goes for names goes for predicates—with some qualification. And so, our answers to the two questions driving this discussion will look familiar. We begin with common nouns.

The common noun "water" is not a name, but, much like one, at some point, and in some manner—perhaps, by speaking it—it was introduced into the language. And, in some manner (perhaps, by speaking it while pointing at a particular body of liquid, though, again, the details do not concern us), let's assume, its extension is fixed (and so, it is settled what "water" means). (As in the case of names, we don't really care about how exactly "water" was introduced, or how its meaning was fixed.)

Of course, there are differences between names and common nouns. For one, the extension of "water" is not what it *names*. The neologist who introduced the common noun was not intending to *name* a particular body of liquid, but instead might have been pointing at it as an exemplar of a property, and somehow thereby fixed its extension to include whatever it is *true of*. (Again, this part of the story is not our focus.)

There are still shared key features, despite these differences. We note the obvious, namely, just as with a name, a common noun can be spelled and pronounced in various ways, and, as a matter of fact, it has been across times and places. And, much like a name, it can be, and has been, misarticulated, if by that is meant the term can be successfully used (tokened) even when its use on a given occasion departs from its customary articulations. This is, indeed, a familiar, and perfectly general, lexical phenomenon, not isolated to names and predicates.

To illustrate, consider the distinct words "bear" and "bare". Were someone to write, "Bare with me!", our reaction would *not* be to ascribe a new meaning to an old word ("bare"), but, rather, to say the speaker misarticulated another old word ("bear"). The speaker *did* request the addressee to *bear* with the speaker, but misspelled the word "bear" as "bare"—a misarticulation. And so, the speaker can be taken not only to mean for the addressee to bear with the speaker*, but also even to have*

said it. (It is also possible, though less likely, that the speaker did, in fact, token "bare", perhaps because she mistakenly believed that "to bare" means *to bear*. Either of these mistakes are possible, but crucially, neither involves assigning *a new meaning* to the term "bare" (cf. Kripke (1977) on speaker reference).)

The difference between expressions and their articulations played a key role in our answer to the first question about names: namely, what does it take to token a name? So, our answer here to our first question about tokening common nouns is going to be the same as the one we gave in the case of names.

Which expression a speaker is tokening depends, on our current account, *on who*, so to speak, introduced the expression into the current *speaker's (mental) lexicon*; if that prior speaker tokened the word "bear," when (perhaps unwittingly) introducing the current speaker to the word which she is now tokening, then that is the word the current speaker is tokening, even if she articulates it as "bare". Or more precisely, if she selects the expression "bear" in her mental lexicon, where this is the expression she was introduced to by another speaker's tokening of "bear" (and so on), then she will have tokened this expression, even if she (mis)articulated it as "bare". So, on our account, one could be tokening, and so, saying that the addressee should bear with the speaker, even if she is misarticulating this as "bare with me". Notice, though, we are not saying the content asserted—*what is said* in the Gricean sense—is determined by speaker intentions. In particular, we are not endorsing Intentionalism about what—which content—the speaker asserts when speaking. Rather, which linguistic form the speaker uses depends on which expression she articulates. Which expression she articulated depends on which expression she tokened, i.e., which expression in her mental lexicon she selected in forming her utterance.

The situation is the same as in the case of names. When the speaker says, "Godel is smart", that the speaker is tokening the name "Godel," rather than, say, "Smith", is a matter of which expression the speaker is actually tokening. But that doesn't mean it's up to the speaker's intentions whom "Godel" names. That is a matter of the meaning of "Godel". A speaker can token "Godel" mistakenly, thinking it named Smith, or mistakenly articulate another name, e.g., "Smith" as "Godel". But neither fact makes it the case that the meaning of the name "Godel" is up to the speaker's intentions.

The point is perhaps even easier to see when the focus shifts to context-sensitive items. So, consider an utterance of "She is happy". That the speaker is using the third person singular female pronoun "she", rather than, say, the male one "he", or the proper name "John", or any other expression, is a matter of which expression is actually being tokened. But whom "she" picks out is *not* a matter of whom the speaker intends to pick out with the expression she is tokening; for example, if

pointing at Mary, "she" will pick out Mary, even if she intends someone else (see Stojnić at al, 2013, 2017). Further, one can assert "She is happy," pointing at a man, because one mistakenly misarticulated the third person singular male pronoun "he". In this case, the speaker doesn't mistakenly believe the man is a female; she just misarticulates the word. (This, we take it, is a common misarticulation for non-native speakers of English whose first languages lack gendered pronouns.) But one can assert "She is happy" tokening the third person singular female pronoun "she" because one mistakenly thinks of the man that he is female. In neither of these cases is it a matter of the speaker's intentions fixing the referent of "he".

It is worthwhile comparing these two types of error. In the first case, the speaker said of a male that he is happy; in the second, the speaker said nothing at all, or something false. In the first case, the speaker is making an articulatory error; in the second, the speaker is making a non-linguistic one. In both, the audience has resources to try to make sense of what the speaker said. They can reason the speaker made a slip, and try to figure out which word the speaker misarticulated, or conclude the speaker accidentally mistook a male for a female. They have to determine the logical form the speaker uttered, or try to make sense of the utterance, by identifying the speaker's background false beliefs. (Note that, with Kripke (1977), we can still maintain that one can 'speaker refer', i.e., manage to convey that the male the speaker "had in mind" is happy (though Kripke doesn't distinguish different sources of error). The audience can figure out that the speaker probably mistakenly used the female gender pronoun to refer to a man, and so figure out the message the speaker intended to convey. But even so, it is crucial to separate disambiguation from meaning (determination). Even if the speaker manages to convey the message she ultimately intended, this is not because some new meaning is attached to "he", just as, in the earlier example, it was not because some new meaning was assigned to "Godel". It is rather because the audience can disambiguate the form the speaker either tokened but misarticulated, or should have tokened save for their erroneous belief that "she" is a male-gendered pronoun, in "He is happy".)

What about our second question concerning successful application? What must speakers know or do in order for their uses of the common noun "water" to succeed in being about anything, and in particular, about water? Put differently, how is successful application achieved for uses of the common noun by speakers who are not neologizing the term—that is, ordinary folk in the same linguistic network?

According to the commonsensical view, competent speakers carry (clusters of) identifying or individuating criteria in their heads that they associate with a word (recall Dummett's claims above). They succeed in talking about something, e.g., with uses of "water", only if whatever is included among this stuff satisfies (a cluster of) the criteria they

associate with "water". This cluster is the meaning of the word. So construed, what speakers must know in order to know the meaning of "water", and so, to fix its extension, is something like: "whatever satisfies "water" does so *only if* it is what fills our oceans and lakes and rivers, comes out of our taps, quenches our thirst, etc."

Of course, not everyone agrees. After all, the information that interlocutors associate with the word "water" can be mistaken, or so incomplete it fails to separate what belongs to the extension of "water" from what does not (H_2O vs XYZ), and yet, it seems, that successful application of the use of the word might still result. As Evans reminds us, "We constantly use general terms of whose satisfaction conditions we have but the dimmest idea. "Microbiologist", "chlorine" (the stuff in swimming pools), "nicotine" (the stuff in cigarettes); these (and countless other words) we cannot define nor offer remarks which would distinguish their meaning from that of closely related words" (Evans 1973). How is this possible?

A familiar response is that, just as through a practice of deference, ordinary folks can use the words they use, so too, through a practice of deference, they can exact successful application of their words as well. If this is correct, then neither the false nor insufficient information in our heads need thwart our successful application of uses of "water". But, while appealing to deference is a common response, how should we understand this sort of deference; viz., deference to whom is relevant? And how can a speaker's deference to anyone help to secure the successful application of an expression, if (as we will maintain) no one need be any *less* mistaken or ill-informed than anyone else?

Here is where a division of labor often enters the discussion. The idea is that in order for there to be a successful application of a word, *somebody* in the network of users must know (a cluster of) necessary and sufficient identifying or individuating conditions for what falls under extension of that word (see, e.g., Putnam 1975). This "expert" needn't be the occurrent user who carries this information. Nor need it even be neologist who coined the term. For a concrete example, consider "water", where all that matters for successful application, on any occasion of use, is that whoever uses it defers to relevant experts about what "water" is true of, or at least about what the relevant individuative nature (property) is of whatever "water" is true of. That is, there has to be some arbiter in possession of relevant knowledge to whom others defer. There has to be an expert.

To elaborate, suppose a speaker carries erroneous or incomplete lexical information about the application of "water". The speaker has heard it used, but misremembered it as being about a liquid fluid state, and so, rules out its gaseous and frozen forms, e.g., or doesn't know enough to distinguish what it applies to from a range of other odorless, tasteless, thirst quenching liquids. Still, if exposed to the word, then, even though confused, or with incomplete knowledge about what it's

true of, its uses can still successfully apply. (For instance, they can successfully say something false if they say "Water is always liquid", or successfully make a request if they say "Give me a glass of water, please".) According to the division of labor thesis, the speaker need only defer to experts on the meaning of "water" for its uses to be successfully applied. Putnam, an early advocate of the thesis writes:

> We could hardly use such words as "elm" and "aluminum" if no one possessed a way of recognizing elm trees and aluminum metal.....Everyone to whom gold is important for any reason has to acquire the word "gold"; but he does not have to acquire a method of recognizing if something is or is not gold. He can rely on a special subclass of speakers. (Putnam 1975)

A commitment to the linguistic division of labor means that a speaker *cannot* enter the network surrounding uses of a noun like "water" unless that speaker defers to "experts" on the meaning of "water" (or on which property it expresses). This is partly what it means to be lexically competent with the word "water". What counts as an expert can vary from context to context. In some contexts, we may care more about underlying composition, and in others, more about functional relations. Different concerns may force us to change allegiances with respect to who the relevant experts are.

So understood, it should be clear that the neologist needn't be an expert. While pointing at some stuff, a neologist may presume that that stuff, and whatever else "water" projects to, shares some property in virtue of which all this stuff has the same composition, and so, thereby falls under the extension of "water". But this does not require the individual to know what that property's composition consists in. That individual may have erroneous or incomplete information about the denotation of "water". However, the underlying assumption is that there are experts somewhere in the network who have identifying or individuating information, and community's deference to them is required (though not sufficient) to account for successful uses of "water".

But why do we need experts? No matter how ignorant or misinformed anyone, *or everyone*, in the network is, including whoever neologized the expression, successful predication can still ensue. (Clearly, we can introduce a term labeling a poorly understood phenomenon, only to learn about the phenomenon later on, with the understanding that we possibly might never master it.) Indeed, Putnam's own paradigmatic example of "water" as used in 1750 is revealing in exactly this respect. Putnam writes (1975):

> In 1750, chemistry was not developed sufficiently to individuate what we call "water" from all other chemical compounds. No one knew about hydrogen and oxygen compounds. Still, when speakers used "water," they succeeded in picking out what we pick out with current uses. That's why it makes sense to say they were wrong about what their uses of "water" applied to, even though these uses still succeeded in picking out water and only water. In short, that's why we can say that we disagree with them about *water*.

How can Putnam reconcile these intuitions with the division of labor thesis? How can membership in the network require a division of labor if all past and present users of the term can be wrong about the composition of its extension? Putnam (1975) attempts to remain committed to linguistic division of labor, even in face of his own "water" counterexamples. He suggests that even in 1750 speakers were deferring to experts, just not past or contemporaneous ones; rather, future ones. (Mysterious!)

So, what can we conclude in the cases of an absence of experts? After all, there may never be an expert, even in principle, among members of our species—if that requires someone who uncovers the nature of the extension of "water". And if the world ended before there were an expert, it's not like everyone would have failed to talk about water.[12] So, what did Putnam intend; is it just a metaphor for *the nature of things?* Since there is a fact of the matter about the nature of whatever "water" picks out, it follows that, even though no science may ever uncover this nature, we can still imagine an omniscient expert who knows all natures.

The problem for Putnam with this suggestion is that it doesn't exploit the expert to determine what "water" picks out. It uses the fact the "water" is true of something-or-other, to determine what it would take to be an expert about *that*. In particular, this way of exploiting expertise doesn't require that anyone actually "possesses a way of recognizing" whether "water" is true of something, at least not anyone in our network (even considered diachronically).

The key idea here is that once the connection between a common noun and its denotation has been established (say, e.g., for "water"), it becomes explorable as to what the nature is of what is picked out by uses of the noun. And though, in some cases, there may be experts about the nature of this property, and though we may defer to them, there is no guarantee that such experts (ever will) exist. But no such guarantee seems necessary in order to secure successful uses of expressions of our language. So, if in order to successfully use a common noun a speaker must defer, then to whom must the speaker defer with a use of "water", if not to knowledgeable experts about the nature of what's in its extension?

Well, on our story, once again, we are assuming that the speaker who introduced the term somehow managed to fix its meaning, and so, its extension. We deny that this speaker did so solely by explicitly intending to pick out some particular sort of stuff or property, since this first speaker was almost certainly wrong about the extension of "water" as well as about nature of its extension, and so might have been every user of the word since. One way around this is just to say the speaker

[12] Does the move to context-specific experts help here? We do not see how. For one, just as there is no guarantee that there is an expert with respect to the nature of a property, there's no guarantee there is an expert with respect to the property relative to some contextually specified purpose we are interested in.

intended to pick out *that thing*, where "that thing" is whatever is the thing that's being actually picked out (if any). Such demonstrative intentions are cheap (easy to form), but they don't provide a rich body of information to be used in identification.

This brings us to our answer to the second question about the successful use of predicates. Accordingly, suppose a progenitor succeeded, despite an abundance of erroneous and/or incomplete information, in introducing a new expression and in fixing its meaning/extension. A network got initiated, where upon all future users of the word can defer (*de facto*) to a chain of speakers back to its initial application ceremony. As before, this requires no *intentional act* of deference on the part of the speaker. All the speaker needs to do is intend to select an expression from her mental lexicon. The expression's identity, and its meaning, is determined by *de facto* deference to the network. And so, the answer to the second question is just as with the answer to the first question about expression tokening; namely, it's what speakers do, not what they know, that enables them to apply words successfully. A speaker selects an *expression* in her (mental) lexicon. In selecting this expression, she *de facto* defers to whomever passed the expression on to her. This doesn't require the speaker or the introducer to be experts or possess any identifying information, or even that the speaker forms an intention to defer to anyone (including experts).

4. *How far does the account extend?*

So far, we have speculated about proper names like "Godel" and common nouns for natural kinds like "water"; but how far can the network model be pushed? Defending his own version of the network model, Burge (1979) argues it has "an extremely wide application," and it does not depend on the kinds of words, say, that "Godel" and "water" are.[13] Indeed, he writes, the network extends to "an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words" (1979). In fact, Burge is clear that the network extends to "any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands" (1979). Similarly, Putnam (1975), though he highlighted natural kind terms, notes that deference is practiced with many other kinds of words as well.

[13] Arguably, Burge would disagree with our non-intentional way of characterizing deference. As explained earlier, throughout most of the literature, it has been assumed that the appropriate kind of deference requires at least an intention to defer. If we are right, even this requirement is too strong. Be that as it may, since our argument crucially relies only on the possibility of ignorance of, and error about, an expression's articulation *and* its meaning, then whenever we have a case of apparent successful use of a term in spite of the possibility of such ignorance and error, our account will equally extend.

But given *our* view of deference, how far does the network model extend? We think the arguments from ignorance and error extend to most (all?) expressions. Clearly, the distinction between expressions and their articulations extends to all expressions. And it seems that, for *any* expression, a speaker can be mistaken about, or ignorant about its articulation. On our account, tokening any expression requires simply selecting it from a mental lexicon. Potential ignorance or error about its identity or articulation are no obstacle to successful tokenings; the expression is individuate by *de facto* deference to the tokening which introduced it into the speakers mental lexicon. Further, we think likewise, for virtually any expression the speaker can successfully *use* it—apply it—regardless of their ignorance or error about their meaning. We have seen how this extends to names and predicates. We think they quite generally, indeed even to connectives. Think about the debate over the meaning and logic of a conditional (cf. Grice (1989a), McGee (1985)) or the issues concerning commutativity of a conjunction in English.[14] Surely, it is not an obstacle to the successful tokening or application of the English conditional, or conjunction, that one might be mistaken about, or even have false beliefs about, some of the inferences that the conditional licenses. While establishing these extensions and what are their virtues in full is something we attempt elsewhere (Stojnić and Lepore, ms), here we note that, as long as the arguments from ignorance and error extend to a class of expressions, it should be clear, so do our answers to (1) and (2).

Conclusion

As sated at the outset, our goals here have been modest. We argued for an account of linguistic deference understood not as an intentional mental act—underscored by an intention to defer—but rather as what we called *de facto* deference—deference as a matter of historical and causal connections that trace the way the linguistic item was introduced into the speaker's mental lexicon. This allowed us to elucidate how speakers can successfully token and apply expressions despite the fact that they might be ignorant, or confused about the expressions' articulation and meaning. To token an expression, and to apply it successfully, speakers don't have to *know* anything; they rather have to do something: they have to select an item from their mental lexicon in forming their utterance. Which item it is that is selected, in turn is determined by the *de facto* deference to the item that was tokened by whomever introduced the speaker to the word. Similarly, its meaning is determined by *de facto* deference to whomever introduced the speaker to the word (and so on, back to the neologizing event). This way of indi-

[14] Some argue that apparent failures of commutativity are due to pragmatic effects (e.g., Grice (1989b)); the proponents of the dynamic semantics for conjunction, in turn, typically argue for a non-commutative meaning for "and" (see, e.g. Groenendijk and Stokhof (1991), or Heim (1982)).

viduating expressions not only explains how one can successfully token and apply a term, despite potential ignorance and error, but allows us to carefully distinguish the interpretive task of disambiguation—the process whereby audience determines which term was uttered—from the metaphysical process of meaning determination. We take this to be a theoretical virtue of our account.

References

Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–121.

Davidson, D. 1979. "Quotation." *Theory and Decision* 11 (1):27–40. (Reprinted in *Inquiries into Truth and Interpretation.* New York: Oxford, 1984: 79–92.)

Devitt, M. 2015. "Should Proper Names Still Seem So Problematic?" In *On Reference*, Andrea Bianchi, ed. Oxford: Oxford University Press: 108–43.

Dummett, M. 1973. *Frege: Philosophy of Language.* Cambridge: Harvard University Press.

_____, 1991. *The Logical Basis of Metaphysics.* Cambridge: Harvard University Press.

Evans. G. 1973. "The Causal Theory of Names." *Proceedings of the Aristotelian Society, Supplementary Volumes* 47: 187–225.

Grice, H. P. 1989a. "Indicative Conditionals." In *Studies in the Way of Words.* Cambridge: Harvard University Press: 58–85.

_____, 1989b. "Logic and Conversation." In *Studies in the Way of Words.* Cambridge: Harvard University Press: 22–40.

Groenendijk J., and Stokhof M. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14(1): 39–100.

Hawthorne J., and Lepore E. 2011. "On Words." *The Journal of Philosophy* 108 (9): 447–485.

Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts doctoral dissertation.

Kaplan, D. 1990. "Words." *Proceedings of the Aristotelian Society*, Supplementary Volumes 64: 93–119.

_____, 2011. "Comments and Criticisms: Words on Words." *Journal of Philosophy* 108 (9): 504–529.

Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

_____, 1977. "Semantic Reference and Speaker reference." *Midwest Studies in Philosophy*.

Lepore, E. 2009. "The Heresy of Paraphrase: When the medium really is the message." *Midwest Studies in Philosophy* 33 (1): 177–197.

Lewis, D. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (3): 221–236.

McGee, V. 1985. "A Counterexample to Modus Ponens." *The Journal of Philosophy* 82 (9): 462–471.

Putnam, H. 1975. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131–193.

Stojnić, U. 2017. "Discourse and Logical Form: pronouns, attention and coherence". *Linguistics and Philosophy* 40 (5): 519–547.

_____, 2019. Ms. "Just Words." Princeton University.

Stojnić, U., and Lepore E., Ms. *Communication in a Sea of Ignorance and Error.* Princeton University and Rutgers University 2019.

Stojnić U., Stone M., and Lepore E. 2013. "Deixis (Even Without Pointing)." *Philosophical Perspectives* 27 (1): 502–525.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Making Meaning Manifest

MARILYNN JOHNSON
University of San Diego, San Diego, USA

In recent work Sperber and Wilson expand on ideas initially presented in Relevance *(1986) and flesh out continuua between showing and meaning, and determinate and indeterminate content. Drawing on Sperber and Wilson's work, and at points defending it from what I see as potential objections, I present a Schema of Communicative Acts (SCA) that includes an additional third continuum between linguistic and non-linguistic content. The SCA clears the way for consideration of what exactly is meant by showing, the motivations of speakers, how affect impacts expression, and metaphor. The SCA allows us to consider not only how but why we engage in certain forms of communicative behavior, and captures the incredible nuance of human interactions: said and meant, linguistic and non-linguistic, determinate and indeterminate.*

Keywords: Sperber and Wilson, Grice, meaning, showing, determinate, indeterminate, linguistic, non-linguistic, metaphor, affect.

1. *Introduction*

Every philosophy of language is distinguished not just by its theoretical core but by the sorts of cases that it considers for explanation. The pragmatic tradition, which includes work by philosophers such as H. P. Grice, Dan Sperber, and Deirdre Wilson, stands apart from predecessors in part because of engagement with how we actually use language, "in the wild"—with meanings and to achieve aims that are not explicitly stated, but suggested or implicated. Language is not best understood in terms of coding meanings, as Sperber and Wilson convincingly argue in the introduction to their 1986 book *Relevance*, but on a continuum with other communicative acts.

The work of Sperber and Wilson builds on the tradition Grice began in the 1950s. In work published in a 2015 edition of the *Croatian Journal of Philosophy* Sperber and Wilson expand on some ideas initially presented in *Relevance.* In (2015) paper Sperber and Wilson expand

their theory to explicitly consider cases of meaning as well as showing, and discuss acts with determinate as well as indeterminate content. Sperber and Wilson's work is remarkable for its willingness to explain acts not just of ordinary utterances but also "ostensive" acts such as sniffing the seaside in a way that makes it clear the sniffer is "sharing an impression" with her audience.

In this same spirit, I expand further on the account presented by Sperber and Wilson, and defend it from what I see as a possible misconstrual of their view. The theoretical framework presented by Sperber and Wilson brings to light some important questions for their account: What does the distinction between meaning and showing amount to? Is this distinction tantamount to a distinction between expressing content linguistically or non-linguistically? Why do we in some circumstances mean/state propositions and why in others do we show evidence? Is this a conscious decision? What is the relationship between conscious awareness and meaning/showing more broadly? I will respond to these questions and will consider a number of communicative acts that go beyond the sorts of cases that are ordinarily considered by philosophers of language—such as utterances that express affective states. I argue that with the clarifications I propose the Sperber and Wilson account has the latitude to account for such acts.

2. *Gricean intentions*

Grice's theory of speaker meaning is known to be complex. As characterized by Sperber and Wilson (2015), on Grice's view:

> In order to mean something by an utterance, the utterer must intend the addressee,
>
> 1) to produce a particular response r
> 2) to think (recognise) that the utterer intends (1)
> 3) to fulfil (1) on the basis of his fulfilment of (2) (118)

What is important about Grice's view is the way that meaning may go beyond the literal words uttered. For example, consider a scenario in which someone taps the person in the row in front of them at the theater and utters

(A) "I cannot see over your hat".

It would be surprising if the person in the hat simply said "Oh I am sorry to hear that, but thanks for letting me know", and turned back around in their seat. The first speaker was not intending to simply inform the hat-wearer of a fact. Here the intended response—which will be readily available to any competent hearer—is the hearer will remove his or her hat. It is by the hearer recognizing that this is what the utterer intends that the hearer will remove his or her hat. That is, to put it in terms of Grice's view as stated above, the hearer will (2) recognize that the speaker intends to get the hearer (1) to remove the hat and will (3) fulfill the request (1), removing the hat, on the basis of the

fulfillment of (2), the recognition of that intention. This sounds complex but any witness to the exchange would expect the hearer to remove his or her hat, an expectation that demonstrates an understanding of such an intention on the part of the speaker.

The complexity of Grice's proposal has led to criticisms. Jennifer Hornsby, for example, writes the following of Grice's theory:

> I think that this ought to seem ludicrous. Real people regularly get things across with their utterances; but real people do not regularly possess, still less act upon, intentions of this sort...notice that an enormous amount would be demanded of hearers, as well as speakers, if such complex intentions really were needed to say things. (Hornsby 2000: 95)

The complexity of the Gricean account does raise questions. Are we supposed to spell out all the intentions required for speaker meaning in our head? If so, need we be conscious of doing this? Wouldn't that take a long time? If not, in virtue of what can it be said that some speaker really has such an intention? Or, to put it in Gricean terms, can there be unconscious m-intentions?

Further complicating things are a number of familiar cases where it seems any relevant intention would need to be more elaborate than the hat case. Metaphors such as,

(B) "Juliet is the sun"

might be taken to express a range of propositions, but not including that Juliet is a giant ball of gas. Must a speaker have intended all of the acceptable propositions the metaphor can be said to express? Is it *that* intention in virtue of which they *are* acceptable? If not, what is the reason for their acceptability?

One case Grice considers is the letter of recommendation example, where an utterer conveys that a job candidate, Mr. X, is no good by writing a very short letter of recommendation stating simply that the candidate is on time and is a competent speaker of English (Grice 1989: 33). In this example, the speaker flouts the maxim of quantity to communicate by conversational implicature (Grice 1989: 33). There are other cases, of a sort that Grice does not consider, where an attitude is conveyed, but it is not by means of conversational implicature (which requires intentional flouting on the part of the speaker).

Slips of the tongue do not fall neatly within the Gricean picture. Consider the following example from Davidson,

(C) 'We are all cremated equal' (Davidson 2006: 251).

Are we justified in coming to the conclusion that this speaker *meant* something about death? Or should we say instead that they intended to say 'created'—not 'cremated'—and thus ignore what seems to be revealed through the utterance?

The following case, in which the speaker reveals a negative attitude, is from István Kecskés,

(D) Roy: Are you okay?
Mary: I'm fine, Roy.
Roy: I would have believed you if you hadn't said 'Roy'.
(Kecskes 2014: 2016)

A proficient speaker will recognize that Mary is not fine. As Kecskés draws attention to with this example, there is something about stating someone's name at the end of such a sentence that expresses displeasure. A noteworthy thing about this case is that it may or *may not* have been Mary's intention to convey her displeasure here. In fact, Mary's intention is not relevant to the determination that the speaker is not fine. This means that this content is expressed by a means other than Gricean implicature of the sort that follows the three-pronged framework, as illustrated by case A.

Case D is one that ordinary hearers can pick up on. There are also cases where some expressed content requires a more trained hearer to pick up on. The following is taken taken from Bezuidenhout (2001), who is expanding on Stern (2000)

(E) A young woman Marie, who is in psychotherapy because she is suffering from anorexia nervosa, tells her therapist that her mother has forbidden her to see her boyfriend. Referring to her mother's injunction, Marie utters:
[1] I won't swallow that
Here 'swallow' is being used metaphorically, and Stern suggests that the content of Marie's utterance (the proposition she expressed) can be paraphrased as
[2] Marie won't accept her mother's injunction.
Given her eating disorder, it seems significant that Marie chose to frame her comment about her mother's injunction by using the word 'swallow'. But once we've accessed the metaphorical interpretation it seems that we've lost the echoes of meaning that might connect what she is saying to her eating disorder and hence to any problems that she might be having with her mother connected to this disorder. (Bezuidenhout 2001: 33–34)

As Bezuidenhout points out in this passage if we interpret metaphors in terms of their literal content then we miss out on shades of meaning that seem to be conveyed by the specifics of the metaphor used. Do we need a theory that allows us to say that Marie really did *mean* something about her eating disorder here, although she may not have consciously intended it? Again, if she did mean something about her mother's eating disorder, it is not because of a complex three-pronged Gricean intention. Indeed, it is precisely her lack of awareness of this connection that a skilled therapist would work to identify and point out to her.

Cases B–E are the sort that can prove problematic for any philosophy of language. Metaphor, as in case B, has received a great deal of

attention in the literature, and slips of the tongue have received a fair amount. Less present in the analytic, and certainly Gricean literature, is consideration of cases such as D and E. I will return to consider these cases in a later section. I will approach them as a part of my proposed Schema of Communicative Acts, which will build on the work of Sperber and Wilson. Before we can get to that point I will present the Sperber and Wilson account.

3. *Sperber and Wilson's proposal*

In their 2015 paper "Beyond Speaker's Meaning" Sperber and Wilson present new applications for their notion of ostensive-inferential communication that go beyond what is normally taken as the purview of philosophy of language. Ostensive-inferential communication makes use of just the first two conditions of Grice's theory of speaker meaning; Sperber and Wilson write that is more "conceptually unified" and "does a better job of explaining how utterances are interpreted than a standard Gricean approach" (117).

On the Sperber and Wilson ostensive-inferential view, in order to mean something by an utterance, the utterer must intend the addressee,

1) to produce a particular response *r*
2) to think (recognise) that the utterer intends (1)

Note here that the third Gricean condition has been dropped. Sperber and Wilson explain their dropping the third clause in the following way:

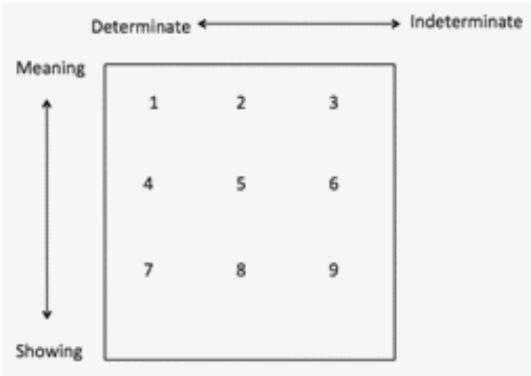
> In characterising ostensive communication, we built on the first two clauses of Grice's definition and dropped the third…because it seemed obvious that there is a continuum of cases between 'meaning that' (typically achieved by the use of language) and displaying evidence that (in other words, showing) and we wanted our account of communication to cover both. (119)

Sperber and Wilson believe that by dropping the third clause—that the recognition of the speaker's intention be *the basis* for a hearer to produce some response—their account covers not only 'meaning that' but 'showing that'.

The central component of the Sperber and Wilson theory—Relevance Theory—is the presumption of relevance. The presumption of relevance is, roughly, the idea that when someone makes an utterance we assume that they have deemed it to be relevant to the conversation, and this knowledge helps us interpret it (Sperber and Wilson 1986). Relevance is one of Grice's four conversational maxims of quantity, quality, relation and manner, which for Grice interact, and the upholding of one often explains why another is violated (Grice 1989; Johnson 2016). In a nutshell of Sperber and Wilson's theory is that relevance alone can do the work that Grice divided into the four maxims. Ostensive-inferential communication is communication that a speaker has deemed relevant.

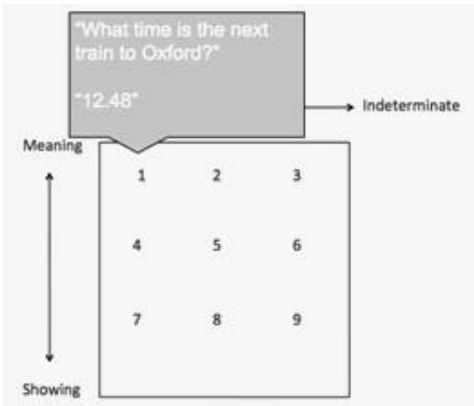
In both the Sperber and Wilson and the Grice characterization, a meaningful utterance is made to "produce a particular response r" in the hearer. This response can be 1) performing a physical action, such as removing a hat, or going away (Grice 1989: 96), or 2) simply coming to have a mental state, such as believing a certain proposition. In other words, the Gricean and Sperber and Wilson accounts can be understood as ways to get others to respond—be that by believing certain things or behaving in certain ways.

Sperber and Wilson go on to consider examples such as 'Juliet is the sun' (2015: 120). Such cases lead Sperber and Wilson to add to their first distinction between showing and meaning—as follows from their dropping of Grice's third clause—with a second distinction, between cases with more or less determinate meaning. A continuum along this distinction is mapped onto the first continuum. They end up with a plane that looks like this:

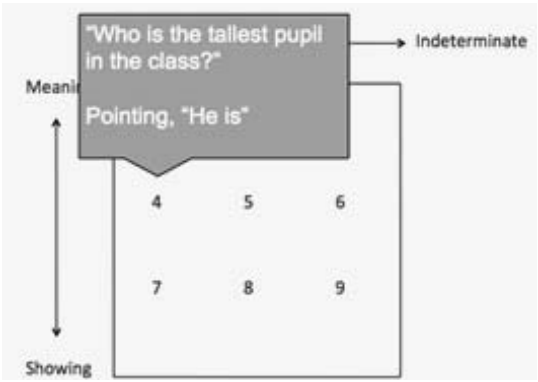


From here Sperber and Wilson proceed to give examples of utterances or behaviors that fall on each of these nine points. These are presented below, beginning with determinate content that is on different points of the meaning-showing continuum (F–H below).

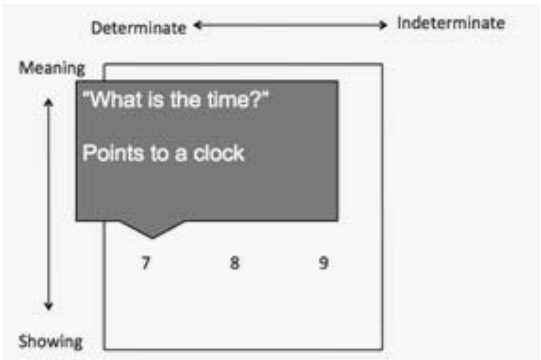
(F)



(G)

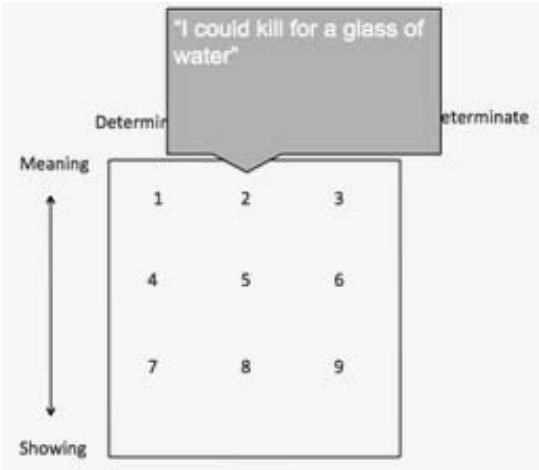


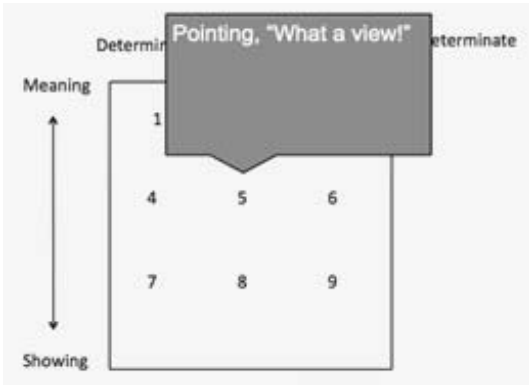
(H)



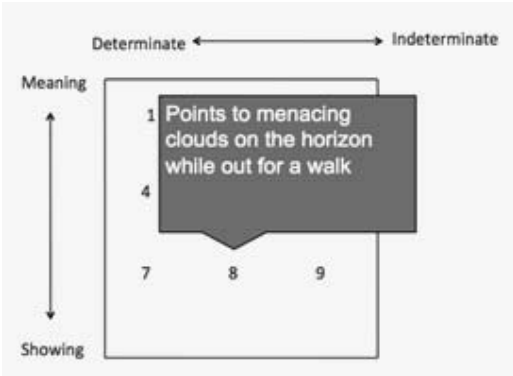
From Sperber and Wilson they present three cases that are between determinate and indeterminate content, and across the meaning-showing continuum (I–K).

(I)



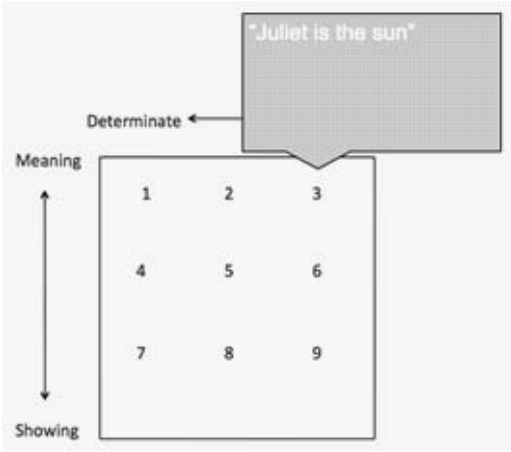


(J)



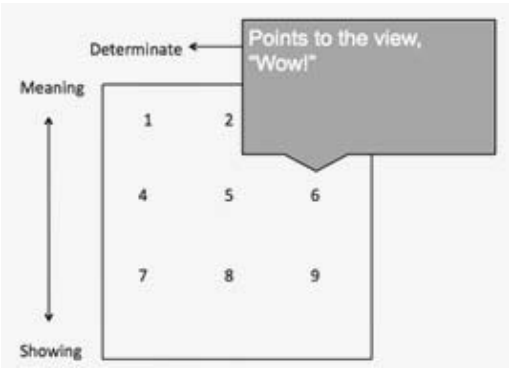
(K)

And lastly, we are presented with indeterminate content, across the meaning-showing range (L–N).

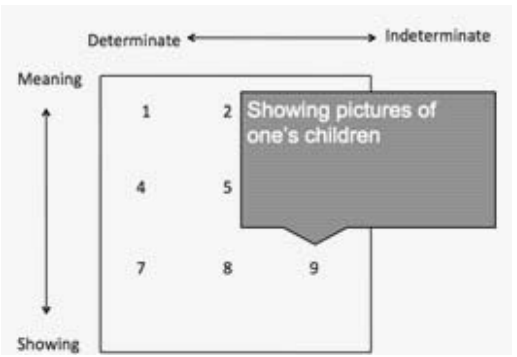


(L)

(M)



(N)



These examples help illustrate what Sperber and Wilson have in mind with these two distinctions between meaning and showing, and between determinate and indeterminate content.

4. *Making manifest and sharing an impression*

With determinate content the response a speaker intends to cause in the hearer is relatively straightforward. With the hat example (A) it was clear that the speaker wanted the hearer to remove his or her hat. With acts on the indeterminate side of the Sperber and Wilson continuum it is much less clear what is going on.

In their 1986 book *Relevance: Communication and Cognition* Sperber and Wilson consider the following case, which is an instance of indeterminate content:

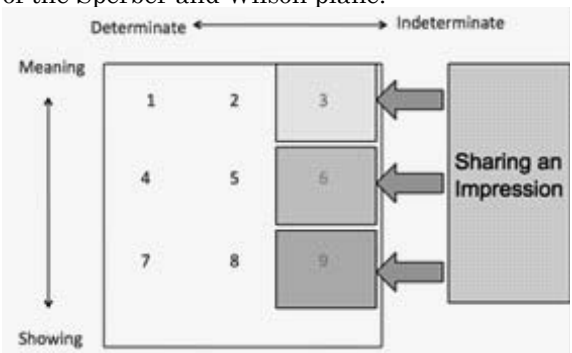
O: Mary and Peter are newly arrived at the seaside. She opens the window overlooking the sea and sniffs appreciatively and ostensively. When Peter follows suit, there is no one particular good thing that comes to his attention: the air smells fresh, fresher than it did in town, it reminds him of their previous holidays, he can smell the sea, seaweed, ozone, fish; all sorts of pleasant things come to mind, and while, because her sniff was appreciative, he is reasonably safe in assuming that she must have

> intended him to notice at least some of them, he is unlikely to be able to pin down her intentions any further. (1986: 55)

In this example Mary behaves in a way that makes it clear that she would like Peter to appreciate the seaside. It is not clear precisely what response she hopes to engender in Peter once he turns his attention to the seaside. If we attempted to spell out which Gricean response, r, Mary has in mind—be it that Peter come to accept some proposition as being true or to perform some action such as taking off a hat—we would fall short.

To address this Mary example[1] Sperber and Wilson present their notion of sharing an impression. They write that if Mary were pressed on what she intended to convey to Peter "one of the best answers" would be that she wanted to share an impression. Cases such as O, where the speaker's meaning is not determinate, cannot be paraphrased without loss (2015: 122).

We can map this notion of sharing an impression on the right side of the Sperber and Wilson plane:



When we express indeterminate content we share an expression.

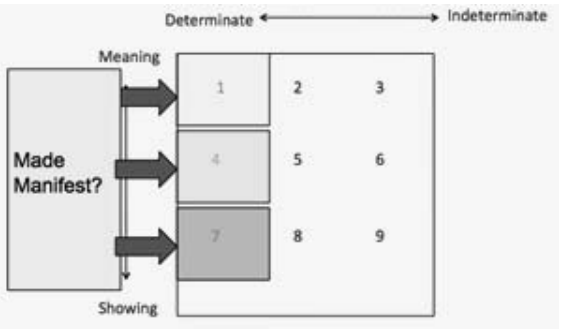
Sharing an impression is contrasted with the notion of making manifest. When some content, p is shown or meant, this is the sort of thing that makes p more manifest on the Sperber and Wilson picture.

They write, "A proposition is manifest to an individual at a given time to the extent that he is likely to some positive degree to entertain it and accept it as true" (134). Manifestness is an *epistemic* notion. In their eyes, "the notion of mutual manifestness is more realistic, more psychologically relevant, and at least as cogent as the notions of mutual knowledge, common knowledge, or common ground" (135). For something to be made manifest it must become salient to the hearer. 'Salience' here is what they called 'accessibility' in *Relevance* (2015: 133). In short,

> Manifestness = epistemic strength + salience (2015: 133)

[1] Unfortunately—and somewhat confusingly given the examples discussed here—Mary or Marie seems to be a popular choice for a female name in hypothetical scenarios; we have seen Mary and Marie already in cases D and E above.

Because manifestness is spelled out in terms of getting a particular proposition across this suggests that it only applies to those instance of meaning or showing that have fully determinate content. For how can a metaphor be made manifest? How can someone "believe or accept it as true" that Juliet is the sun? If manifestness is the sort of thing that can be applied only to utterances and behaviors with determinate content, then we see that manifestness applies to only a certain area of the plane, and on the opposite side from sharing an impression, as I have shown below.



5. *Linguistic and non-linguistic content*

Having presented the Sperber and Wilson framework I will now turn to my proposed addition to it. In their paper, Sperber and Wilson write that 'meaning that' is "typically achieved by the use of language" (119). They do not say that use of language is a necessary or sufficient condition for 'meaning that'. However, all the examples Sperber and Wilson give in their schema of 'meaning that' are linguistic (Examples F, I, and L above). All the intermediary cases are both linguistic and non-linguistic, pointing in conjunction with uttering (Examples G, J, and M above). And all the cases of 'showing that' are non-linguistic (H, K, and N above).

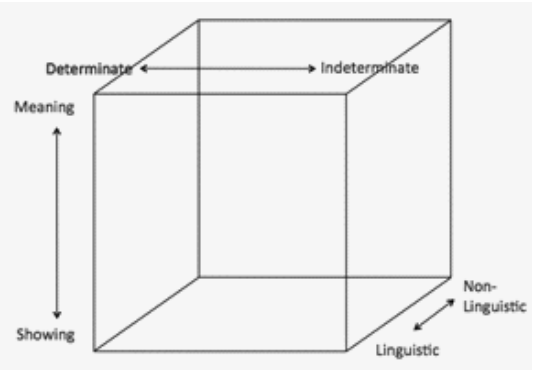
This could be taken to suggest that the distinction between showing that and meaning that is ultimately a distinction between expressing content linguistically and non-linguistically. We might wonder whether 'displaying evidence that' can be achieved by linguistic means and whether 'meaning that' can be achieved by non-linguistic means. What are the consequences of this for a theory of speaker meaning, if any?

Despite their examples perfectly mapping on to a linguistic/non-linguistic distinction in this way, it seems that Sperber and Wilson do not want us to understand 'meaning that' and 'showing that' as a contrast between linguistic and non-linguistic reasons. Again, they do not say that use of language is a necessary or sufficient condition for 'meaning that', merely saying it is "typical" of 'meaning that'.

If this is right, this suggests that there is another continuum between linguistic and non-linguistic cases of showing and meaning that

could be mapped onto the Sperber and Wilson framework as a third dimension. The resulting schema is what I call the Schema of Communicative Acts or SCA.

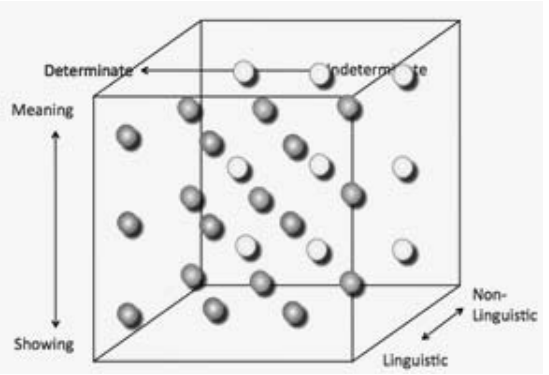
Schema of Communicative Acts (SCA)



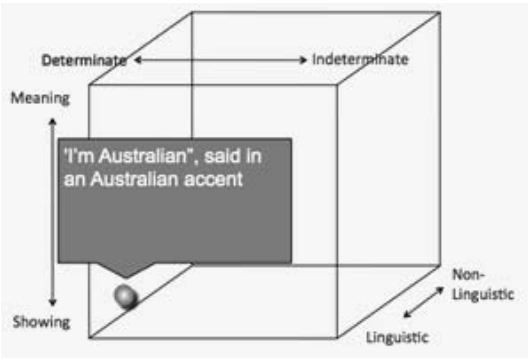
Having this as a third dimension could help to distinguish the contrast between meaning/showing from expressing content linguistically/non-linguistically and better showcase the full range of possible cases of communication.

What we would want now that the new SCA framework is on the table is 27 cases, one for each point of intersection of the three variables. If this cannot be done it puts pressure on the idea that the meaning/showing distinction is not tantamount to a distinction between linguistic and non-linguistic reasons for coming to act.

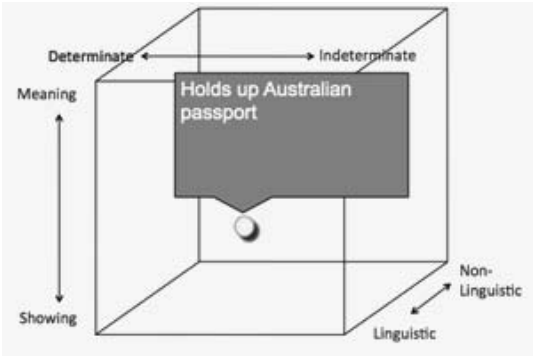
Schema of Communicative Acts With 27 Intersections



We can find instances of determinate showing that are linguistic as well as non-linguistic (J, K).

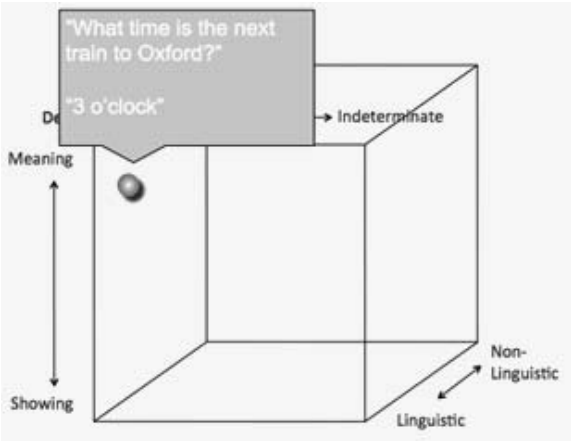


(J)

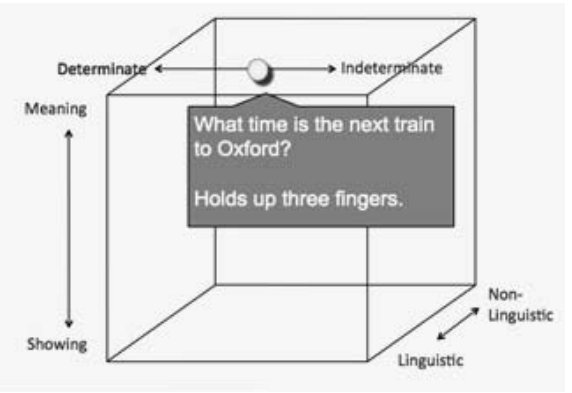


(K)

Likewise, we can find determinate cases of meaning that are linguistic, as well as non-linguistic (L, M).



(L)



(M)

I will not present 27 cases here but what I have begun indicates that it can be done for all 27 intersections. These examples across all three dimensions of the SCA point us away from concluding that the meaning/showing distinction is a linguistic/non-linguistic distinction, as it may have appeared given the examples Sperber and Wilson provide.

6. *On showing*

Once we have clarified that the distinction between meaning and showing isn't tantamount to a distinction between linguistic and non-linguistic content, another question arises, pertaining to what exactly is meant by 'showing'. Showing is said to be "displaying evidence that" (2015: 119). However, this only pushes the question back. What exactly counts as "displaying"? More to the point, how intentional must showing or displaying be, and must acts of showing follow the Sperber and Wilson two-pronged framework for ostensive-inferential communication?

In colloquial use, showing can be intentional or unintentional, as in "your undershirt is showing". If showing is understood in the ordinary sense, it is safe to say that we often show things that are not relevant to the current situation. There certainly are things that are gotten across with utterances that might seem best classified as perhaps unintentional showing, or revealing, as with "I'm fine, Roy". In other words, to put it in Sperber and Wilson's terms, showing, as it is ordinarily understood, does not seem to follow the presumption of relevance.

Taking 'showing' as something looser than a technical term that follows the presumption of relevance, it is clear that we sometimes show—or convey—things we 1) intend to conceal or 2) are unaware of revealing, as in M.

M. Let's say that a man, Antonio, goes to the Metropolitan Museum of Art one day. He is given a pin that says 'MET' that he buttons onto his shirt. He later leaves the museum and rides the subway to Lincoln Center where he attends a performance of the New York Philharmonic Orchestra. Antonio sees a friend at the

> concert and this friend says, "So, you went to the Met today?" Antonio replies, "How do you know?"

At the moment when he saw his friend was Antonio showing that he had been to the Met that day? Was he *displaying evidence that* he had been to the Met that day? Do we need to know more about his mental state? In other words, does showing require an intention?

These questions are important because, with showing, there seems to be a tenuous link between a conscious intention on the part of the agent and what is communicated (that is, gotten across to an interlocutor).

Perhaps an issue at hand in assessing the Met case is one of temporality. It is almost impossible for one to produce an utterance without awareness that one is producing an utterance (although it is possible to construct limited cases). Because of this fact we can presume that a speaker has deemed any utterance to meet some intention *now*. It is this fact that leads to the presumption of relevance. However, with Antonio wearing the pin at the Met the matter is thornier. At the moment he put on the pin we might say he intended to show he had paid the admission fee. We might even say he had an intention that this information continue to be available to a viewer for the duration of his visit. Is such an intention required for showing?

Recall that on the Sperber and Wilson ostensive-inferential view, the utterer must intend the addressee,

> 1) to produce a particular response *r*
> 2) to think (recognise) that the utterer intends (1)

Perhaps Sperber and Wilson wish to restrict their account of showing to those acts that satisfy these conditions. However, I believe that it is more constructive and has more explanatory power if we say that acts of showing need not meet these two conditions. Indeed, the best account of meaning and showing seems to be that meaning must satisfy these two conditions, but *showing need not.* The most explanatorily robust account of showing requires *no intention on the part of the shower.* In addition to what I see as the other benefits of this position, this account of showing, understood in a less restricted way, can account for more examples and more closely aligns with our colloquial use of the term.

Sperber and Wilson may reject this proposal and advocate instead for showing to be understood as a technical term that applies only to ostensive-inferential communication. If so, we need to know more about how to treat cases such as M. On my proposal, what we say about Antonio is straightforward: he is showing that he went to the Met all day, although he is not aware of showing for the majority of the time. If we limit showing to ostensive-inferential communication that meets the two-pronged framework some other treatment of this case is needed. If we say that showing only applies to some early moment of the Met pin application, it seems very difficult to pinpoint when this would be, and why.

Showing in case M is a process with lasting effects. These effects may or may not be intentional. Or, if they were initially intentional, may not

still be intentional by the time they are interpreted. A complete account of showing would include an explanation of how much awareness on the part of the speaker is required for it to be a genuine case of showing.

7. *Motivations*

A further question that arises from this closer consideration of the showing-meaning continuum and the SCA more broadly is about speaker motivations. Why would someone choose to convey meaning in one coordinate or another? The decision to use linguistic or non-linguistic means is perhaps specific to situations. If I am in Croatia, I may hold up the letter 3 to order more glasses of wine for the table rather than speak, because I do not know the language. I may say "excuse me" loudly to someone who is in my way but looking in the other direction. If I am a dancer or a visual artist, my work will be conveyed through non-linguistic means because my training is on one side of this continuum.

The determinate vs. indeterminate continuum is about the nature of the content itself. If someone chooses to express indeterminate content—be it by a sniff at the seaside, a metaphor, a poem, or an abstract painting—this is about the message itself (or here range of messages).

The decision to show or mean, via linguistic or non-linguistic means is a subsequent question about how to get that across. Recall that manifestness, the successful outcome of expressing determinate content, is an explicitly epistemic notion, the extent to which, for any given proposition, the interlocutor "is likely to some positive degree to entertain it and accept it as true" (Sperber and Wilson 2015: 134). Why would someone, on an occasion, choose to provide direct evidence of some fact rather than expect that their communicative intention alone would be enough to cause some response, r, in the hearer? The answer has to do with how they expect they will be interpreted.

Donald Davidson considers this point in his paper "A Nice Derangement of Epitaphs". He writes,

> An interpreter has, at any moment of a speech transaction, what I persist in calling a theory…I assume that the interpreter's theory has been adjusted to the evidence so far available to him: knowledge of the character, dress, role, sex, of the speaker, and whatever else has been gained by the speaker's behavior, linguistic or otherwise. As the speaker speaks his piece the interpreter alters his theory. (2006: 260)

As Davidson writes, an interpreter decides how to interpret on the basis of assessing "character, dress, role, sex, of the speaker, and whatever else has been gained by the speaker's behavior, linguistic or otherwise" (2006: 260). As Davidson later notes, the speaker's theory about the interpreter's theory shapes how he chooses to attempt to convey his meaning.

I recently had a student who told me that she had to miss class because of jury duty. I said that was fine and that she should get the notes from another student. She later emailed me a photo of her jury

summons. I did not require extra evidence to believe that she had jury duty. However, she felt the need to show me direct evidence.

The fact that the student believed my recognition of her intention was insufficient for me to believe she had jury duty is likely the result of her having experienced a failure to achieve a certain result by such means in the past. Thus, learning from this experience, any rational communicator would move down the axis from meaning to showing.

That is, we might say that a speaker who chooses to provide direct evidence when his or her intention would be sufficient has had their communicative behavior modified by what has been called "testimonial injustice" (Fricker 2007)—when the interpreter's "theory" (Davidson 2006) "causes a hearer to give a deflated level of credibility to a speaker's word" (Fricker 2007: 1). Sperber and Wilson's framework spells out of manifestness as an explicitly *epistemic* notion. This understanding of what we aim to achieve when expressing determinate content pushes us to consider the social factors that shape how a speaker would go about achieving their intended result. These social factors affect where an act will fall on the meaning-showing continuum.

8. *Expression and affect*

I began this paper by considering examples including 'I'm fine, Roy' and 'I won't swallow that'. Are we now in a position to resolve any of these confounding questions related to these utterances that I posed at the start? First it will helpful to map Sperber and Wilson's notions of making manifest and sharing an impression onto the 3D continuum I have proposed.

What seems to be special about Marie uttering "I won't swallow that" and Mary uttering "I'm fine, Roy" is that both speakers seem to be showing or *revealing* an emotional state that they are not aware of—in the case of Marie in her therapist's office—or may be aware of but suppressing—in the case of Mary speaking to Roy. The propositional content Mary utters with "I'm fine" contradicts with what she shows by uttering "Roy". To address what is going on such cases—and why they might be special—we must engage with work on consciousness and emotions from psychology and philosophy of mind.

There are a wide range of positions on the relationship between emotions, consciousness, affect, and utterances. Wittgenstein writes that if humans did not show outward signs of pain such as groaning or grimacing "it would be impossible to teach a child the use of the word 'tooth-ache'" (Wittgenstein 1958: 257). This account makes central the ways we *show* some of our internal bodily states.

This showing of external bodily states plays an important role in how we make hypotheses about the mental life of others. A psychotherapist may, for instance, make the assessment that a patient is in denial if they are laughing while describing the death of a parent (Jewett 1982).

Some theories of emotion place the subjective affective phenomenol-

ogy—not its visible manifestation—at the center. Jesse Prinz has argued that emotions are what he calls "embodied appraisals" and that all emotions "potentially occur with feelings of bodily changes" (Prinz 2006: 91). He is also explicit to note that on his view "all emotions can be conscious" (Prinz 2006: 91) but does not claim that all emotions *must be* conscious all the time (Prinz 2006: 201–202). Others defend the cognitive view of emotions—that to be in a mental state such as fear, is to be consciously experiencing a perceived danger (LeDoux 2017: 303).

There is a wide variety of viewpoints on whether or not emotions must be conscious. Thus, to explain the Mary and Marie cases in terms of emotions would be to muddy the waters with a number of theoretical commitments on the very point we would like to clarity. We can instead talk in terms of affect, which "can designate the whole subject matter we are discussing here: emotions, moods, feelings" (Damasio 2000: 342). Such a move is an attempt to be agnostic as to the details of the theoretical commitments made by the philosophers of mind I appeal to here.

We must, however, be conscious of an affective state for us to verbally state as much. Philosopher of mind David Rosenthal (2006) writes,

> Suppose I am angry at you for doing a certain thing. If my anger is conscious, I might explicitly report the anger, by saying 'I'm angry with you.' Or I might express my anger nonverbally, say, by some facial expression or body language. … when I nonverbally express my anger, the anger may or may not be conscious ...when I say 'I am angry' I report my anger; I do not verbally express it.. (316)

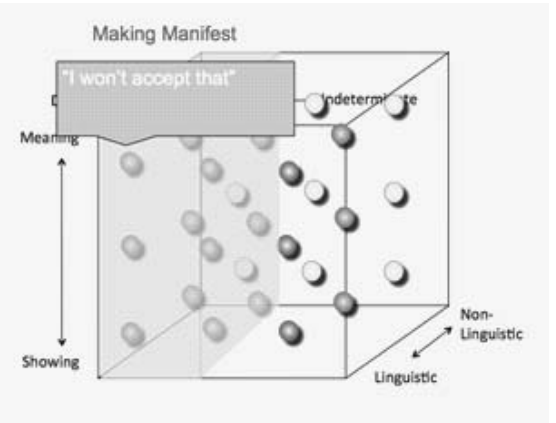
Rosenthal here introduces a distinction between "reporting" affective states and "expressing" affective states. Reporting an affective state requires awareness of that state, where expressing that affective state does not require awareness of it.

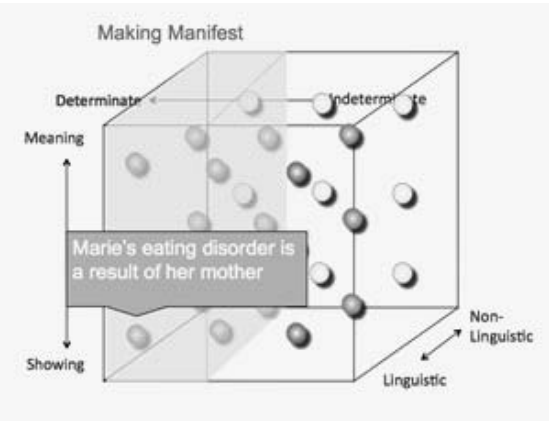
Perhaps we have a similar distinction that can be made between the sorts of contents that are meant and those that are shown. We might extend Rosenthal's account and conclude that although things that are meant must be conscious, those that are shown need not be.

Such a move would, however, require a reply to the sorts of questions I posed about the nature of showing earlier. If showing need always be intentional this move could not be made.

However, if such a move could be made, it could be brought back to deal with cases such as Marie and her mother. We could say that Marie *showed* that she ties her mother's being overbearing to her eating disorder, consciously or subconsciously, but not that she *meant* this by her utterance. (The linguistic vs. non-linguistic addition I suggested clears the way for this; otherwise we cannot have linguistic showing).

We can map these two contents onto the SCA as follows:





The bottom example is an instance of revealing an unconscious state. On this proposed model and understanding of showing, unconscious states may only be shown, and not meant. You cannot mean something you are unaware of meaning. You can show something you are unaware of showing.

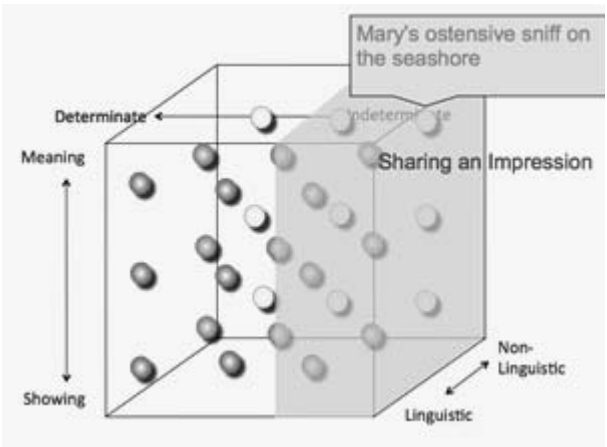
Expression of affective states is difficult to suppress (Argyle 1975: 111–112; Damasio 2000). Neuroscientist Antonio Damasio writes, "We are about as effective at stopping an emotion as we are at preventing a sneeze. We can try to prevent the expression of an emotion and we may succeed in part but not in full" (Damasio 2000: 49). If we think of certain utterances or parts of utterances as difficult to stop as sneezes, then they clearly do not follow the presumption of relevance.

Because of this the explanation for why we produce language that reveals affective states should be understood to be different from language that is costly. To put it in Sperber and Wilson's terms: the revealing of affective states does not seem to follow the presumption of relevance.

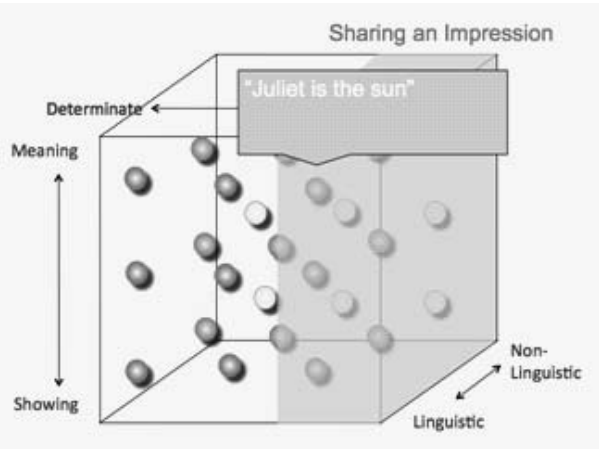
Not all utterances—or all parts of all utterances—are produced because of the intended effect on the hearer. Some of them are driven by affect. This awareness allows us to say something constructive about Mary and Marie from cases D and E. It also is the key piece to explaining utterances that otherwise have no clear intended effect—such as ranting about a bad day or a recent comment by the president. Utterances of this sort have their genesis more in the resulting effects on the speaker—not on the hearer—although a speaker may or may not be consciously aware of this. The continua of the SCA have provided the framework for the discussion of such complex utterances.

9. *Metaphor*

For my final section I will return to consider the case of metaphor I posed at the start and see how it can be treated within the SCA. Sperber and Wilson's original (1986) example of sharing an impression was Mary sniffing ostensively on the seashore. This was an instance of indeterminate meaning, and it is non-linguistic, and so will be slid back on the third proposed plane of the SCA.



'Juliet is the sun', an instance of linguistic indeterminate meaning, would fall into the following space on the proposed continuua:



An utterance such as 'Juliet is the sun' is an instance of sharing an impression because it cannot be paraphrased without loss of meaning—Sperber and Wilson's "test" for indeterminate content, borrowed from "the Romantics" (Wilson 2011).

Although I advocate for the modified Sperber and Wilson framework and believe it is a powerful tool that can be used to helpfully map and analyze utterance types, this does not presuppose the Sperber and Wilson account of metaphor. Metaphors such as 'Juliet is the sun' have been seen as problems because on the Gricean account, for a speaker to have a meaning intention a speaker must have a complex three-pronged intention with respect to the response r they intend the speaker to have. With metaphor it is hard to imagine what this would look like.

Metaphors have been raised as a problem on this view because it seems improbable that a speaker who utters a metaphor has an intention that includes *all* the meanings we would want to say are expressed by a metaphor. As I posed rhetorically at the start, if a speaker does *not* have such an intention, on the Gricean view, in virtue of what can we say that an utterance containing a metaphor has such meaning?

Griceans have responded to the apparent quandary presented by the complexity of metaphors by 1) weakening the requisite intentions, 2) oversimplifying their account of metaphors, or 3) positing dubious mental contents. This leaves one with the impression that there is some problematic ad hoc shifting taking place. Grice himself recognized the apparent problem for his view writing that some utterances may be understood as expressing an open disjunction of propositions (Grice 1989: 40; 120). However, this seems to pose a problem for what sort of mental state this would require on the part of the speaker.

Sperber and Wilson's account of metaphor is idiosyncratic in its own way. On the relevance theoretic picture, metaphors are on a continuum with hyperbole (Wilson 2011). Deirdre Wilson writes that 'John is a giant' "would count as hyperbole if taken to mean that John is very tall for a human" and would count as a metaphor "if taken to mean

that John stands out for other reasons than simply his height" (Wilson 2011: 181). If it seems that the Sperber and Wilson account reduces metaphor to nothing special, that is because it does—explicitly. Sperber and Wilson embrace this, writing in their "Deflationary Account of Metaphors" that "there is no specific mechanism to metaphor, no interesting generalization that applies only to them" (Sperber and Wilson 2008: 84).

Instead, the relevance theory account of metaphor posits encyclopedic entries for concepts such as "giant" and "sleep" (Sperber and Wilson 2004; Wilson 2011). To interpret a metaphor is to choose amongst these encyclopedic entries. For example, to interpret a metaphor such as 'The audience slept through the lecture' involves choosing between sleep meaning to "a. become mentally disengaged, b. lose interest in one's surroundings c. become motionless and unresponsive, d. gradually lose consciousness, e. undergo physical changes (snoring, slowed heart-rate, deep breathing, etc.)" (Wilson 2011: 188). It is not clear how this "encyclopedia entry" would come to be a part of a hearer's mind, how discrete these categories are, or how it could work for all metaphors containing 'sleep', including novel ones.

Such accounts of metaphor fail to account for much of the richness of metaphor—albeit willingly on the part of Sperber and Wilson. Other accounts treat metaphor as something special and may seem more satisfying because of this. For instance, in the work of Dick Moran (1989) metaphors are special in virtue of their "framing effects". According to Moran, when we encounter a metaphor such as 'Jack is a refrigerator', we cannot help but conger up a mental picture that frames Jack in some way as a refrigerator. On the Moran view, these mental effects are akin to the way we can shift to "see an aspect"—viewing Wittgenstein's duck-rabbit as a duck or a rabbit (Moran 1989: 89). To hear 'Jack is a refrigerator' is to shift from viewing Jack as an ordinary man to "see an aspect" of him in some way as a refrigerator. As Elizabeth Camp (2017) has pointed out in later work on metaphors as insults, these framing effects may be the reason such statements are not fully cancellable.

To understand metaphor in terms of sharing an impression, on the right side of the SCA, is not to come down in favor of one theory of metaphor or another.

It is not clear whether or not Sperber and Wilson are attempting to revisit and revise their previously presented account of metaphor when they present it as sharing an impression. Based on what they argued in their 2008 and 2011 "deflationary" accounts of metaphor it is difficult to see how metaphor is an instance of indeterminate content on their view. For, as argued by Wilson (2011) with the 'sleep' example metaphor *does* have determinate content, and we use the presumption of relevance to pick that content out from a finite number of encyclopedic entries. On its face this view of metaphor is quite different from sharing an impression which does not have determinate content' after

'quite different from sharing an impression'. This may be an inconsistency in what Sperber and Wilson have said about metaphor across different papers, or a misconstrual to be ironed out. If it is an inconsistency we can stick with what they argue in the 2015 paper and perhaps understand metaphor in terms of an account that more resembles Moran's framing.

Part of the apparent problem that metaphor presents for the Gricean seems to disappear when we stop seeing metaphor as expressing a range of propositions, and see it instead in terms of framing effects. It seems much more plausible that a speaker could have an intention to frame something in a way—Jack as a refrigerator—than that this speaker has a range of propositions in mind. To invite a hearer to picture this frame presents a nice parallel with Mary sniffing ostensively at the seashore.

Either way, although metaphor-qua-problem-case-for-Grice tends to be clustered in a certain part of the cube (top right), it is clear that the degree to which some intention is conscious is *distinct* from the meaning-showing, determinate-indeterminate, or linguistic-nonlinguistic continuua. Seeing this can allow us to disentangle questions about the degree to which some intention is conscious from where the corresponding utterance falls within the proposed quality space.

10. *Conclusion*

A full account of our communicative practices will be mindful of what these distinctions mean for

1) our theories of meaning and
2) our explanations of why we engage in certain communicative acts, including showing.

The ability to handle a wide range of cases is a strength of the SCA. Sperber and Wilson's work shows the power of applying philosophy of language grounded in Grice to an array of cases, and their 2015 framework—which I hope to have strengthened—has great potential for theorizing not just about language, but about meaning-making, and the conscious and unconscious things we show, in general. Such work allows us to ask not only *how* but *why* we engage in certain forms of communicative behavior, and captures the incredible nuance of human interactions: said and meant, linguistic and non-linguistic, determinate and indeterminate.

References

Argyle, M. 1975. *Bodily Communication. International* University Press.

Bezuidenhout, A. 2001. "Metaphor and What is Said: A Defense of a Direct Expression View of Metaphor." *Midwest Studies in Philosophy* 25: 156–186.

Camp, E. 2017. "Why Metaphors Make Good Insults: Perspectives, Presupposition, and Pragmatics." *Philosophical Studies* 174 (1): 47–64.

Davidson, D. 2006. "A Nice Derangement of Epitaphs." In E. Lepore and K. Ludvig (eds.). *The Essential Davidson.* Oxford: Oxford University Press.

Damasio, A. 2000. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness.* New York: Mariner Books.

Fricker, M. 2006. *Epistemic Injustice.* Oxford: Oxford University Press.

Grice, H. P. 1989. *Studies in the Way of Words.* Oxford: Oxford University Press.

Hornsby, J. 2000. "Feminism in Philosophy of Language." In M. Fricker and J. Hornsby (eds.). *The Cambridge Companion to Feminism in Philosophy.* Cambridge: Cambridge University Press: 87–106.

Johnson, M. 2016. "Cooperation with Multiple Audiences." *Croatian Journal of Philosophy* 16 (47): 203–227.

Jewett, C. 1982. *Helping Children Cope with Separation and Loss.* Boston: The Harvard Common Press.

Kecskes, I. 2014. *Intercultural Pragmatics.* Oxford: Oxford University Press.

_____, 2016. "Intracultural Communication and Intercultural Communication: Are They Different?" Lecture. June 10, 2016. University of Split, Croatia. *7th International Conference on Intercultural Pragmatics and Communication.*

LeDoux, J. 2017. "Semantics, Surplus Meaning, and the Science of Fear." *Trends in Cognitive Science* 21 (5): 303–306.

Moran, R. 1989. "Seeing and Believing: Metaphor, Image, and Force." *Critical Inquiry* 16 (1): 87–112.

Prinz, J. 2006. *Gut Reactions: A Perceptual Theory of Emotion.* Oxford: Oxford University Press.

Rosenthal, D. 2006. *Consciousness and Mind.* Oxford: Oxford University Press.

Solomon, R. 2001. *True to Our Feelings: What Our Emotions Are Really Telling Us.* Oxford: Oxford University Press.

Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition.* Oxford: Blackwell Publishing.

_____, 2004. "Relevance Theory." In L. Horn and G. Ward (eds.). *The Handbook of Pragmatics.* Oxford: Blackwell: 607–632.

_____, 2015. "Beyond Speaker's Meaning." *Croatian Journal of Philosophy* 15 (44): 117–149.

Wilson, D. 2011. "Parallels and Differences in the treatment of metaphor in relevance theory and cognitive linguistics." *Intercultural Pragmatics* 8: 177–196.

Wittgenstein, L. 1958. *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Basil Blackwell.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

The Problem of First-Person Aboutness

JESSICA PEPP*
Uppsala University, Uppsala, Sweden

The topic of this paper is the question of in virtue of what first-person thoughts are about what they are about. I focus on a dilemma arising from this question. On the one hand, approaches to answering this question that promise to be satisfying seem doomed to be inconsistent with the seeming truism that first-person thought is always about the thinker of the thought. But on the other hand, ensuring consistency with that truism seems doomed to make any answer to the question unsatisfying. Contrary to a careful and enticing recent effort to both sharpen and escape this dilemma by Daniel Morgan, I will argue that the dilemma remains pressing both for broadly epistemic and broadly causal-acquaintance-based accounts of the aboutness of first-person thought.

Keywords: First-person thought, aboutness, reference determination, acquaintance, introspection.

1. *Introduction*

The topic of this paper is the question of in virtue of what first-person thoughts refer to what they do (or, for those who prefer not to use 'refer' when speaking about thoughts, in virtue of what first-person thoughts are about what they are about). This can seem like an odd question, because it seems so obvious that first-person thoughts are about the person thinking them. Being about the thinker of the thought seems to be part of what it is to be a first-person thought. But its being obvious that first-person thought is always about the person thinking it does not make it obvious in virtue of what this is the case. (Any more than its being obvious that some uses of "Aristotle" refer to a certain

* Acknowledgments: For helpful comments on earlier versions of this work, I am grateful to audience members at the Dubrovnik Workshop on Philosophy of Language, the Higher Seminar in Theoretical Philosophy at Lund University, and the Perspective in Language and Thought workshop at Uppsala University.

philosopher makes it obvious in virtue of what this is the case.) Let us call the question of in virtue of what a first-person thought refers to what it refers to (or is about what it is about) the *Question of First-Person Aboutness*. My focus here will be on a dilemma arising from this question. On the one hand, approaches to answering this question that promise to be satisfying seem doomed to be inconsistent with the aforementioned truism that first-person thought is always about the thinker of the thought. But on the other hand, ensuring consistency with that truism seems doomed to make any answer to the Question of First-Person Aboutness unsatisfying.

Contrary to a careful and enticing recent effort to both sharpen and escape this dilemma by Daniel Morgan (2015), I will argue that the dilemma resists this effort. At present, I do not see a good way of escaping it. I find this troubling in part because I am drawn to a certain general view of what it is for things (at any rate, concrete things, human beings included) to be genuinely thought about, or "in" our thoughts. This view holds that one requirement is that the thinker has a perception-based form of acquaintance with the things thought about. This view in turn requires an answer to the Question of First-Person Aboutness that seems certain to be speared on the first horn of the dilemma. The type of answer required by the acquaintance view of reference determination is different from the type of answer that Morgan defends against the dilemma. But they face similar challenges stemming from the idea that first-person thought is guaranteed to be about the person thinking it. Moreover, it is not clear that either view can overcome these challenges. My goal here is to show the power of the dilemma by homing in on the types of cases that make the horn of violating the guarantee of reflexive reference the sharpest. It is hard to see how a satisfying answer to the Question of First-Person Aboutness can avoid this problem. But giving up on the search for a satisfying answer is also unappealing. I conclude that the dilemma remains, with the acquaintance view of reference-determination as one of its hostages.

The plan for the paper is as follows. In section 2, I will set out the general acquaintance-based view of what it is for things to be in thought that I favor, and that I take to be threatened by the case of first-person thought. In section 3, I will explain how first-person thought threatens this view, as well as broadly epistemic views about what determines the reference of first-person thought. In the central part of the paper, section 4, I will refine the challenge from first-person thought, taking account of Morgan's effort to disarm the challenge to broadly epistemic views of reference determination, and showing its persistence for both acquaintance-based and broadly epistemic views. In section 5, I will consider François Recanati's answer to the Question of First-Person Aboutness, arguing that it falls on the second horn of the dilemma articulated in the opening paragraph. Section 6 concludes.

2. *Extended acquaintance requirements on being in thought*

I will start by setting out the form of acquaintance requirement on a thing's being thought about that I favor. When it comes to *thinking about something,* or *something's being thought about* or *being in thought,* we can make a broad distinction between a more and less substantial form of this relation. For instance, there seems to be a difference between thinking that there is a unique individual who is the oldest human currently living and thinking about Kane Tanaka, in particular. In the first case, one is, in some sense, thinking about Kane Tanaka, since she is (at the time of writing, and trusting Wikipedia) the oldest human currently living. But this seems quite different from thinking about Kane Tanaka because you are looking at her, or talking to her, or remembering her. In the latter cases, Kane Tanaka in particular seems to figure in your thought in a more substantial way than in the former.

It is not easy to make this intuitive difference precise. Bertrand Russell used a notion of "acquaintance" to do so. He argued that for a thing to be in one's thought in the second, more substantial way, one had to be "acquainted" with it. On one ordinary understanding of "being acquainted" with something, there is something intuitive about this requirement, since one need not be acquainted (in that ordinary sense) with Kane Tanaka to think that there is a unique oldest currently living human, or to go on to think things about whoever satisfies that condition. But to think about her as a result of seeing her or remembering her, one would need to be (in the ordinary sense) acquainted with her. Russell, of course, was not using this ordinary notion, but understood acquaintance as such an unimpeachable epistemic relation that even those who see and talk to Kane Tanaka are not acquainted with her and can think about her only in the same insubstantial way as those who consider that there is a unique oldest currently living human (i.e., they can think of her only "by description") (Russell 1905 and 1910–11).

Nonetheless, a tradition stemming from Russell retains the idea that a useful line can be drawn between the more and less substantial ways for a thing to be in thought by appeal to a (less epistemically demanding) notion of acquaintance. A common denominator for this tradition, articulated by Robin Jeshion, is that the relevant kind of acquaintance must satisfy the following condition, which she calls the "Standard-Standard on Acquaintance": "One can be acquainted with an object O only by perception, memory, and communication chains" (Jeshion 2010: 109).[1] To spell this out a little more, we can enumerate the kinds of events or episodes that can give a thinker acquain-

[1] The tradition endorsing an acquaintance requirement on the substantial way of having a thing in thought, and understanding acquaintance in accord with the Standard-Standard, has been called "the extended acquaintance tradition" (Dickie 2016) and "causal acquaintance" (Hawthorne and Manley 2012).

tance with an object, according to the Standard-Standard. These are: (i) instances of perceiving the object, (ii) instances of being referred to the object[2] via a communication chain originating in someone's perception of the object and (iii) instances of remembering the object, where the memory derives either from one's past perceptions of the object or from one's past uptake of the object via being referred to it. I will call these *S-acquaintance instances* (the "S-" is to signify that they meet the Standard-Standard).

In an earlier paper I distinguished two ways of using the Standard-Standard notion of acquaintance to articulate a requirement on having a thing in thought in the more substantial way just alluded to (Pepp 2019). These two ways differ with respect to how they spell out what this "more substantial" way of having a thing in thought amounts to. On the first kind of acquaintance requirement, the "more substantial" way is for the thing to *figure in the content* of one's thought, and acquaintance with the thing is required for this. On the second kind of acquaintance requirement, the "more substantial" way is for the thing to be in one's thought *in a non-satisfactional way* (i.e., for it to be in one's thought not in virtue of satisfying a condition that is also in thought), and acquaintance with the thing is required for this.

A requirement of the second kind is:

> *Non-satisfactional Acquaintance Requirement (NAR):*
>
> For a concrete object to be thought about in a non-satisfactional way, it must be thought about partly in virtue of one or more S-acquaintance instances.

Note that this requirement is restricted to *concrete* (as opposed to abstract) objects. The restriction sets to one side the challenge that abstract objects cannot be perceived, so if they can be in thought in a non-satisfactional way, then the acquaintance requirement fails. This is a serious and interesting challenge to a general acquaintance requirement on being thought about in a non-satisfactional way. Addressing it requires taking up the broader question of how it is possible to think and know about abstract objects if they are causally inefficacious. This paper leaves that question for another time and focuses on whether an acquaintance requirement on being in thought non-satisfactionally is defensible *even for concrete objects*.

Another thing to note about NAR is that it not only requires that S-acquaintance instances coincide with non-satisfactional thought about concrete objects, but that it is *in virtue of* these S-acquaintance instances that concrete objects are thought about in a non-satisfactional way (cf. Jeshion 2010: 69). The idea behind NAR is that S-acquaintance is part of the mechanism of reference for thoughts that are non-satisfactionally about concrete objects. Part of what binds these thoughts to the

[2] By "being referred to the object," I mean roughly what Bach (2008) means by it: in understanding someone's use of a word to refer to an object, one is *referred* to that object.

objects they are about is the connection of these thoughts to thinkers' perceiving, being referred to, or remembering the objects. Of course, there are different views about what sort of connection is required so as to spell out the full mechanism. All that NAR requires is that S-acquaintance figures in that story.

3. The challenge from first-person thought

NAR is a plausible principle. Many of the kinds of examples that have been brought to bear against acquaintance requirements on "singular thought" are cases in which an object is clearly being thought about in virtue of satisfying a condition invoked in thought, and yet it is claimed to be intuitive that the content of the thought is singular with respect to that object—it is that object, not the condition that brings it into thought, which figures in the content of the thought (see Pepp 2019 for discussion). Since these are clearly not cases of objects being thought about non-satisfactionally, they do not threaten NAR. Instead, NAR is threatened by cases in which it seems that a thought is about a concrete object both non-satisfactionally and not in virtue of S-acquaintance with the object. In the aforementioned paper I identified two such challenges: one based on cases, adduced by David Kaplan and Imogen Dickie, in which a thinker seems to think non-satisfactionally of an object in virtue of perceiving *evidence* of the object but not the object itself (see Kaplan 2012: 144 and Dickie 2016: chapter 6); the other based on first-person thought. My focus in this paper is the challenge to NAR from first-person thought. In particular, I aim to get clear on what the heart of the challenge is.

3.1. *The structure of the challenge to NAR*

The challenge to NAR from first-person thought is based on the following three claims:

1. First-person thoughts are about concrete objects (i.e., human beings).
2. First-person thoughts are not about particular human beings satisfactionally, i.e. in virtue of those human beings satisfying conditions that figure in the thoughts.
3. First-person thoughts are not about particular human beings in virtue of the thinkers' S-acquaintance with those human beings.

If all three of these claims are true, then first-person thoughts are counterexamples to NAR. By "first-person thoughts," I refer to the kind of thoughts we express in language using the grammatical first person. (This is not a definition, but only a way of pointing you to the thoughts in question.) Examples include my occurrent thought that I am tired or my standing belief that I was born in Boston.

I am inclined to accept claims 1 and 2. Concerning 1, it is compelling that my first-person thoughts are about the same thing that certain

third-person thoughts on the part of other thinkers are about: namely *me*, a certain human being. Whatever I, this human being, ultimately *am* metaphysically speaking, I also find it compelling that this individual is just as perceptible as, say, tables and chairs, and hence is concrete enough for purposes of the present discussion.

To reject 2, one might, in a Russellian vein, argue that first-person thoughts are about particular human beings in virtue of those human beings satisfying conditions that figure in a thinker's thought such as the condition of *being the person experiencing this*, where "this" anchors the condition in a particular mental episode (or other unit of mentation). I am not drawn to such an approach, for two reasons. First, it is not clear why thought about experiencing a mental particular would be prior to first-person thought. Indeed, it seems more likely that the idea or concept of someone's experiencing a mental particular would be derived from first-person thought about one's own experiences, at least ontogentically. Second, even if first-person thought is satisfactional in this way, the reduction relies on mental items being thought about non-satisfactionally. Thus, a defender of NAR will face the task of arguing that we are S-acquainted with our own mental episodes, which arguably would be a way of being S-acquainted with ourselves.[3]

Claim 3, it seems to me, is the most promising of the three claims for a defender of NAR to reject. It is a (negative) partial answer to the Question of First-Person Aboutness. Thus, to reject Claim 3 would be to defend a broad positive answer to that question: namely, that first-person thoughts are about what they are about partly in virtue of being based upon a perceptual, memory, or communicative connection of the right sort to the object. "Connection of the right sort" is a placeholder: what counts as the right sort of connection could be specified in different ways. One approach would be to fill the placeholder with some sort of epistemic restriction, so that only connections that provide epistemic benefits (such as enabling the thinker to gain knowledge or true beliefs about the object) qualify.[4] A different approach would be to fill the placeholder with a less epistemically loaded restriction, perhaps requiring the connection to be information-carrying only in the sense that it allows the object to make some sort of cognitive impact on the thinker, whether or not this enables epistemic advance.[5] Let us call this broad view of reference determination for thoughts—however "the right sort of connection" is ultimately spelled out—*Aboutness through S-acquaintance*. Another broad kind of view of reference determina-

[3] Alternatively, in a Fregean vein, one might argue that first-person thoughts are about particular human beings in virtue of those human beings satisfying conditions imposed by private, primitive self-concepts that are present in thought. The problem here is that it is not at all clear what sort of conditions these would be.

[4] François Recanati (2012) calls such relations "epistemically rewarding." (I will return to Recanati's own answer to the Question of First-Person Aboutness in section 5 below).

[5] See Julie Wulfemeyer (2017) for development of such a view.

tion for thoughts (which might overlap with Aboutness through S-acquaintance) is what I will call *Aboutness through Epistemic Gain*. On this kind of view, a thought is about what it is about partly in virtue of being based upon the thinker's ways of gaining knowledge (or justified true belief, or some other epistemically positive status) about that thing.

Gareth Evans famously defended (albeit tentatively) a version of Aboutness through Epistemic Gain for first-person thoughts.[6] But familiar problems for the view can make it seem hopeless.[7] These problems stem from hypothetical cases designed to show that if Aboutness through Epistemic Gain were correct for first-person thoughts, then some first-person thoughts would fail to be about the person thinking them. But it is a truism that first-person thoughts are always about the person thinking them, so Aboutness through Epistemic Gain cannot be right when it comes to first-person thoughts. The hypothetical cases used to support this argument may be divided into two types, what I will call *absences cases* and *diversion cases*. They serve equally well as problems for Aboutness through S-acquaintance as they do for Aboutness through Epistemic Gain. Thus it is useful for someone like me, who is inclined to defend NAR, to consider the implications of these cases for S-acquaintance-based views.

Before I introduce the two types of case, it is worth a brief glance at the most prominent alternative to Aboutness through Epistemic Gain (or to Aboutness through S-acquaintance) for the case of first-person thought. I will call this alternative *First-Person Aboutness by Reflexive Rule*. This is the view that first-person thoughts are about what they are about in virtue of being governed by the rule that they refer to whoever thinks them. This view is not threatened by the kinds of cases I am about to describe. But, as Morgan convincingly argues, nor does it provide a satisfying answer to the Question of First-Person Aboutness (see Morgan 2015: 1801–1802). This is because it is not clear in what sense first-person thoughts are governed by such a rule. If to say that they are governed by this rule is just to say that, in fact, first-person thoughts always refer to the one who thinks them, then the Reflexive Rule view *describes* the reference of first-person thoughts but is silent about what *determines* that reference—i.e., about *in virtue of what* first-person thoughts always refer to the one who thinks them. In other words, the Reflexive Rule view on this interpretation does not answer the Question of First-Person Aboutness at all. On the other hand, understood as an answer to this question, the Reflexive Rule view seems false. It might be accepted that rules (i.e., the conventions of language) make it the case that uses of the pronoun "I" refer to the one who uses the word, but no such conventional rules govern thoughts.[8] Thus, if the

[6] In Chapter 7 of *Varieties of Reference* (1982).

[7] John Campbell (1994) and Lucy O'Brien (2007) lay out these problems.

[8] Morgan also considers whether Peacocke's (2008) view should be seen as

cases I am about to describe tempt one to resort to Aboutness by Reflexive Rule, it should be held in mind that this is tantamount to giving up on answering the Question of First-Person Aboutness. This is the second horn of the dilemma of which the problems about to be raised are the first horn.

3.2. *Absence and diversion cases*

Now to the cases. First let us consider absence cases. Elizabeth Anscombe described the following scenario:

> And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other. Now I tell myself 'I won't let this happen again!' (Anscombe 1981)

As this is a "sensory deprivation" scenario, the subject is presumably not receiving information about herself either via external senses or via bodily senses (e.g. proprioception, kinaesthesia, nociception). It seems that having one's external senses cut off would also entail that one is not being referred linguistically to oneself via a communication chain originating in someone's perception of one. Evans added to Anscombe's scenario the possibility that the person could also have amnesia and thus not be receiving any information about herself via memory (Evans 1982: 215). This seems to leave no instances of S-acquaintance for a first-person thought to be based on. Nonetheless, says Evans, the person in this scenario "may still be able to think about himself, wondering, for example, why he is *not* receiving information in the usual ways." If there is first-person thought about oneself in the absence of any S-acquaintance with oneself on which the thought could plausibly be based, this is a problem for Aboutness through S-acquaintance and for NAR. The case also seems like a problem for Aboutness through Epistemic Gain, since the subject's ways of gaining knowledge of herself are disabled.

Next let us consider diversion cases. David Armstrong suggested the following scenario:

> We can conceive of being directly hooked-up, say by transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movements of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs. (Armstrong 1984: 113)

In this case, the subject is receiving information via proprioception—a likely kind of perception to determine the reference of first-person thoughts, given that it is a sense dedicated to perceiving the perceiver. But the information she receives is not about herself, but about some-

treating first-person thought as having its reference determined by a primitive rule of reflexive reference (as opposed to merely being correctly described as always referring to the thinker). He concludes that this is not clear, and that if the rule is treated by Peacocke as a primitive determiner of reference, it is not clear why we should accept this.

one else. If she then has the first-person thought, based on this information, that she is walking, and if the reference of this thought is determined partly by the perceptual connection on which it is based, it would seem that this first-person thought should be about the other person. Or, at least, there should be some uncertainty regarding whom it is about. But this seems wrong: the thought is about the person thinking it. This suggests that the thought's connection to the thinker's perception and memory, or to her ways of gaining knowledge, is not part of what makes the thought be about what it is about. This calls Aboutness through S-acquaintance, together with NAR, and Aboutness through Epistemic Gain into question.

4. *Refining the challenge from first-person thought*

Morgan mounts a strong defense of Aboutness through Epistemic Gain, arguing that absence cases are not as much of a problem as they appear to be, while diversion cases are harder to deal with but still leave various options open for defenders of Aboutness through Epistemic Gain. My own investigations in this section will suggest that the situation is the opposite, both for Epistemic Gain theorists and for S-acquaintance theorists. Diversion cases can be handled, while absence cases, properly described, show the core of the problem posed by first-person thought for these two kinds of view about reference determination.

4.1 *Absence cases*

I will begin with absence cases. It is notable that in presenting their scenarios of sensory and memory deprivation, both Anscombe and Evans describe the subject's first-person thought as a reaction to her situation. Anscombe's subject thinks (what she might express in language as) "I won't let this happen again." Evans's subject wonders "why he is *not* receiving information in the usual ways." Calling attention to the reaction that a subject would have to finding herself in a deprivation scenario makes it intuitive that someone in such a situation would, and *a fortiori* could, have first-person thoughts. But it should also lead us to question whether the cases described by Anscombe and Evans are really cases in which *all* perceptual, memory, and communicative connections of the right sort to determine the aboutness of first-person thoughts are absent. For if the first-person thoughts that a subject would have in these scenarios are reactions to her situation, then this suggests that she is somehow aware of, or receiving information about, how things are with herself.

In this vein, Morgan suggests that the subjects in these scenarios remain able to introspect—to "rely on [their] direct way of gaining knowledge of [their] own mental properties" (Morgan 2015: 1804). It is plausible that their first-person thoughts are based upon this way of gaining knowledge of themselves. For the purposes of defending About-

ness through S-acquaintance (and thereby defending NAR), one would have to argue that the kind of introspection on which these thoughts are based is plausibly perceptual, or at least enough like perception that it does not violate the spirit of S-acquaintance.[9] It seems to me that this condition will be met if the kind of introspection on which these thoughts are based is a means of detecting pre-existing mental states, properties or events.[10] There is reason to suppose that the first-person thoughts we imagine people having in these scenarios would indeed be based upon the detection of a prior mental property or condition: the property or condition of *not* being perceptually aware in the usual ways and *not* having memories.

Of course, these scenarios are not really so easy to imagine, and it is not entirely clear what we are supposed to imagine, especially concerning the subject's loss of memory. Are we to imagine her lacking all forms of memory—episodic memory, semantic memory, working memory, procedural memory and so on—or only some sub-class of these? I am not sure what the mental life of someone lacking all of these would or could be like, including whether or not they would or could have first-person thoughts (or any thoughts at all), especially when also deprived of all perceptual stimulus. But to make the best case against NAR, one might stipulate that the subject lacks all memory that counts as S-acquaintance with herself. (This would mean that her first-person thoughts could not satisfy NAR by their connection to memory instances of S-acquaintance with herself.) So, she might still remember things such as that Paris is the capital of France, or how to ride a bike. If so, then it might be suggested that she could have what Tyler Burge calls "cogito-like thoughts" (Burge 1988). These are thoughts in which one thinks a thought as part of the act of self-ascribing that thought. For example, the subject might think that she is thinking that Paris is the capital of France, where the thinking that Paris is the capital of France occurs as part of the thinking that she is thinking this. This would be a first-person thought, but it would seem to involve no detection of a prior mental condition of thinking that Paris is the capital of France.

Given the case as we have now described it, it strikes me as unclear whether the subject could have cogito-like thoughts. It is not so much that anything obviously prevents it. However, in considering what might prompt such thoughts in a subject with no perception of herself,

[9] This is not critical for Morgan's purposes, since he is defending a version of Aboutness through Epistemic Gain: the view that the reference of first-person thoughts is determined by our "ways of gaining knowledge of ourselves." These ways of gaining knowledge need not be exhausted by instances of S-acquaintance. He notes that even if "it is wrong to think of introspection as a faculty that is just like vision, except that it is trained on the mind," introspection could still be appealed to as a way of gaining knowledge of ourselves (2015: 1805). However, as we will see in the discussion below, it will not help Morgan's defense to appeal to forms of introspection that are radically dissimilar to S-acquaintance.

[10] Eric Schwitzgebel (2016) calls this the "detection condition" on introspection.

including no detection of her own pre-existing mental states, and no memories of such perception or detection, it is difficult for me, at least, to have a firm intuition that such thoughts could come about in her.

Nonetheless, it would be simple to modify the case further so as to remove this uncertainty. We can stipulate that not only does the subject lack all forms of S-acquaintance with herself, but she is also being artificially stimulated in such a way as to produce perceptual and introspective experience of the usual kind. In other words, we can make the case more like a Matrix or Cartesian evil genius scenario, although we specify that the subject is not even able to detect the mental states induced in her (anything she seems to detect in this way is fabricated). Let us call this an *illusory absence case*. A subject in such a situation presumably could have a cogito-like thought. This is a first-person thought. Intuitively, it is about the one who thinks it. But it cannot be about her in virtue of her S-acquaintance with herself, because she lacks any such S-acquaintance.

Illusory absence cases seem to me to provide the core challenge to Aboutness through S-acquaintance and NAR. Do they challenge Aboutness through Epistemic Gain to the same extent? It might seem that Aboutness through Epistemic Gain is on better footing here, since it seems able to admit cogito-like thoughts as ways of gaining knowledge about oneself. If cogito-like thoughts are ways of gaining knowledge about oneself, then these thoughts can serve as determiners of the reference of first-person thoughts on an Epistemic Gain view. By contrast, the advocate of Aboutness through S-acquaintance cannot appeal to cogito-like thoughts as determiners of reference, since they do not involve S-acquaintance.

But it is questionable whether the epistemic view really has an advantage here. There is an air of circularity in the claim that a thinker's first-person thought is about herself in virtue of her being the person she can gain knowledge about by thinking a thought as part of self-ascribing that thought—that is, by thinking first-personally that she is thinking that thought. To treat cogito-like thoughts as a reference-determining form of self-knowledge amounts to saying that a thinker's first-person thoughts are about her in virtue of its being *her* about whom she has first-person thoughts. So it seems to me that the defender of Aboutness through Epistemic Gain should be loath to appeal to cogito-like thoughts as a way of securing first-person reference in illusory absence scenarios. To do so is to follow the Reflexive Rule theorist in giving up on the effort to say in virtue of what first-person thoughts are about what they are about.

In sum, absence cases as they are usually described are not definitive counterexamples either to NAR or to broadly epistemic views of first-person reference. However, when they are built up into illusory absence cases, it is difficult to resist the conclusion that first-person thought can occur and be about the one who thinks it without this aboutness being in virtue of either the thinker's S-acquaintance with

herself or ways of gaining knowledge about herself. This, it seems to me, is the core problem for efforts to explain the aboutness of first-person thought in either way, and hence for NAR.

4.2 *Diversion cases*

As I mentioned above, Morgan thinks that diversion cases are more troublesome for Aboutness through Epistemic Gain than absence cases because, while absence cases can be dealt with by appeal to introspection, diversion cases require an explanation of why first-person thoughts are not about the sources of the information on which they are based (Morgan 2015: 1806). In the last section I argued that in fact absence cases cannot be effectively dealt with by appeal to introspection, neither by a defender of Aboutness through Epistemic Gain, nor by a defender of Aboutness through S-acquaintance. By contrast, I think diversion cases can be dealt with by rejecting the explanandum. That is, we do not have to explain why first-person thoughts in diversion cases are not about the sources of the information on which they are based, because it is not obvious that, in general, they are not about these sources. Instead, while it is highly intuitive that in such cases first-person thoughts *are* about the person thinking them, it is not uniformly so obvious that they are not *also* about the person from whom the information originates.

Let me first acknowledge that in a case like the one from Armstrong cited above, it seems pretty clear that in imagining the subject thinking that she is walking, or has crossed legs, we would (in Lucy O'Brien's words) "surely take it that I am thinking, probably falsely, about myself, rather than thinking truly about the person who was the source of the information."[11] This intuition seems solid about this particular case, but I think variants on the case provoke less certainty, in particular about the claim that the thinker is *not* thinking truly about the person who was the source of the information. For instance, imagine a similar case of receiving proprioceptive information from someone else's body, but imagine this happening while one is doing a mindfulness exercise. One pays close attention to (what one takes to be) the position of one's body, carefully observing (what one takes to be) the angle of one's elbow, the tension in one's wrist muscles, and so on. One has various first-person thoughts as a result of this attentive study, thinking that one's elbow is bent exactly ninety degrees, that one's wrist muscles are just tense enough to support a press-up, and so on. One makes a concerted effort to *get it right*. It seems at least somewhat plausible that although the person in this case is clearly thinking falsely about herself, she is also thinking truly about the other person (that is, if her proprioceptive judgments are accurate with respect to their body).

We might also consider a case in which the subject not only receives

[11] O'Brien (2007: 39).

proprioceptive information from another person's body, but also nociceptive information, and quasi-memory of both types of perception from that other person. Imagine that on the basis of this information she has the first-person thought that the position she has (what she takes to be) her legs in now is more painful than the one they were in the last time she sat on a chair. Here, too, it seems plausible that in some sense she is thinking truly about the person who was the source of the information, even while also thinking falsely about herself. I suspect that the more information channels are diverted to the other person in an imaginary case, the stronger will be the sense that the thinker's first-person thoughts are in some sense about that other person, in addition to being about herself.

These cases may be compared to cases involving linguistic reference that have been called cases of "partial reference" or "multiple reference".[12] In these cases, a speaker's confusion of two objects makes it plausible that she refers to both of them with her use of a name or demonstrative, and thus that she may say something true about one of them while simultaneously saying something false about the other. Michael Devitt illustrates the phenomenon using names: if I know Devitt has a cat called "Nana" and Devitt points out to me a Persian cat who is not Nana and says, "This is Nana," then if I later say, "Nana is a Persian," I am referring by "Nana" both to Nana, whom I heard about earlier from Devitt, and to the cat I was shown, who is not Nana. I am speaking truly of the latter, but falsely of the former. As Devitt puts it, "there is only one strong [basic intuition]: the 'total performance' involves elements of truth and falsity" (1981: 145). Susanna Siegel gives the following example (adapted from one used by Sydney Shoemaker in a different context):

> You are a salesman in a tie store. By reaching past an opaque door into a display case, you put your hand on a blue silk tie. At the same time, another salesman is reaching through the cabinet and touching a red silk tie. Through the glass top of the cabinet, you can see the red tie being held by the other salesman, whose arm looks like yours. You mistake his hand for yours and you believe that you are the one touching the red tie. You say to a customer, who was looking in another direction for a red silk tie, "This one is red." (Siegel 2002: 10–11)

Siegel points out that there are three things we might say about the use of the demonstrative 'this' in such a case: it refers to one of the ties but not the other, it fails to refer, or it refers to both ties. The advantage of the third option—the use of the demonstrative has multiple reference—is that it respects the intuition that the salesman says something true about the red tie, while also saying something false about the blue tie.

[12] The term "partial reference" comes from Hartry Field. (1973). It is developed for the case of singular terms by Michael Devitt (1981: 145ff). Susanna Siegel (2002) introduces the term "multiple reference" for the same phenomenon and applies it to demonstrative reference.

If it is right that in some diversion cases first-person thoughts are about the person the thinker receives information from *in addition to* being about the thinker, then such cases are not necessarily a problem for either Epistemic Gain or S-acquaintance views of the reference of first-person thoughts. For this is consonant with the claim that S-acquaintance, or ways of gaining knowledge, are partially determinative of that reference. Still, it might be objected that the diversion cases are importantly different from Devitt's and Siegel's multiple/partial reference cases. In particular, the latter are cases about which it might be said that a single linguistic utterance is used to express two different thoughts. In Devitt's example, the speaker intends to say both that the cat she has been told about named "Nana" is a Persian and that the cat she was shown is a Persian. In Siegel's example, the speaker intends to say both that the tie he is touching is red and that the tie he is looking at is red. But in the diversion cases, what is at issue is not a linguistic expression that might be of two thoughts but a single, first-person thought. (Likewise, such single, first-person thoughts seem to be what are expressed using first-person linguistic expressions.) What, then, could factor such a thought into one instance of thinking falsely about oneself and one instance of thinking truly about someone else?

One might appeal to the claim that these first-person thoughts are not based upon identifications to argue that no such factoring is available. Consider that the tie salesman's utterance of "This is red" arguably expresses a single, perceptual-demonstrative thought based on the speaker's overall perceptual awareness. Even if this is a single thought, it is based upon the thinker's identification of the tie he is seeing with the tie he is touching. This makes it plausible that the speaker is in some sense expressing two thoughts, one about each tie. By contrast, it is an often-noted feature of first-person judgments based on proprioception, nociception and the like that they are not based on identifications. In thinking that I am walking or have crossed knees, I do not think that the person whose body I proprioceive has crossed knees, identify that person as myself and thereby think that I have crossed knees.

This seems right, but I wonder whether even thoughts like the one in Siegel's example are generally based on identifications. No doubt some of the time one consciously judges that something one is touching is the same thing that one is seeing, but much of the time our different sensory modalities seem to be integrated without any such conscious judging. As I type, fairly quickly, but not wholly by touch, I do not consciously judge that the keys I am hitting are the ones I am seeing, I just let my hand-eye coordination do its work. Suppose my coordination is off and in some instant the key I feel is not the one I (briefly) visually attend to. Having recently spilled cola on the keyboard, I think to myself that this key will be sticky. If the key I feel is in fact not sticky but the key I see is, it seems just as reasonable a conclusion here as in the tie case that my thought is partly true—about the key I see—and part-

ly false—about the key I feel. This does not seem to rely on my having made a judgment identifying the key I feel with the key I see. Maybe it is nonetheless right that I believe that the key I see is the one I feel under my index finger, even though I do not consciously judge this. But then it does not seem less right that I believe that the person whose body I proprioceive as having a hard but movable surface under its index finger is me. This is not a conscious identification judgment that I make, but I believe it, at least in the sense that if I were now told that I am in an Armstrong-style situation I would change my belief. This does not strike me as importantly different from the sense in which I believe that the key I am touching is the one I am visually attending to.

Thus, I think we can allow that S-acquaintance with, and ways of gaining knowledge of, individuals other than oneself can be partly determinative of the reference of first-person thought.[13] It is to be expected that intuitions will vary about in which cases first-person thought is partly, or additionally, about someone other than the thinker, while also being about the thinker. We find similar variations in intuitions about in which cases a speaker is referring (in language) to more than one object. But as long as it is plausible that such multiple or partial reference can occur in first-person thought, there is not a need for defenders of the S-acquaintance and Epistemic Gain views to explain away the (supposed) fact that it does not occur.

4.3 *Total diversion cases*

Appeal to multiple or partial reference is thus a viable way of defending NAR in the face of diversion cases. One type of diversion case it does not seem to help with, however, are cases of *total diversion*, in which a subject has no S-acquaintance with herself because all of her usual ways of perceiving, remembering and detecting herself are diverted to another individual.

[13] Morgan argues that in diversion cases the diverted senses are *not* ways of gaining knowledge of the other individual. A rough way of putting his suggestion is that for the subjects in diversion cases, faculties like proprioception may be supplying information from the other individual in the purely causal, non-epistemically beneficial way mentioned in section 3.1 above. But they are not enabling knowledge about the other individual because it is not their function to provide awareness of bodies other than one's own (2015: 1806–1807). This is in the same way that vision, for example, does not function to provide awareness of the surfaces of our retinas, but of distal objects, even though in the purely causal sense our visual experiences give us information about the surfaces of our retinas. Thus, if a scientist artificially produced an image on one's retina, the resulting visual experience would not be a way for one to gain knowledge of the state of one's retina, even though it would carry information about this state. In effect, this suggestion assimilates diversion cases to absence cases, treating diverted senses in the same way as artificially stimulated ones. If this is the right thing to say on behalf of Aboutness through Epistemic Gain regarding diversion cases, then again the core problem for the view is displayed by illusory absence cases.

Imagine Beth, an ordinary human being living an ordinary life. Now imagine Ann, another human being who is "hooked up" to Beth's body as in Armstrong's case, but more comprehensively. Not only does Ann receive proprioceptive information from Beth's body, she *only* receives such information from Beth's body. Moreover, Ann receives perceptual information only from Beth's body, seeing what Beth sees, hearing what Beth hears, tactilely perceiving what Beth touches, and so on. In addition, Ann quasi-remembers only Beth's memories. Finally, all of Ann's perception-like awareness of mental states and properties (i.e., detection forms of introspection) is detection of Beth's mental states and properties.[14] This seems to be enough to establish that any S-acquaintance instances on which Ann's first-person thoughts could be based are instances of S-acquaintance with Beth. So if Ann could think first-person thoughts in this scenario, then if NAR is correct, those thoughts should be about Beth.

This case is similar to an illusory absence case, except that Ann's perception and memory has its source in another individual, Beth, rather than in a fabrication. Just like in the illusory absence case, it seems that Ann could think first-person thoughts in this scenario. For Ann is not deprived of the kind of stimulus and information that might be needed to prompt first-person thought. However, it also seems that at least some of the first-person thoughts Ann would have in this scenario would not be about Beth. In particular, it seems that Ann could easily be prompted to think cogito-like thoughts, such as thinking that she is thinking that grass is green. These thoughts would intuitively be about Ann, not about Beth. (After all, the thought is intuitively true, but Beth, we may assume, is not thinking that grass is green.) But if they were about a particular thing partly in virtue of the thinker's S-acquaintance with that thing, then they could only be about Beth.

One can imagine a philosopher of an enactivist or embodied cognition bent arguing that in such a case, the thinker of the cogito-like thought is not Ann, but the odd combined entity of Ann and Beth. Since Ann's cognitive and perceptual activity is so seamlessly and comprehensively integrated with Beth's, the thinker who is Ann is now Ann as augmented by Beth. So when this thinker thinks a cogito-like thought, the thought is about this augmented thinker. It may be about this augmented thinker in virtue of the augmented thinker having S-acquaintance with itself, since it has S-acquaintance with Beth who is a part of itself.

But we can cut off such avenues by taking Beth out of the picture and turning this into an illusory absence case. In such a case there is no other individual who could be a part of the relevant thinker and with

[14] Here I follow Armstrong in taking it that it is "perfectly conceivable that we should have direct awareness of the mental states of others" (1993: 124), at least as long as that direct awareness is understood as detection-style introspection. (Armstrong suggests that in a case where one has such awareness of another's mental states we might call it 'telepathy' instead of 'introspection.')

whom that thinker has S-acquaintance. So it seems the thoughts could not be about the thinker in virtue of S-acquaintance with the thinker. Hence the total diversion case either is a real problem for Aboutness through S-acquaintance and for NAR, or leads us back to the illusory absence case that is a real problem. For the same reasons as laid out above, it is also a problem for Aboutness through Epistemic Gain. But again, the problem stems from the thinker's lack of S-acquaintance with herself (or of ways of gaining knowledge of herself that could determine first-person aboutness). It does not stem from her having these links to others.

5. *Prospects for a satisfying reflexive rule account?*

At this point, it might seem inevitable that for first-person thought, Aboutness through Epistemic Gain and Aboutness through S-acquaintance should be rejected (the latter taking NAR down with it). This would leave us with Aboutness by Reflexive Rule, and the choice of either accepting that there is nothing more to say about in virtue of what first-person thoughts refer to what they do, or the task of developing a more substantive picture of what it is for a type of thought to be governed by a reference-determining reflexive rule.

One such picture of the latter sort is offered by François Recanati (2012, 2014).[15] Recanati claims that first-person thoughts refer to what they do in virtue of involving a certain type of indexical mental file which he calls the "SELF file." Like all indexical files in Recanati's picture, a SELF file refers to a particular thing in virtue of the file's having the function of exploiting a certain epistemically rewarding ("ER") relation between the subject in whose mental architecture it appears and that thing. In the case of SELF files, this epistemically rewarding relation is *identity*. Identity is an epistemically rewarding relation for a thinker in virtue of making certain kinds of knowledge possible for her given her cognitive equipment. For those whose cognitive equipment includes faculties like proprioception, kinaesthesia and introspection, identity is epistemically rewarding and hence fit to be exploited by a mental file. Since SELF files refer to particular objects in virtue of having the function of exploiting the identity relation between subjects and those objects, and since the identity relation relates any subject to herself, it follows that whenever a SELF file is used (i.e., whenever anyone thinks a first-person thought), it refers to the thinker.

This seems to me to be an attempt at giving substance to the view that first-person thought is governed by a reflexive reference-determining rule. Recanati says that the functions of mental indexical files like the SELF file play the same role for these files as conventional mean-

[15] Recanati describes his view as an epistemic view and as in agreement with Evans and Morgan in taking an epistemic approach. But it seems to me that in fact his approach is better described as a reference rule view. I will elaborate on this as the section develops.

ings play for linguistic indexicals, namely: "through their functional role, mental file types map to types of ER relations, just as, through their linguistic meaning (their character), indexical types map to types of conextual relation between token and referent" (2012: 60). So a SELF file, because its functional role is to exploit the identity relation for information, refers to the object identical to the subject. In this way, functional role provides a substantive notion of governance by a rule, doing for thoughts what conventional meanings are supposed to do for language. Recanati's picture adds substance to First-Person Aboutness by Reflexive Rule by taking the claim that first-person thought is about whatever the thinker bears the identity relation to and saying *in virtue of what* this is the case. The story is that it is the case in virtue of the identity relation being epistemically rewarding given the cognitive equipment of the thinker.

Is this a satisfactory substantive notion of first-person thoughts being governed by a reflexive rule? I do not think it is entirely satisfactory, for the following reason. In illusory absence and total diversion cases, faculties such as proprioception and kinaesthesia do *not* make the identity relation epistemically rewarding. These faculties do not enable Ann to gain knowledge of the person to whom she is identical; if they enable her to gain knowledge of anything, it is Beth. They do not enable the person in the enhanced Matrix scenario to gain knowledge of the person to whom she is identical. Of course, if one includes cogito-like thoughts as types of introspection, one can say that introspection still allows these subjects to gain knowledge about the people to whom they are identical. But again, that these thoughts are ways for these subjects to gain knowledge about themselves depends on these subjects' first-person thoughts being about themselves. So this appeal to introspection cannot be used as part of a satisfying explanation of in virtue of what first-person thought is about what it is about.

In what sense, then, can a mental file possessed by thinkers in these situations function to exploit the epistemically rewarding relation of identity? Recanati might say that a thinker in an illusory absence situation deploys a malfunctioning token of the file-type SELF. The token would be malfunctioning because it fails to meet the "normative requirement corresponding to the function of the file," which is that "the subject should stand in a suitable ER relation to some entity (the referent of the file)." Nonetheless, the file-type might be tokened (and, presumably, refer to the individual to whom the thinker bears the identity relation) in the absence of such a relation so long as there is "a presumption that the normative requirements are (or will be) satisfied" (Recanati 2012: 63–64). The problem with this approach, it seems to me, is that there might not be a presumption that identity is an epistemically rewarding relation—for instance, the subject might be convinced that she is in an illusory absence scenario—yet cogito-like thoughts still seem thinkable.

Perhaps a more promising reply is to emphasize the importance of the *type* of file of which individuals' SELF files are tokens. It is the function of this *type* of file to exploit the relation of identity, which is epistemically rewarding given the cognitive equipment of some relevant *type* of thinker in which this type of file occurs.[16] As Recanati puts it, "Mental files are typed according to the type of ER relation they exploit." But here the question arises as to whether individual mental files are of the type SELF in virtue of exploiting the ER identity relation, or whether individual mental files exploit the ER identity relation in virtue of being of the type SELF. If the former, then it is not clear that subjects in the illusory absence cases have what could properly be considered tokens of the type SELF. This would imply that they do not think first-person thoughts, which seems to be false. If the latter, then the Question of First-Person Aboutness is going to be answered roughly as follows: a first-person thought refers to the one who thinks it in virtue of using a mental file of a type that exploits the ER relation of identity. But there is no longer anything to say about in virtue of what the first-person thought uses a mental file of that type. It is not in virtue of the thought's involving a file that exploits the identity relation, given that in illusory absence cases it seems subjects could not have such a file. So we seem to be stuck with the answer that a first-person thought is about the one who thinks it in virtue of being the type of thought that is about the one who thinks it.

6. *Conclusion: The crux of the problem*

Let us take stock. We saw that NAR requires that for first-person thought, the right account of reference determination must be based upon instances of S-acquaintance. But there are familiar problems for epistemic views of reference determination for first-person thought which seem to apply equally to S-acquaintance views. These problems might augur for rejection of such views (and NAR to boot), but as Morgan has argued, the alternative reflexive rule approach to the reference of first-person thought seems not to answer the question of reference determination at all. This provides motivation to try to overcome the familiar problems for epistemic views or S-acquaintance views. In digging in to these familiar problems, I have found that as they are usually presented, they are not such big problems after all. But variant presentations of them do reveal a serious problem: it is not clear how either epistemic or S-acquaintance accounts of the reference of first-person thoughts can allow for cogito-like thoughts to be about their thinkers in illusory absence cases. It is not helpful to point out that cogito-like thoughts are ways of gaining self-knowledge and can be included in an epistemic account of first-person reference determina-

[16] What might the relevant type of thinker be? Perhaps human beings? Or rational beings?

tion. For what makes these mechanisms be ways of gaining knowledge about a particular individual is that the thinker's first-person thoughts are about that individual.

So it seems to me that there is a real problem for both epistemic accounts of first-person reference determination and S-acquaintance based accounts. Thus, there is a real problem for NAR. However, there is also a real problem with dismissing these accounts: it seems to leave us with no account of what makes first-person thoughts be about what they are about. Maybe the problems for the alternative accounts are insuperable enough that we will in the end be driven to accept a kind of primitivism about the reference of first-person thoughts. This would in turn demand a basic, albeit limited, exception to NAR. I am not tempted by this option, but at present I am also not sure how to solve the problems for S-acquaintance and epistemic views of the reference of first-person thought. I hope that the above explorations have at least succeeded in properly carving out these challenges.

References

Anscombe, E. 1981. "The First Person." In her *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers*. Vol. II. Oxford: Basil Blackwell.

Armstrong, D. M. 1968/1993. *A Materialist Theory of the Mind*. London: Routledge.

_____, 1984. "Consciousness and Causality." In D. M. Armstrong and N. Malcolm (eds). *Consciousness and Causality*. Oxford: Basil Blackwell.

Bach, K. 2008. "On Referring and Not Referring." In J. Gundel and N. Hedberg (eds.). *Reference: Interdisciplinary Perspectives*. Oxford: Oxford University Press.

Burge, T. 1988. "Individualism and Self-Knowledge." *Journal of Philosophy* (85): 649–663.

Campbell, J. 1994. *Past, space and self*. Oxford: Clarendon Press.

Devitt, M. 1981. *Designation*. New York: Columbia University Press.

Dickie, I. 2016. *Fixing Reference*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Field, H. 1973. "Theory Change and the Indeterminacy of Reference." *Journal of Philosophy* (70): 462–81

Hawthorne, J. and Manley, D. 2012. *The Reference Book*. Oxford: Oxford University Press.

Jeshion, R. 2010. "Singular thought: acquaintance, semantic instrumentalism, and Cognitivism." In R. Jeshion (ed.). *New Essays on Singular Thought*. Oxford: Oxford University Press.

Kaplan, D. 2012. "An Idea of Donnellan." In J. Almog and P. Leonardi (eds.). *The Philosophy of Keith Donnellan*. Oxford: Oxford University Press.

Morgan, D. 2015. "The Demonstrative Model of first-person thought." *Philosophical Studies* 172: 1795–1811.

O'Brien, L. 2007. *Self-Knowing Agents*. Oxford: Oxford University Press.

Peacocke, C. 2008. *Truly understood*. Oxford: Oxford University Press.

Pepp, J. 2019. "Principles of Acquaintance." In T. Raleigh and J. Knowles (eds.). Acquaintance: New Essays. Oxford: Oxford University Press.

Recanati, F. 2012. *Mental Files*. Oxford: Oxford University Press.

_____, 2014. "First person thought." In P. Engel et al. (eds.). *Liber Amicorum*. <https://www.unige.ch/lettres/philo/publications/engel/liberamicorum>.

Russell, B. 1905. "On Denoting." *Mind* 14: 479–493.

_____, 1910–11. "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11.

Schwitzgebel, E. 2016. "Introspection." *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/introspection/>.

Siegel, S. 2002. "The role of perception in demonstrative reference." *Philosophers' Imprint* 2 (1): 1–21.

Wulfemeyer, J. 2017. "Bound Cognition." *Journal of Philosophical Research* (42): 1–26.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Aristotle's Perceptual Optimism

PAVEL GREGORIĆ
Institute of Philosophy, Zagreb, Croatia

In this paper, I would like to present Aristotle's attitude to sense-perception. I will refer to this attitude as "perceptual optimism". Perceptual optimism is, very briefly, the position that the senses give us full access to reality as it is. Perceptual optimism entails perceptual realism, the view that there is a reality out there which is accessible to our senses in some way or other, and the belief that our senses are veridical at least to some extent, but it is more comprehensive than that. For instance, a perceptual optimist does not admit such things as qualities which are perceptible in principle but not by us or bodies too small to be perceptible. In this paper I argue that Aristotle is a perceptual optimist, since he believes that reality, at least in the sublunary sphere, is indeed fully accessible to our senses. In the first and largest part of this paper, I will show, in seven distinct theses, what Aristotle's perceptual optimism entails. In the second and shorter part, I will put Aristotle's position in a wider context of his epistemology and show why it was important for him to be a perceptual optimist.

Keywords: Senses, direct realism, qualities, sensibles, veridicality, Democritus**.**

I.

I suppose it is uncontroversial that Aristotle's universe is a universe of substances and their attributes. It is equally uncontroversial that Aristotle's universe is divided in two rather different parts, the sublunary and the supralunary. Both parts of the universe are composed of material substances, that is bodies, and their attributes. However, the sublunary part is marked by all sorts of changes and transformations, which are maintained in everlasting order by the circular motions of celestial bodies and their immaterial unmoved movers. All bodies in

the sublunary world are made of elements, each element featuring a combination of two qualities: hot or cold, dry or moist.[1]

So, all bodies that populate the sublunary world necessarily have a pair of these elementary qualities—some degree of hotness or coldness, some degree of dryness or moistness. Now, these elemental qualities are tactile qualities, that is qualities essentially picked up by the sense of touch. Aristotle says that tactile qualities are the distinctive characteristics of bodies *qua* bodies.

> The distinctive characteristics of the body *qua* body, are tactile. By distinctive characteristics I mean those by which the elements are distinguished—hot and cold, dry and moist—and about which we spoke earlier in our discussions about the elements (*On the Soul* II.11, 423b27–29).[2]

So, the sense of touch puts us in contact with the most fundamental qualities, that is the qualities of the elements from which the whole sublunary world is built.[3]

More to the point, Aristotle thinks that human beings have especially refined sense of touch and he connects that with our intelligence.

> In the other senses humans fall short of many other animals, but in regard to touch they achieve greater precision than the others. Hence the human being is the most intelligent of animals. (*On the Soul* II.9, 421a21–23)[4]

There is a complicated story in Aristotle about why humans have especially refined sense of touch and why it makes them the most intelligent of all animals (*phronimōtaton tōn zōiōn*). Suffice it to say that this has something to do with the heart, which is not only the central organ in Aristotle's theory, but also the proper sense organ of touch. Aristotle argues that the human heart is composed of flesh made from the finest mixture of elements, its hotness is well balanced by respiration through our large lungs and by the inherent coldness of our large brain, and also the central position of the heart inside an erect body of human beings relieves it of the pressure from the upper parts, so it can

[1] Fire is thus hot and dry, air is hot and moist, water is cold and moist, earth is cold and dry. The transformation of elements is effected by way of preserving one and replacing the other elementary quality, e.g., air turns into water when it replaces hotness with coldness while preserving moisture.

[2] All translations of Aristotle are mine. Apart from the Revised Oxford translation of Aristotle's Complete Works in Barnes (1984), I consulted three recent English translations of *De anima*, Shields (2016), Reeve (2017) and Miller (2018).

[3] Apart from the two elementary qualitative ranges (hot-cold, dry-moist), Aristotle sometimes adds further qualitative ranges to the domain of tactile qualities, such as soft-hard, rough-smooth, light-heavy. Aristotle's understanding of sensible qualities stands in stark contrast with that of Democritus, who says: "For by convention sweet, by convention bitter, by convention hot, by convention cold, by convention color, but in reality atoms and void" (fr. B125 Diels-Kranz, Taylor's translation); see also fr. A37 quoted below in n. 20.

[4] See also: *On the Sense* 4, 441a1–2: "...the sense of touch is most precise in comparison with all the other animals." *Parts of Animals* II.16, 660a12–13: "The human beings are the most perceptive of animals with respect to the tactile sense."

function optimally.[5] In any case, Aristotle seems to believe that our exceedingly refined sense of touch guarantees that we get the elementary qualities right.

Of course, bodies may have further qualities which are related to the other senses as tactile properties are related to the sense of touch. Such properties Aristotle calls "special sensibles." Special sensibles are properties which are perceived directly, or in themselves (*kath' hauto*), and they are accessible to one special sense only, e.g., colours are accessible only to the sense of sight and sounds only to the sense of hearing. As such, special sensibles are properties with reference to which each special sense is *defined*, e.g., the sense of sight is essentially the ability to perceive colours.[6] Whatever else is perceived by sight, it is perceived by way of, or in accompaniment of, colours that are being seen. To return to my main point, apart from the tactile qualities, most bodies have colours—or show colours of other bodies, in the case of transparent bodies—some of them have flavours, some emit smells and some produce sounds when struck. Equipped with the five senses—touch, taste, smell, hearing and sight—we can access all the aforementioned properties of bodies. I shall return to this point presently.

Another important thing about Aristotle's theory of perception is his talk of "reception of form without matter," of the sense becoming "like" its object, and of the "identity" between the sense and its object. Here are some representative passages:

> That which can perceive <i.e. a sense> is in potentiality like that which can be perceived <i.e. sensible> is already in actuality, as it has been said. For the former is affected when it is not like the latter, but after it has been affected it has become like it and similar to it. (*On the Soul* II.5, 418a3–6)

> We should assume, then, concerning all sense-perception that a sense is that which can receive perceptible forms without matter. (*On the Soul* II.12, 424a17–18)

> Now, then, by way of summarizing the things which have been said concerning the soul, let us say again that the soul is in a sense all existing things; for what exists is either objects of perception or objects of thought; and knowledge <i.e. a fully actualized faculty of thought> in a way *is* the objects of knowledge, and perception <i.e. a fully actualized faculty of perception> in a way *is* the objects of perception. (*On the Soul* III.8, 431b20–28)

These and related passages have been widely discussed by scholars, some arguing that the eye becomes literally red when we see an apple ("literalism"), others arguing that there is no physical change that underlies an act of seeing an apple, at any rate not in the way Aristotelian matter underlies form ("spiritualism"), still others that neither of these

[5] For the composition of the heart and the erect posture of human beings, see Aristotle's *Parts of Animals* II.1, III.4 and IV.10, Gregoric (2007: 40–51) and Gregoric (2005).

[6] In fact, there are two types of special sensibles of the sense of sight, colours and phosphorescent things, the former requiring light and the latter requiring darkness; cf. Gregoric (2018).

two positions is quite right.[7] Without taking a definite position on this long and often very subtle debate, I think Aristotle's position is as follows. When we see an apple, our sense of sight takes on the red colour of the apple without taking on the apple's matter. It is not that our sense of sight takes on some sort of copy of the apple's red colour, some sort of representation or encoded information which then gets suitably interpreted. Rather, our sense of sight takes on the very property instantiated in the apple, the token of red colour that is in the apple, and likewise with the other senses and their special sensibles. I understand Aristotle's talk of "taking on the form without matter" and the sense becoming "like" the object, then, as a strong version of direct realism about the perception of special sensibles. On that point I side with a number of scholars who take Aristotle to be a direct realist, though this is not entirely uncontroversial.[8]

If the sense takes on the token quality out there, or to the extent that it does, the sense cannot go wrong about it. Indeed, Aristotle writes at several places in *On the Soul* and in other works, that the senses do not go wrong concerning their special sensibles.

> "That which can be perceived" <i.e. a "sensible"> is spoken of in three ways: in two ways it is perceived in itself, and in one accidentally. Of the first two, one is special to an individual sense and the other common to them all. By "special" I mean that which cannot be perceived by another sense and concerning which there cannot be deception, as sight is of colour, hearing of sound, taste of flavour, whereas touch has several different qualities. But each sense discriminates concerning these qualities and is not deceived that there is colour nor that there is sound, but what or where is the colored object, or what and where is the object that emits sound. (*On the Soul* II.6, 418a8–16)

> That is why the senses are deceived about these <viz. the common sensibles>, but are not deceived about the special sensibles, e.g. sight about colour or hearing about sound. (*On the Sense* 4, 442b8–10)

> As for truth, to show that not everything that appears is true: first, perception, at least of the special sensible, is not false, though appearance is not the same thing as perception. (*Metaphysics* IV.5, 1010b1–3)

Although Aristotle does not quite say so, I take it that the senses are veridical concerning their respective special sensibles *because* there is a sort of identity between the senses and their objects in acts of perception.

[7] For a thorough overview of the debate, with a detailed map of different positions, see Caston (2005).

[8] See, e.g., Owens (1981), Burnyeat (1992), Broadie (1993). One source of controversy are the passages in which Aristotle describes the special sensible as a *logos* between the two extremes on a qualitative spectrum, on the one hand, and the sense as a *logos* and a "mean" affected by the sensibles, on the other hand. This allows for an interpretation according to which perception consists in the sense instantiating the same *logos* that the object instantiates with the special sensible. Though this is not quite the same as representationalism, it is not direct realism, either. For a defence of this sort of view, see Caston (2005: 299–315). See also Caston (1998) for his earlier challenge to the view that Aristotle was a direct realist, with an interesting response by Putnam (2000).

At one point, however, Aristotle says that the senses are subject to error *in the smallest degree* with regard to the special sensibles:

> Perception of the special sensibles is true, or is subject to falsity in the smallest degree. Second, perception is of that to which the special sensibles accidentally belong; and already here it is possible to be mistaken. For there is no mistake in that it is white; but that the white is this or other, there is mistake. Third, perception is of the common sensibles which accompany the accidental sensibles to which the special sensibles belong (I mean, for instance, motion and magnitude); concerning these it is in fact especially possible to fall into error with respect to perception. (*On the Soul* III.3, 428b18–25)

The qualification in the first sentence most probably refers to abnormal circumstances, such as illness or fatigue, special condition of the sense organ, unusual state of the medium, large distance and other unfavourable conditions of perceiving. In normal circumstances, however, a sense gets its special objects right, and I take it that it gets them right because it is *in-formed* by them, for the sense takes on the very sensible form of the object.[9]

Let me now briefly pause to state the first three components of Aristotle's perceptual optimism.

First, all material substances necessarily have some properties that directly, in themselves, activate our senses. In other words, there are no material substances in the sublunary world which are fundamentally imperceptible, that is imperceptible because they do not have any special sensibles. I will call this the "universal perceptibility thesis."[10]

Second, the perceptible properties that directly, in themselves, activate our senses—that is the special sensibles—are as real as the material substances to which they belong, and they are perceived because the special senses take them on and become identical with them in acts of perception. This is the "direct realism thesis".

Third, because the senses take on special sensibles and become identical with them in acts of perception, there is no room for error, at least in normal circumstances. I propose to call this the "qualified perceptual veridicality thesis."

I call this veridicality thesis "qualified" for two reasons. First, because Aristotle admits abnormal circumstances in which the senses can go wrong about their special sensibles. Second, because Aristotle recognized other types of sensible items, namely the common and the accidental sensibles, with regard to which the senses can and often do go wrong.

The common sensibles are properties such as shape, size, motion and number, which are perceived insofar as they accompany special sensibles, and they invariably do accompany special sensibles. We cannot perceive white without this white being of a certain shape and size,

[9] Recent literature on the subject includes Johnstone (2015) and Koons (2018).

[10] This thesis stands in stark contrast with the teaching of Democritus: "For by convention sweet, by convention bitter, by convention hot, by convention cold, by convention color, but in reality atoms and void" (fr. B125 Diels-Kranz). See also fr. A37 quoted below in n. 20.

in motion or at rest, one or many. They are called "common" because they are perceived by two or more special senses. There are different views as to how precisely the common sensibles are perceived, but everyone agrees that the senses need to be unified in some way or other in order to grasp the common sensibles.

Accidental sensibles are substances and their locations,[11] but presumably also classes and relations, possibly even facts. All such things are perceived insofar as a set of special and common sensibles accidentally happens to be this or that. Because a certain combination of colours of some shape and size happens to be Thomas, I perceive Thomas. There are various ways to understand this. Some scholars think that accidental perception is not perception at all, but a way of reporting perceptual events, some think that this is a sort of "association of ideas" which requires either a minimal conceptual apparatus or an involvement of non-rational capacities such as memory and imagination (*phantasia*), and still others construe it as a genuine sort of perception.[12]

In any case, I should like to emphasize that Aristotle is not a Protagorean relativist or an Epicurean who subscribes to the view that all perceptions are true.[13] No, there is only one type of sensible items which we get right, according to Aristotle, namely the special sensibles, and we get them right only in normal conditions; that is why this is a *qualified* perceptual veridicality thesis. But it is a veridicality thesis nonetheless, at the very fundamental level.

Aristotle's perceptual optimism runs much deeper than these three theses. In *On the Soul* III.1, Aristotle raises the question why we have more than one special sense.

> Could it be in order that the accompanying and common sensibles (e.g. motion, magnitude and number) may be less likely to escape our notice? For if there were only one sense—say, sight of white—the common sensibles would rather have escaped our notice and would seem to be the same because colour and magnitude always accompany one another. But in fact, since the common sensibles are found also in the other type of sensible <i.e. magnitude accompanies not only colours but also tangible properties>, this makes it clear that each of them is different. (*On the Soul* III.1, 425b5–11)

According to this passage, then, we have a plurality of senses in order to increase the accuracy of perception of the common sensibles, with respect to which perception is most likely to go wrong. The gist of Aristotle's argument seems to be the following. Every time we perceive a colour, we perceive a patch of some shape and size, it is either one or many, moving or resting. If we had only the sense of sight, the argument goes, there would be nothing to make us aware of the fact that

[11] See, e.g., Aristotle's examples in *On the Soul* II.6, 418a16–17, 20–23.

[12] The classic paper on accidental sensibles is Cashdollar (1973). A discussion of different positions on accidental perception, with extensive bibliography, can be found in Perälä (forthcoming).

[13] See, e.g., Lee (2005: 133–180), Striker (1977) and Everson (1990).

colour is a different property from shape, size, number and motion. But, as things are, we have the sense of touch too, and every tangible quality that we feel also comes with some shape and size, one or many, moving or static. Because shape, size, number and motion accompany not only colours but also tangible qualities, we realize that they are in fact different properties from both colours and tangible qualities.

On the face of it, this is not a convincing argument. Why could not one realize that colours are different from shapes and sizes by noticing that colours of a certain shape and size can change while the shape and size remain the same, as when a chameleon turns from brown to green? Or by noticing that a certain shape and size change while the colour remains the same, as when one moulds a chunk of wax? Aristotle might respond to this that such cases would inform the perceiver that colours and shapes can vary independently of one another, but not that they are two independent *types* of properties. To understand that, the perceiver needs to have access to shapes as they accompany tangible qualities and realize that the shape which accompanies a colour of an object is the very same property that accompanies the tangible qualities of that object.[14]

Still, Aristotle's argument explains, at best, why we have two senses—touch and sight—not why we have all five of them. Indeed, the other three senses are not particularly good at perceiving the common sensibles, anyway. I mean, smell or taste hardly allow us to perceive much of the common sensibles. The real and more fundamental reason why we have five senses is, no doubt, to enable us to receive the five different types of special sensibles: colours, sounds, odours, flavours and tangible qualities. Aristotle does not say so in as many words, but this is clearly what follows from his teleological framework.[15]

Now, to understand the extent of Aristotle's perceptual optimism, it is important to observe that the five different types of special sensibles, for which we have five different senses, are *all* such properties that exist in the universe. That is to say, there are no further properties of this sort, some sixth type of special sensible which defines some sixth sense that we do not happen to be endowed with. This is what follows from Aristotle's argument against the existence of a sixth sense from the beginning of *On the Soul* III.1, 424b22–425a13.

This passage reveals two crucial things for my story. First, Aristotle is convinced that each sense is receptive of the *whole* range of qualities that fall under its province.

[14] George Berkeley famously denied that the shape or size we see is in fact the same property as the shape or size we feel. For instance, in §127 of his *Essay towards a New Theory of Vision*, he wrote: "The extension, figures, and motions perceived by sight are specifically distinct from the ideas of touch called by the same names, nor is there any such thing as one idea or kind of idea common to both senses."

[15] See, e.g., *On the Sense* 1, 436b10–437a17 and 5, 444b19–20, *History of Animals* IV.8.

> As things are, we have perception of everything of which touch is the sense, for all tangible qualities are perceptible by us by means of the sense of touch. (*On the Soul* III.1, 424b24–25)

What Aristotle is saying here, I take it, is that there is no type of tangible quality such that it falls outside of the range to which our sense of touch is receptive. Admittedly, the same applies to the other senses, e.g. there is no shade of colour which is invisible to us. So, something like infrared or ultraviolet is out of the question for Aristotle. The sense of sight is sensitive to *all* colours there are.

Of course, this does not mean that animals or individuals within the same species do not differ in the sharpness of their sense of sight. Indeed, Aristotle thinks that people with blue eyes have better sight in the dark, whereas people with dark eyes have better sight in light.[16] Moreover, some people can see farther than others and others have a higher resolving power at close distances.[17] In all such cases, sharpness of sight has something to do with the constitution of the sensorium—that is the eye as the peripheral sense organ, the blood or *pneuma* as the internal medium of transmission, and the heart as the central sense organ.

Despite these variations across species and among individuals of the same species, however, none of the senses is fundamentally lacking by being restricted only to a part of the range of its corresponding special sensible. Rather, each sense is receptive of the *full range* of qualities that constitute its special sensible (and with reference to which each sense is defined and about which it does not go wrong in normal circumstances). Let us call this "the full-range receptivity thesis."

The second thing that Aristotle's argument reveals is even more astonishing from a modern point of view. Aristotle maintains that there are no other than the five senses.[18] His argument goes like this. If there were an extra sense, there would be an extra sense organ. But sense organs—or their crucial parts which are receptive of special sensibles—can only be made of simple bodies. There are only four simple bodies in the sublunary sphere: earth, fire, air and water. Now, earth either cannot serve as a sense organ, or else it enters the constitution of the sense-organs of the contact senses, touch and taste. Similarly, fire either cannot serve as a sense organ, or it is common to all the sense-organs, given that all sensitive beings are warm. This leaves us with air and water. Being transparent and thus receptive of colours (and easily ensconced), water is used up for the sense-organ of sight. Being conductive of sounds, air is used up for the sense organ of hearing, and either

[16] *Generation of Animals* V.1, 779b12–780a25. See also *Generation of Animals* V.1, 780a25–36 and *Parts of Animals* II.13, 657a31–34 for the thinness of skin surrounding the eye contributing to the sharpness of sight.

[17] *Generation of Anima*ls V.1, 780b14–781a13.

[18] Aristotle's argument can be interpreted also modally, to the effect that there can be no other than the five senses; cf. Shields (2016: 255–257). This should be contrasted with Democritus' fr. A116 (Diels-Kranz): "Democritus says there are more senses <than the five>, for irrational animals, wise men and gods."

water or air is used up for the sense organ of smell. So, given that there are no other simple bodies, there can be no other sense organs, and hence there can be no other senses.[19] In Aristotle's own words:

> Consequently, if there is no other <simple> body and no quality such that it does not belong to any of the <simple> bodies in this world, no sense would be left out. (*On the Soul* III.1, 425a11–13)

This is not a particularly convincing argument. If one simple body can serve as an organ of two senses, e.g. water as an organ of sight and hearing, why can it not serve as an organ of three or more senses? The prominent place of this argument in the treatise *On the Soul*, ushering in a new stage of Aristotle's account of the perceptual faculty of the soul at the beginning of Book III, suggests that he found it rather important. And we can see why it is important for Aristotle to rule out the possibility that there are senses beyond the familiar five ones: if another special sense existed, it would be defined with reference to its range of special sensibles of which we have no idea, and this would mean that bodies have properties which are fundamentally perceptible—but not by us. In other words, this would mean that there is a whole segment of reality to which we humans have no access. And if there were such a segment of reality, we would rightly question whether the rest of our knowledge of the world is correct. That is to say, if there were a segment of reality to which we have no access, that would mean that our inductions are seriously compromised, which in turn means that we may not have gotten all the universals, or that the universals we did get may be incomplete or ill-founded. Admitting the sheer possibility of an extra sense, then, would compromise Aristotelian science and make it vulnerable to sceptical objections.

Aristotle's perceptual optimism runs still deeper. In the following pages I would like to discuss two further component theses. Both of these additional theses are found in passages from Aristotle's less known work, the short treatise *On the Sense and the Sensibles* (*De sensu et sensibilibus*) from the collection *Parva naturalia*. In this treatise Aristotle raises various problems related to the senses and the special sensibles. One of the problems is whether there are invisible magnitudes.

This problem is mentioned for the first time in chapter 3 of *On the Sense*, where Aristotle discusses various theories of colours.

> Hence, if it is not possible for a magnitude to be invisible, but rather every magnitude is visible from some distance, the superposition theory too might pass as a theory of mixture of colours. Indeed, on the juxtaposition theory too, there is nothing to prevent some combined colour from appearing to viewers at a distance. That there is no magnitude such as to be invisible has to be discussed later on. (*On the Sense* 3, 440a26–31)

It is clear from this passage that Aristotle thinks that there are no invisible magnitudes. He explicitly says that "every magnitude is vis-

[19] See also *On the Sense* 5, 444b19–20.

ible from some distance". (For all practical purposes, we can replace his term "magnitude" here with the term "body".) One should be reminded that, according to Aristotle, a body is visible on account of colour—either its own or colour of other bodies seen through it, if the body is transparent. Of course, colours are always accompanied by some shape and size, they are either moving or at rest, but these common sensibles are not visible without colours.

The promise of a fuller discussion of this problem is met in Chapter 6 of *On the Sense*. There Aristotle frames the question explicitly with reference to all special sensibles. He wonders "if every body is infinitely divisible, are sensible qualities also infinitely divisible, for example, colour, flavour, smell, sound, heavy and light, hot and cold, hard and soft?" (445b3–6). Could Aristotle really mean to say, quite contrary to plain common sense, that there are no bodies so small as to escape being seen, heard, felt, etc.? Yes, he could, though his view is quite nuanced. Here is the whole passage:

> Since, then, the properties must be spoken of as species, though continuity is always present in them, we must take into account that potentiality is different from actuality. And for this reason, when one sees a grain of millet, a ten-thousandth part of it escapes notice, even though sight traversed it, and the sound within a quarter-tone escapes notice even though one hears the entire melody which is continuous. It is the interval between the extremes which escapes notice. Likewise with very small parts in the case of other objects of perception, too. Namely, they are potentially visible, but not actually, as long as they are not separate. For a foot length is potentially present in a two-foot length, but actually only after it has been removed. It is reasonable to suppose that, when they are separated, such tiny increments would be dispersed into their surroundings, like a flavoured droplet poured into the sea. However that may be, since the increment of sense-perception is neither itself noticeable nor separable (for the increment is potentially present in the more precise sense-perception), it is not possible to perceive actually such a tiny object of perception, either. However, it remains perceptible nonetheless; namely, it is so potentially already, and actually when added <to a larger object that actualizes one's sense>. (*On the Sense* 6, 445b29–446a15)

This is a difficult passage, but my understanding of it, in a nutshell, is as follows. Aristotle argues that something can be too small to be *actually* visible—his example is a ten-thousandth part of a grain of millet—while remaining always *potentially* visible. And it remains potentially visible in two ways. First, it remains potentially visible *while* integrated with the whole grain, because we actually see the whole grain, not an aggregate of parts, though of course the grain is potentially divisible into parts, and when we actually divide the grain into parts, we then see these parts. Second, a ten-thousandth part of a grain of millet, if we somehow managed to separate it off from the whole—Aristotle seems to suggest—would remain only potentially visible, because it would be immediately "dispersed to its surroundings." I take Aristotle to be saying that such a tiny part would immediately merge with another body

in its surrounding, thus becoming only potentially visible in the first sense, as a part of this body with which it merged. Again, we would perceive the whole of this body, and only potentially its parts.

Thus, Aristotle accommodates the common-sense view that there are bodies too tiny to be actually seen, yet he prevents the inference that, therefore, there are bodies which are fundamentally invisible, that is bodies which are not characterized by colours. Of course, this is precisely what the ancient atomists advocated, namely that there are imperceptibly small bodies of different shapes, sizes and motions, but no colours, flavours or temperature.[20] Aristotle, by contrast, believed in the qualitative world from the smallest to the largest of scales.[21]

So, the sixth component of Aristotle's perceptual optimism is the thesis that there are no bodies fundamentally inaccessible to our senses.

The seventh and the last component of Aristotle's perceptual optimism that I wish to discuss is the thesis that there are no imperceptibly short intervals of time. Aristotle argues in support of this thesis in chapter 7 of *On the Sense*. This thesis is part of his reply to the problem of simultaneous perception. The question is whether two special sensibles can be perceived at the same time, which seems to be a problem for the individual senses, because Aristotle argued that only one thing can bring about one act of perception at one time. One possible way out of this problem is to propose that, in fact, we cannot perceive two special sensibles at the same time, but if the time between perceiving one and perceiving the other is too short to be perceptible, it will seem to us that we perceive two special sensibles at the same time. However, Aristotle does not like this solution precisely because he does not like the idea of imperceptibly short intervals of time.

Aristotle supplies two arguments against imperceptibly short intervals of time. Here is the first argument:

> For if, when someone perceives himself or anything else in continuous time, it cannot at that time escape his notice that he exists; but if there is within the continuous time some part which is so short as to be entirely imperceptible, it is clear that at that time it would escape his notice that he himself exists, sees and perceives. (*On the Sense* 7, 448a26–30)

The gist of this argument is that, if there were imperceptibly short intervals of time, we would not perceive anything in such intervals,

[20] See, e.g., fr. A37 (Diels-Kranz), which comes from Aristotle's lost treatise *On Democritus*: "Democritus thinks that substances (viz. atoms) are so small as to elude our senses, but they have all sorts of forms and shapes and differences in size. So he is already enabled from them, as from elements, to create by aggregation bulks that are perceptible to sight and the other senses."

[21] One might think that instruments such as the microscope and the telescope disprove Aristotle's thesis. However, they only redefine the threshold between actual and potential perceptibility, but do not eliminate it. Aristotle could point out that the bodies we see through a microscope or a telescope are coloured much like the bodies we see around us with the naked eye. So, far from undermining his sixth thesis, the instruments actually support it. Of course, the telescope would create problems for Aristotle on different grounds.

and hence we would not be aware of our own existence—we would not be conscious—in such intervals. However, our awareness of our own existence is continuous and uninterrupted, hence our perception is continuous and uninterrupted, therefore there are no imperceptibly short intervals. Of course, few of us today will find this argument convincing, not only because we know for certain that there in fact are intervals of time too short to be detected by our unaided senses, but also because few of us would be prepared to use the subjective diachronic unity of consciousness as a criterion of objective states of affairs in the world.

The second argument is perhaps less naive and certainly more elaborate.

> Moreover, there would be neither a time in which he perceives nor a thing that he perceives, except perhaps in the sense that he sees in some part of the time or sees some part of the thing—if indeed there is any magnitude, either of time or of the thing, which is entirely imperceptible due to its smallness. For if he sees the whole line and perceives it in the corresponding continuous time, he does not see it by means of some part of it. Let CB, in which he does not perceive, be removed <from the whole interval AB>. Then perception of the remaining part of the interval <i.e. AC>, or of what is perceived in that part of the interval, is like perceiving the whole earth on account of perceiving this particular part of earth, or like walking the whole year on account of walking in this particular part of year. But in CB he perceives nothing. Therefore, because he perceives in some part of the whole interval AB <viz. in AC>, he is said to perceive in the whole interval and the whole corresponding thing. And the same holds also in the case of AC. For one always perceives in some part of the interval and some part of the corresponding object, whereas the whole can never be perceived. Therefore, all things are perceptible, though they do not appear as large as they are. (*On the Sense* 7, 448a30–b13)

This is a *reductio ad absurdum* argument which can be reformulated as follows. Take a perceptible interval of time AB. That interval is perceptible because in its duration we perceive some one object, say line XY. Now, take out an imperceptibly short segment of the interval AB, let us call it CB. In CB, then, we do not perceive anything. In other words, in CB we do not perceive any part of line XY. Well, then, what happens in the remaining interval of time, AC? Clearly, in AC we perceive some part of line XY, let us call it XZ. Of course, once we admitted an imperceptibly short segment of the whole interval AB, we must admit it also for interval AC. Removing the imperceptibly short segment of AC, in the remaining part of it we perceive only a part of XZ, and so on *ad infinitum*. What follows is that the whole line XY can never be perceived—if imperceptibly short intervals of time are admitted. Indeed, nothing can ever be perceived, since any perceptible interval of time can be divided into an imperceptible interval and the correspondingly shorter perceptible interval. Therefore, there are no imperceptibly short intervals of time.

The second argument will probably remind the reader of Zeno's paradoxes and Aristotle's solution to it. As is well-known, Aristotle tackled

the paradoxes by arguing that magnitudes (bodies, spatial extensions, temporal intervals) are infinitely divisible only in potentiality, not in actuality. However, it is important to note that Aristotle's denial of infinite divisibility is very different from that of the atomists. Atomists denied infinite divisibility because they thought there was a ground level at which magnitudes cannot be further divided, that is the level of atoms of matter, atoms of spatial extension, atoms of time.[22] By contrast, Aristotle was a staunch continuist who thought that a magnitude cannot possibly be built from items that are not magnitudes. If something is a magnitude, Aristotle thought, in principle it is divisible. Atoms, being in principle indivisible, are not magnitudes. And you can never get a magnitude from items that are not magnitudes: a line is a not a collection of points, a place is not a collection of indivisible locations, a time-interval is not a series of indivisible "nows", and likewise a body is not an aggregate of uncuttable atoms. Similarly, a perceptible interval—that is a period of time in which we perceive something—does not consist of imperceptibly short intervals. This is the seventh and the last thesis that I propose to identify as constitutive of Aristotle's perceptual optimism.

Observe that the case of imperceptibly short intervals of time is parallel to the case of imperceptibly small bodies. Aristotle would happily concede that in any given interval of time there are segments that are only potentially perceptible, just as in any given body there are parts that are only potentially perceptible, but he would deny that any segment or part is so small as to be fundamentally imperceptible. In fact, the sixth and the seventh thesis go together. There are no fundamentally imperceptible intervals of time *because* there are no fundamentally imperceptible bodies. For, if there were actual imperceptibly short intervals of time, there would have to be actual imperceptibly small parts of bodies that are grasped in such intervals. However, since there are no imperceptibly small bodies in actuality, there cannot be imperceptibly short intervals of time in actuality, either.[23]

To summarize, I have identified seven theses that constitute Aristotle's perceptual optimism:

1. Universal perceptibility—all bodies have some special sensibles and are hence fundamentally perceptible.
2. Direct realism—special sensibles are real and the senses become "like" them.
3. Qualified perceptual veridicality—in normal circumstances the senses do not go wrong about their special sensibles.
4. Full-range receptivity—the senses are receptive of the whole spectrum or range of qualities that constitute their special sensible.

[22] This is certainly true for Epicurus, whereas it is an open question whether the earlier atomists argued for atomism of space and time.

[23] Not just the last two theses, but every single one of the seven identified theses constitutive of Aristotle's perceptual optimism seems to go against the teaching of ancient atomists.

5. No sixth sense—there are no extra senses and hence no extra ranges of qualities that would constitute their special sensibles.
6. No bodies fundamentally inaccessible to our senses—there are no bodies, regardless of their size, such that we cannot perceive them at least potentially.
7. No imperceptibly short intervals of time—there are no intervals of time, regardless of their length, such that nothing can be perceived in their duration.

I hope to have shown that Aristotle believed the universe, or at any rate its sublunary sphere, to be fully accessible to our senses. There are no scales, no unknown qualities, and no unknown ranges of otherwise familiar qualities, that are inaccessible to our senses. The qualities that exist and that are open to us, are knowable for what they are. In normal conditions we get them exactly as they are. Because we get these qualities right, we have a solid basis for perceiving correctly other types of properties too, although that may require some honing of our senses. That is to say, we can and naturally do improve our perception of the common sensibles as we become more experienced perceivers, and I suppose the same goes for the accidental sensibles. And because our perception is fundamentally veridical, Aristotle can rest assured that our knowledge based on perception is sufficiently well-founded.

II.

Aristotle's perceptual optimism is part of his general cognitive optimism: Aristotle believes that human beings can, in principle, know everything there is to be known in the universe. This is the view he shared with Plato, who divided the universe into the world of changing material objects that we perceive and the world of unchanging immaterial objects, called "forms" or "ideas", that we grasp by thinking. Aristotle diverged from Plato as to how material objects and forms are related and also how perception and thought are related. Very briefly, for Aristotle forms are the internal causes of material objects, not separately existing objects; and thinking is *founded* on perception, not something best performed *independently* from perception, as Plato had argued. According to Aristotle, if we perceive a sufficient number of objects and facts in a certain domain, if we remember them in an organized way, and if we then start to inquire about the causes of these objects and facts, we will naturally come to have an intellectual grasp of the relevant forms and of the explanatory relations among them, and that is precisely what it means to think (*noein*) in the primary sense of that verb. So, to grasp forms it is not that we must emancipate ourselves from the senses, recollect and engage in rigorous dialectical reasoning, as Plato had taught, but, on the contrary, we must first and foremost engage in extensive and systematic use of our senses.

If we fail to use the senses to acquire relevant data, not only does the move from the perception of particulars to the grasp of universals,

essences and explanatory relations among them become deeply problematic, but also an understanding of the universals, essences and explanatory relations among them is undermined. This is how Aristotle puts it:

> It is evident also that if some perception is wanting, some knowledge must also be wanting—knowledge which it is impossible to get if we learn either by induction or by demonstration, if demonstration depends on universals and induction on particulars, if it is impossible to study universals except through induction (...) and if it is impossible to make an induction without having perception, for particulars are grasped by perception. It is not possible to get knowledge of these items—neither from universals without induction nor through induction without perception. (*Posterior Analytics* I.18, 81a38–b9)

Similar empiricist statements can be found in several other places in Aristotle's works, most famously in *Posterior Analytics* II.19 and *Metaphysics* I.1. Despite such statements, however, it would be a mistake to call Aristotle an empiricist, since he agrees with Plato that there can be no scientific knowledge without grasping forms, and grasping forms is the task of a special and entirely independent cognitive capacity called "intellect" (*nous*), which requires development and which is, when fully developed, infallible.[24] It is important to point this out, because Aristotle is far from thinking that scientific knowledge (*epistēmē*) is reliable simply because and insofar as the senses supply correct data. Reliability of scientific knowledge is based, according to Aristotle, on the infallibility of the intellect at least as much as on the veridicality of the senses for supplying correct data. So, even though scientific knowledge can never be reduced to the correct use of the senses, the senses nonetheless have to be veridical for scientific knowledge to obtain. To quote one of the leading contemporary interpreters of Aristotle:

> Any truths that mortal minds may contemplate are obtained, directly or indirectly, by way of the five senses; and it is highly plausible (to say no more) to suppose that the objects of the mind's contemplation will be true only if the perceptual reports from which they were somehow obtained are also true. Thus rational creatures like us cannot achieve the good unless their senses are veridical. But nature does nothing in vain. Hence the senses are veridical. (Barnes 1987/2014: 608)

This is a statement of general cognitive optimism of the distinctly Aristotelian variety. Plato was also a cognitive optimist, as I have pointed out, but his cognitive optimism was based solely on the intellect, that is on recollection and dialectic in emancipation from the senses, whereas Aristotle's cognitive optimism was based to a large extent also on the senses.

If all our knowledge is ultimately founded on the senses, as Aristotle thought, the senses had better be veridical, at least at some fundamental level. The belief in veridicality of the senses, supported by direct realism and universal perceptibility, is the cornerstone of Aris-

[24] For a fuller presentation of Aristotle's position, see Frede (1996).

totle's perceptual optimism. However, even with these first three theses granted, knowledge would still be on shaky foundation, (i) if there were some qualities inaccessible to us but accessible to living beings endowed with extra senses, (ii) if familiar qualities had parts of their ranges inaccessible to our senses, (iii) if there were bodies in principle bereft of qualities that directly stimulate our senses, or (iv) if there were imperceptible periods of time, that is periods of time in which things can be or happen in ways that are inaccessible to us. To exclude these possibilities, and thus to give perception as solid grounding as possible, these additional four theses were needed. So, Aristotle was quite an optimist regarding perception in order to provide as secure foundation for scientific knowledge as possible, while at the same time avoiding the extreme view of the relativists and the Epicureans who claimed that all perceptions are true.

The preceding discussion allows us to conclude that Aristotle's perceptual optimism was a reaction to two varieties of perceptual pessimism, Democritus' and Plato's. Democritus argued that the ultimate constituents of reality are corporeal (atoms) and that we have only indirect access to them, through reason. This puts a great strain on Democritus' theory of knowledge, of which he seems to have been acutely aware, and which made him something of a cognitive pessimist.[25] Plato, by contrast, taught that the ultimate constituents of reality are incorporeal and ontologically independent of bodies (forms), and he argued that we have direct access to them, through intellect, which made him a cognitive optimist. Aristotle embraced Plato's cognitive optimism, but rejected his perceptual pessimism. This had something to do with the fact that Aristotle agreed with Plato that the ultimate constituents of reality are incorporeal forms, but he disagreed that forms are independent of bodies. If forms are found in bodies, as the organizing principle that determines the shape and behaviour of bodies, forms cannot be discovered and understood except through perception. However, Aristotle readily admits that perception itself is not sufficient for this task. One needs to have intellect, too.[26]

[25] There are various takes on Democritus' epistemology, as one can see from an informative overview in Lee 2005: 188 n. 31, but my claim finds support in several fragments from Diels-Kranz: "In reality we know nothing, for truth is in the depths" (B117); "By this principle man must know that he is removed from reality" (B6); "Yet it will be clear that to know what kind of thing each thing is in reality is impossible" (B7); "That in reality we do not know what kind of thing each thing is or is not has been shown many times" (B8); "The argument too shows that in reality we know nothing about anything, but each person's opinion is something which flows in" (B9).

[26] Earlier versions of this text were presented as a paper at the conference "Experience and Reasoning in Scientific Methodology: Between Antiquity and the Early Modern Period" in Prague (9–11 May 2019) and as an invited lecture at the University of Oslo (13 June 2019). I am grateful to the audiences at both events, especially to Matyaš Havrda and Robert Roreitner in Prague and to Thomas K. Johansen and Franco Trivigno in Oslo, for their incisive comments and encouragement. I owe thanks also to Filip Grgić, Istvan Bodnar, Klaus Corcilius, Stephan Herzberg and Arnold Brooks, whose remarks helped me clarify certain

References

Barnes, J. (ed.). 1984. *The Complete Works of Aristotle. The Revised Oxford Translation*, 2 vols. Princeton: Princeton University Press.

_____, 1987. "An Aristotelian way with scepticism." In M. Matthen (ed.). *Aristotle Today. Essays on Aristotle's Ideal of Science*. Edmonton: Academic Printing and Publishing: 51–65; reprinted in *Proof, Knowledge, and Scepticism*. Oxford: Oxford University Press, 2014: 602–616.

Broadie, S. 1993. "Aristotle's Perceptual Realism." *Southern Journal of Philosophy* 31: 137–159.

Burnyeat, M. F. 1992. "Is An Aristotelian Philosophy of Mind Still Credible? A Draft." In M. Nussbaum and A. Oksenberg Rorty (eds.). *Essays on Aristotle's* De Anima. Oxford: Clarendon Press: 15–26.

Cashdollar, S. 1973. "Aristotle's Account of Incidental Perception." *Phronesis* 18: 156–175.

Caston, V. 1998. "Aristotle and the Problem of Intentionality." *Philosophy and Phenomenological Research* 58: 249–298.

_____, 2005. "The spirit and the letter: Aristotle on perception." In R. Salles (ed.). *Metaphysics, soul and ethics in ancient thought: Themes from the work of Richard Sorabji*. Oxford: Clarendon Press: 245–320.

Diels, H. and Kranz, W. (eds.). 1959. *Die Fragmente der Vorsokratiker. Zweiter Band*. 9th ed. Berlin: Weidmann.

Everson, S. 1990. "Epicurus on the truth of the senses." In. S. Everson (ed.). *Companions to ancient thought 1: Epistemology*. Cambridge: Cambridge University Press: 161–183.

Frede, M. 1996. "Aristotle's Rationalism." In M. Frede and G. Strike (eds.). *Rationality in Greek Thought*. Oxford: Clarendon Press: 157–173.

Gregoric, P. 2005. "Plato's and Aristotle's Explanation of Human Posture." *Rhizai* 2: 183–196.

_____, 2007. *Aristotle on the Common Sense*. Oxford: Oxford University Press.

_____, 2018. "Aristotle's Transparency: Comments on Katerina Ierodiakonou, 'Aristotle and Alexander of Aphrodisias on Colour.'" In B. Bydén and F. Radovic (eds.). *The Parva naturalia in Greek, Arabic and Latin Aristotelianism*. Zurich: Springer: 91–98.

Johnstone, M. A. 2015. "Aristotle and Alexander on Perceptual Error." *Phronesis* 60: 310–338

Koons, B. 2018. "Aristotle's Infallible Perception," *Apeiron* 2018 (ahead of print).

Lee, M.-K. 2005. *Epistemology after Protagoras: Responses to Relativism in Plato, Aristotle, and Democritus*. Oxford: Clarendon Press.

Miller, F. D., Jr. 2018. *Aristotle: On the Soul and Other Psychological Writings*. Oxford: Oxford University Press.

Owens, J. 1981. "Aristotelian Soul as Cognitive of Sensibles, Intelligibles, and Self." In J. R. Catan (ed.). *The Collected Papers of Joseph Owens*. Albany: State University of New York Press: 81–98.

points and avoid some errors, and to two anonymous referees who made helpful suggestions. This work has been fully supported by Croatian Science Foundation under the project IP-2018-01-4966.

Perälä, M. Forthcoming. "Aristotle on Incidental Perception." In C. Thomsen Thörnqvist and J. Toivanen (eds.). *Sense-perception in the Aristotelian Tradition*. Leiden: Brill.

Putnam, H. 2000. "Aristotle's Mind and the Contemporary Mind." In D. Sfendoni-Mentzou (ed.). *Aristotle and Contemporary Science*. Vol. 1. New York: Peter Lang: 7–28.

Reeve, C. D. C. 2017. *Aristotle: De Anima*. Indianapolis–Cambridge: Hackett.

Shields, C. 2016. *Aristotle: De Anima*. Oxford: Oxford University Press.

Striker, G. 1977. "Epicurus on the Truth of Sense Impressions." *Archiv für Geschichte der Philosophie* 59: 125–142; reprinted in *Essays on Hellenistic Epistemology and Ethics*. Cambridge: Cambridge University Press, 1996: 77–91.

Taylor, C. C. W. 1999. *The Atomists: Leucippus and Democritus*. Toronto: University of Toronto Press.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Burge on Mental Causation

MARKO DELIĆ
University of Split, Split, Croatia

The article discusses Tyler Burge's views concerning the debate about the causal efficacy of mental properties, as found in his article "Mind-Body Causation and Explanatory Practice." Burge argues that a proper understanding of kind-individuation and causal explanation in science gives strong prima facie reasons for believing that mental and physical properties are not mutually exclusive. He does so by analysing the strength of two metaphysical theses which standardly underlie the debate—token physicalism and the "Completeness of physics." I present his analysis and argue that without an account of mental causation, his analysis does not support the conclusion that mental and physical properties are not mutually exclusive. Also, I question the methodological adequacy of Burge's analysis for scientific practice.

Keywords: Burge, mental causation, psychology, neuroscience, causation, explanation, overdetermination, properties, kinds.

1. *Introduction*

For several decades now philosophers of mind have been struggling with the question of whether, and how could mental states (and events), in virtue of their mental properties (such as their intentional or phenomenal properties) exert any causal influence on the world. The worry that they do not exert any such influence is suggested by two independently plausible metaphysical theses. First of these, the "Completeness of physics" (CP) states that *all physical effects are fully determined by law by a purely physical prior history* (Papineau 2000: 179). According to the second thesis, the "Irreducibility of the Mental" (IM) mental properties are not reducible to physical properties (Putnam, Fodor). Now, if every physical event (a certain behaviour) is completely determined by prior physical events (states of the body or the central nervous system), and the mental properties (a certain intentional content) of those events cannot be reduced to their physical properties, it seems

that mental properties are excluded from being a possible cause of that piece of behaviour.[1]

The problem has come to be known as the "Exclusion Problem" (EP), most famously associated with the work of Jaegwon Kim. EP presents itself as a problem for the "non-reductivist" types of physicalism. While discussing the causal efficacy of the mental, Tyler Burge develops his views by analysing the so-called "Token-Identity Physicalism" (TIP). Being a non-reductivist position, TIP can generate the EP. In his article "Mind-Body Causation and Explanatory Practice" Burge views the "Mental Causation Debate"[2] (MCD) as a result of metaphysical theses which, contrary to the conviction of many contemporary philosophers, do not possess a justified theoretical motivation in the actual sciences of cognition. He proposes shifting MCD from the terrain of metaphysics to the terrain of actual cognitive sciences, whose practice Burge considers to be the main umpire in MCD. Burge's strategy is twofold. First, Burge tries to undermine TIP on standard externalist intuitions, thus discouraging the motivation for the identification of mental and physical events. Second, he tries to bring the notion of causal powers closer to the epistemic endeavours of actual scientific practice and argue that the indispensability of mental vocabulary in psychological explanation warrants our belief in the efficacy of mental properties and makes MCD redundant.

The first section of the paper will present the EP, and TIP as understood by Burge. The second section will consider Burge's attempts to undermine the theoretical motivation of TIP. The third section will be concerned with Burge's understanding of causal powers and the strength of CP as a premise in EP. Finally, the last section will try to show that, contra Burge, the metaphysical debate about the causal efficacy of the mental (MCD) has real motivation in the actual practice of cognitive sciences. Also, I will argue that Burge's understanding of the "autonomous coexistence" of psychological and neuroscientific explanation cannot be defended without an account of mental causation and giving an adequate account of mental causation forces Burge back into the very thing he aims to undermine—the MCD.

[1] The present article is based on a talk given at the annual conference of the *Society for the Advancement of Philosophy* "New topics in Philosophy" (Zagreb, 2018). Also, I would like to thank Dunja Jutronić, Dario Škarica and Ljudevit Hanžek for commentary and support in writing this article.

[2] The phrase Mental Causation Debate (from now on MCD) is due to Tim Crane (Crane 1995). I will use it to refer to the general debate concerning the causal efficacy of the mental.

2. *The exclusion problem*

Token identity physicalism advocates a weaker type of identity between the mental and the physical. It can be described as consisting of the next two theses:[3]

1) For every mental event x there exists a physical event y such that x=y
2) Mental properties M of x cannot be reduced to physical properties P of y (IM)

The first theses makes TIP a physicalist position by giving ontological priority to the physical (Crane 2003) while the second thesis makes TIP a non-reductive version of physicalism. The second thesis (IM) is suggested by independently plausible theses such as the *multiple realization thesis* (Putnam 1967; Fodor 1974) according to which a single mental type can be realized by different physical types; or the *explanatory gap* which exists between first-person phenomenal descriptions and physical descriptions of those experiences (Nagel 1974; Levine 1983). One way for the exclusion problem to arise is to adopt a view of causation which treats events as being causally efficacious not in virtue of them simply being events but in virtue of the properties these events possess (Crane 2003).[4] Then, if the physical properties of a certain event, a piece of behaviour for example, are caused exclusively by the physical properties of prior events, as CP suggests, it follows that the mental properties of those prior events are excluded from being possible causes of that piece of behaviour.[5]

Burge's way of handling EP is to undermine the metaphysical basis which generates it, thus discouraging MCD. This is expressed in the *motto* at the beginning of his paper:

> Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little. (Burge 2007: 344)

3. *Undermining token-identity physicalism*

Burge's first step is to show that the motivation for identification of mental and physical events is not justified. For Burge, that amounts to showing that the respective sciences, psychology as the study of the mental and neuroscience as the study of the nervous system, individuate their kinds in an essentially different way. While neuroscience in-

[3] Only first of these is necessary for token physicalism. In conjunction with the second it becomes a non-reductive version of physicalism.

[4] This is also how Burge sets the exclusion problem in one version (Burge 2006: 346).

[5] Burge notices that this way of understanding CP leaves open the possibility that, although mental properties cannot be the causes of physical properties, they could nevertheless cause other mental properties. However, this would, as he immediately notices, severely limit the causal efficacy of the mental. After all, mental properties are invoked to explain behaviour.

dividuates kinds *narrowly*, only with respect to the intrinsic, bodily properties, psychological kinds are individuated with reference to their intentional content, which is relational, that is, environmentally dependent. This view is supported by well-known externalist thought experiments. Here, I will use Burge's "aluminium-twalum" example. We imagine a person A whose environment contains the metal aluminium. When A interacts or thinks about aluminium his thoughts are about the aluminium present in his environment. Now, imagine a person B who is physically identical to person B, only whose environment contains a different metal, *twalum*. Twalum is (phenomenologically and practically) indistinguishable (to both A and B) from aluminium, yet it is a metal of a different chemical kind. What seems to follow is that the respective contents of A's and B's mental states is different. While A's thoughts are about aluminium, B's thoughts are about twalum, despite the physical states of A and B being identical (Burge 2007: 316–317).

What is suggested by the thought experiment is that a certain physical event-token, which is a plausible candidate for the identification with a mental event-token, can have different intentional contents on different instantiations. This, according to Burge, is enough to show that mental events cannot be identical to physical events (Burge 2007: 350–351).

Since externalism *per se* is not the topic of this article, we will not concern ourselves with the objections raised against it here. However, in his paper, Burge discusses an objection made to his argument by Donald Davidson. Since Burge's reply reveals his understanding of kind-individuation, it will be illuminating to present it here.

Davidson argues that broad identification of kinds doesn't refute token-identity. He gives the example of a sunburn. While sunburns are individuated environmentally (with reference to an ultraviolet radiating object, most commonly the Sun), it is still plausible to identify every token of a sunburn with a certain physiologically specified state of the skin. Burge agrees with Davidson but rejects his analogy of sunburns and mental events on epistemic grounds. The difference is, Burge argues, that sunburns can be identified in purely physiological terms, without any reference to a potential environmental cause, while mental kinds do not admit such identifications. This in turn, is grounded in the fact that physiological descriptions of sunburns provide *systematic* and *explanatory* ways of individuating sunburns, while descriptions of brain states upon which mental kinds (plausibly) depend do not (Burge 2007: 352–353).

> The system of intentional content attribution is the fundamental means of identifying intentional mental states and events in psychological explanation and in our self-attributions. In fact, we have no other systematic way of identifying such states and events. (Burge 2007: 354)

Physiological properties, unlike intentional properties, do not allow for a systematic and explanatory way of individuating mental kinds.

Therefore, Burge believes it to be theoretically unmotivating to insist on TIP.

4. *Questioning the role of CP*

Similar considerations underlie Burge's analysis of the role CP has in EP. In section 1, CP was presented in terms of event properties. The physical properties of a certain event are completely determined (or have their probabilities completely determined) by the physical properties of prior events. This is the same as saying that an event can cause physical effects only in virtue of its physical properties. To see how Burge understands CP and its strength as a premise in EP, we need to explain how Burge understands causation. Burge believes the best way to understand causation is to see how causal explanation works in actual science. He discusses the notion of *causal powers*. For Burge, the causal powers of an event are determined by the properties which are relevant in describing the patterns of causation in which the kind of that event enters into. Since these patterns are different, depending on the explanatory aims of specific sciences, the properties relevant for describing them will differ too (Burge 2007: 346–347).

Applying the analysis to our present subject, CP amounts to claiming that the patterns of causation identified in physical explanation need to invoke only (and exclusively) physical properties of a certain event. If, however, the patterns of causation are different, as they are in psychology, the restriction that only physical properties are relevant in determining the causal powers of an event becomes unwarranted, since the properties (and thus the causal powers of the event) needed to explain these patterns, change. The only way to insist on such a restriction would be to show that mentalistic (psychological) discourse, that is, the patterns of causation psychology describes, is either non-descriptive or non-causal, which is, as Burge points out, far from being the case (Burge 2007: 347). If the battle is fought on the grounds of actual scientific practices, as Burge believes it ought to be, mental properties are easily defensible.

> If physical events have mental properties, one is not entitled to the view that only physical properties (properties specified in the physical sciences or in ordinary physicalistic discourse) determine all the causal powers of a physical event *(as opposed to merely all the causal powers associated with physicalistic explanations of the physical event*), unless one can show that mentalistic explanation is either non-causal or fails to describe patterns of causal properties. For the causal powers of a physical event that is mental might include possible effects that are specified in mentalistic explanation. No one has shown that mentalistic explanation is either non-causal or non-descriptive. Nor is either view plausible. (Burge 2007: 347)

What this shows is that only if the effect is specified as physical (as belonging to the patterns of causation described by physics) mental properties are excluded. If it is specified as mental (as a piece of inten-

tional behaviour, for example), the mental properties easily find their way back into the adequate explanation.

One problem that can be brought against such an analysis is that it is not certain whether it respects the so-called *No-overdetermination principle,*[6] which is commonly invoked in discussions about mental causation. The principle states that physical effects cannot be overdetermined by physical and mental causes. A certain effect is said to be metaphysically overdetermined if it has two or more sufficient, but metaphysically independent (Loewer 2015: 51). A common example of overdetermination is the case when there are two shooters, each of whom kills a victim. Overdetermination is only possible if the two cause are metaphysically independent. In the case of mental events, however, the mental and physical properties are not metaphysically independent an (Loewer 2015: 51)—hence the name of the principle—*No-overdetermination*. Assuming physicalism, the dependence relation is usually described as that of *supervenience* of the mental properties on physical ones, or *realization*[7] of mental properties by physical properties. Given this assumption, the two types of properties are again in the state of competition for a certain effect. And given CP, the mental properties come out as inefficacious.

Burge rejects this argument by arguing that the view of causation it presupposes simply begs the question against the efficacy of mental properties. The view he has in mind, I believe, is similar to an account of causation which views causation as a matter of "transference of quantities" such as energy and momentum (from now TQ).[8] Burge finds such an account adequate for physical explanation, but problematic as a model for causal explanations as found in psychology:

> Why should mental causes alter or interfere with the physical system if they do not materially consist in physical processes? Thinking that they must, surely depends on thinking of mental causes on a physical model—as providing an extra 'bump' on the effect. The idea seems to be that a cause must transfer a bit of energy or exert a force on the effect. (Burge 2007: 358)

If causation is understood as transference of quantities, mental causes are surely excluded, since, as Burge argues, they do not materially consist in such processes. Presupposing TQ thus begs the question. What is needed to motivate CP in the kind of way which excludes the mental, Burge believes, is to show that a physical model of causation (such as TQ) is appropriate for psychological explanation. Only then would one be in position to claim that the mental and physical somehow interfere or overdetermine a given effect. However, he finds no support for such a view:

[6] I take the term from Heil and Robb (2019), although my construal of the principle here differs from the one they give.

[7] For supervenience, see Kim (1993). For realization Polger and Shapiro (2016).

[8] See Dowe (2000) for a discussion of such theories of causation.

> But whether the physical model of mental causation is appropriate is, again, part of what is at issue. As we have seen, one can specify various ways in which mental causes 'make a difference' which do not conflict with physical explanations. The differences they make are specified by psychological causal explanations, and by counterfactuals associated with these explanations. Such 'differences' made by psychological causes do not require that gaps be left in physical chains of causation. They do not seem to depend on any specific assumptions at all about the physical events underlying the mental causes. (Burge 2007: 358–359)

If the causal explanations found in physics and psychology do not interfere, Burge concludes, there is no reason to believe that the properties these explanations invoke are mutually exclusive, as the principle of *No-overdetermination* suggests.[9]

5. *A weakness in Burge's analysis*

Although being fairly sympathetic with Burges's way of analysing the role CP has in EP, in this section I will try to express my worries with his analysis. I will try to show that his analysis does not support the conclusion that mental and physical properties are not mutually exclusive and thus cannot serve Burge's intention to present MCD as a misguided discussion.

As seen in the previous section, the way Burge tries to dispense with MCD is to show that causal explanation in psychology is not incompatible with a view of causation associated with natural sciences, thus showing that the worries surrounding the *No-overdetermination* principle are badly grounded. He concludes:

> The upshot of this reasoning is that we have no ground for assuming that the failure of mental causes to interfere in the physical chain of events must be explained in terms of mental causes' consisting in physical events. Interference would be surprising, given antecedent assumptions about mental and physical explanation. So non-interference is in no need of explanation in ontological terms. (Burge 2007: 359)

What Burge seems to be claiming is that overdetermination is problematic only if causation is understood on a model such as TQ. Only then the mental and physical somehow interfere or overdetermine a physical effect. But, as Burge argues, since such a model of causation is not appropriate for psychological explanation, one cannot presup-

[9] In his article, Burge makes no reference to TQ specifically. Here I use it since it is suggested by the quoted passage of Burge's article (Burge 2007: 358). Whether this account is appropriate as a model for explanations of physical events is questionable (see Dowe 2000 and Loewer 2015: 54). Furthermore, since the relationship between psychology and neuroscience is what is at issue here, Burge could have picked interventionism (Woodward 2003) which is popular among mechanistically-oriented philosophers of cognitive science. Nothing, however, depends on what account is taken as a model for physical causation, neither for Burge's argument, nor for my objection in the next section. Here, TQ can be understood, somewhat broadly, as standing for any account of causation which will be shown as appropriate for explanations of physical events (in natural sciences).

pose that it excludes mental causation. The only way one could show that TQ excludes the mental would be to show that that explanations in natural sciences and psychology interfere and this is not the case. I agree with Burge on this point. However, I do not agree with the jump Burge makes from the premise that physical and psychological explanations do not interfere to the conclusion that the properties these explanations invoke are not mutually exclusive. To show that this is the case, Burge would have to give an account of causation which would be *both appropriate for psychological explanation and compatible with a model of causation such as TQ*. And such an account is exactly what is missing in Burge's analysis. Without an account of mental causation, I see no justifiable reason for Burge to uphold such a "compatibility" between different types of causation. What I am claiming is that the burden of argument is on Burge. Although he is correct in claiming that TQ is not sufficient to exclude mental causation, he is not also justified in claiming that TQ is compatible with mental causation. Such compatibility requires a positive argument and Burge does not give one. Burge tries to argue that non-interference of psychological and physical explanation warrants our belief in the compatibility of mental and physical causes. But this is not the case. Non-interference of psychological and physical explanation only support a negative conclusion, that is, the conclusion that TQ need not exclude other types of causation. But it does not support the conclusion that TQ is compatible with other types of causation (His insistence on the inadequacy of TQ as model for psychological explanation only makes this point more evident). This conclusion requires an account of mental causation, but Burge fails to provide one.[10]

If this reasoning is correct, it shows that Burge cannot avoid the problems associated with the *No-overdetermination* principle without giving an account of mental causation and its relationship with physical causation. But devising such an account of causation for psychological explanation amounts to nothing less than collapsing back into MCD, the very thing that Burge tried to present as futile. To see that this is so, one needs only notice that a huge portion of the literature concerning MCD is explicitly devoted to developing accounts of causation which would satisfy these constraints (Crane 2006: 1124). Such are, for example, accounts of causation which appeal to counterfactual dependence (Loewer 2015), structural causes (Dretske 1988) or program explanation (Jackson and Petite 1990). Each of these aims to provide an account of causation which would make mental causation compatible with a physical model of causation.

[10] This objection is similar to the worries Kim has with Burge's analysis: "The issue is how to make our metaphysics consistent with mental causation, and the choice that we need to make is between various metaphysical alternatives, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other" (Kim 1998: 62).

The problem with Burge's attempt to present MCD as a misguided discussion can thus be formulated as a dilemma. If Burge does not provide an account of mental causation, he lacks a strong argument for the compatibility of mental and physical causation. On the other hand, if he tries to provide such an account, he ends up participating in MCD.

In addition, some portions of Burge's article can even be interpreted as occupying a position in MCD. Burge's central view, according to which the causal efficacy of properties depends on the way events are kind-individuated overwhelmingly reminds of the so-called *dual explanandum* versions of mental causation, according to which the mental and physical properties of an event are causally efficacious for different properties of the effect.[11] At one point, Burge says:

> ...we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. (Burge 2007: 359)

or in the passage quoted in the previous section, when Burge says the following:

> For the causal powers of a physical event that is mental might include possible effects that are specified in mentalistic explanation. (Burge 2007: 347)

If so, Burge's analysis inherits the problems associated with these kinds of strategies. Unsurprisingly, the main problem dual explanandum strategies face is the worry that they do not respect the *No-overdetermination* principle (Robb and Heil 2019), the very same principle, I argued, Burge cannot adequately overcome without giving an account of causation both appropriate for psychology and compatible with physical models of causation.

6. *Conclusion*

I will conclude by a brief methodological consideration. At several points in his article, Burge expresses his belief in some sort of metaphysical dependence between the mental and the physical:

> There is certainly reason to believe that underlying our mental states and processes are physical, chemical, biological, and neural processes that proceed according to their own laws. Some such physical processes are probably necessary if intentional (or phenomenal) mental events are to be causes of behavior. (Burge 2007: 349)

On the other hand, however, his analysis of TIP and MCD supports a view of psychology and neuroscience as being importantly different scientific enterprises, whose taxonomies and explanations differ in important, even unbridgeable ways. The problem with such a view is that it leaves the relationship between psychology and neuroscience highly problematic. Since both of these are concerned with understanding cognition and behaviour, to insist that they can proceed on completely

[11] See Robb and Heil (2019) for an overview of such approaches.

separate courses would be to advocate something in the spirit of averrosian "double truth theory," according to which it is possible for religion and philosophy to arrive at mutually contradicting but true knowledge. At one point, Burge seems to be fine with such a view:

> Maybe science will never make use of anything more than limited correlations with the lower, more automatic parts of the cognitive system. Maybe identities or part–whole relations will never have systematic use. Maybe the traditional idea of a category difference will maintain a presence in scientific practice. (Burge 2007: 360)

Some philosophers argue for a different picture of the relationship between psychology and neuroscience. For example, Bechtel (2008: 71) argues that cognitive scientists use identities between mental and brain processes as heuristics which then serve to improve both the psychological and neuroscientific research. In a similar spirit, Polger and Shapiro (2016: 168–169), following Churchland (1986) see the relationship between the two as one of coevolution and interplay in which both behavioural experiments and neuroscientific manipulations converge in advancing our understanding of cognitive phenomena. If that is the case, as Burge himself seems to accept at one point (Burge 2007: 381), then the various problems which MCD identifies, such as the relationship between different causal models or the question of metaphysical dependency, far from being theoretically unmotivating, come out as important in understanding the relationship between psychology and neuroscience. However, what this relationship will turn out to be, I agree with Burge, is an open question.

References

Bechtel, W. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.

Burge, T. 2007. *Foundations of Mind: Philosophical Essays, Volume 2*. New York: Oxford University Press.

Crane, T. 1995. "The Mental Causation Debate." *Proceedings of the Aristotelian Society* 69: 211–36.

_____, 2003. "Mental Causation." In L. Nadel (ed.). *Encyclopedia of Cognitive Science*. London: Macmillan: 1120–1125.

Churchland, P. 1986. *Neurophilosophy*. Cambridge: MIT Press.

Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.

Fodor, J. 1974. "Special Sciences: Or the Disunity of Science as a Working Hypothesis." *Synthese* 28: 97–115.

Jackson, F. and Petit, P. 1990. "Program Explanation: A General Perspective." *Analysis* 50: 107–117.

Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

_____, 1998. *Mind in a Physical World*. Cambridge: MIT Press.

Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354–361.

Loewer, B. 2015. "Mental Causation: The free lunch." In T. E. Horgan, M. Sabates, and D. Sosa (eds.). *Qualia and mental causation in a physical world: Themes from the philosophy of Jaegwon Kim*. Cambridge: Cambridge University Press.

Nagel, T. 1974. "What is it like to be a Bat?" *Philosophical Review* 83: 435–456.

Papineau, D. 2000. "The Rise of Physicalism." In M. Stone and J. Wolff (eds.). *The Proper Ambition of Science*. New York: Routledge: 174–208.

Polger, T. W. and Shapiro, L. A. 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.

Putnam, H. 1975. "The Nautre of Mental States." In H. Putnam (ed.). *Mind, Language and Reality: Philosophical papers*, vol 2. Cambridge: Cambridge University Press: 429–440.

Robb, D. and Heil, J. 2019. "Mental Causation." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), forthcoming URL = <https://plato.stanford.edu/archives/sum2019/entries/mental-causation/>.

Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Heda Festini's Contribution in the Research of Croatian Philosophical Heritage

LUKA BORŠIĆ and IVANA SKUHALA KARASMAN
Institute of Philosophy, Zagreb, Croatia

In this text we offer an overview of Festini's works on history of Croatian philosophy. The article is divided in five parts in which we discuss Festini's attitude towards Croatian Renaissance philosophers, eighteenth and nineteenth century Croatian philosophers, and two philosophers from the twentieth century (Vladimir Filipović and Marija Brida). Majority of Festini's texts were published in the journal Prilozi za istraživanje hrvatske filozofske baštine.

Keywords: Heda Festini, Croatian philosophy, history of philosophy.

1. *Introduction*

Heda Festini devoted a great part of her writings and research to the history of Croatian philosophy. It should be particularly emphasized that she published articles on some well-known philosophers, such as the versatile Renaissance philosopher Frane Petrić, as well as on some unexplored and almost unknown Croatian philosophers, such as Pietro Botturin, Antun Petrić, Juraj Politeo, Albin Nađ and Jure Pulić. The originality of her approach set the bar quite high and gave general directions on how to do research in the history of Croatian philosophy for researchers to come.

2. *Festini on the Renaissance Croatian philosophers: Grisogono, Petrić and Skalić*

In 2009 in the journal *Filozofska istraživanja* Festini's article on the Renaissance philosopher Federik Grisogono (Zadar, 1472—Zadar, 1538) was published under the title "Grisogonov iskoračaj u novu znanost" ("Grisogono's leap forward towards a new science"). In it Festini claims

that in Grisogono's book *Astronomsko zrcalo* (*Astronomical Mirror*) we can find some features of modern science and for this reason Grisogono should be considered to be one the most important Croatian philosopher since he anticipated some of the changes that will happen in science. In her conclusion, Festini lists several features of Grisogono's philosophy that are akin to some of the features of modern science. Particularly, his geometry shows traces of a non-Euclidean approach, as well as his understanding of force (Festini 2009: 732).

A lager part of Festini's opus is devoted to the Renaissance philosopher from Cres, Frane Petrić (Cres, 1529—Rome, 1597). The first Festini's article on Petrić, "Frane Petrić o principima historijskog istraživanja iz perspektive problematičke povijesti" ("Frane Petrić on the principles of historical research from the perspective of problematic history"), was published in *Prilozi za istraživanje hrvatske filozofske baštine* (further, *Prilozi*) in 1979. In this article Festini shows how Petrić's interpretation of history in its problematic environment is reflected in his texts.

In 1995 Festini published "Još jedan pokušaj talijanizacije Petrića" ("One more attempt at Italianization of Petrić") in the journal *Filozofska istraživanja*. In it Festini responds to a writer from Rijeka, Giacomo Scotti, who published a text in which he argued that Petrić was an Italian philosopher. Festini's arguments were twofold. On the one hand, the fact that Patrić wrote only in Latin and Italian does not contradict his Hercegovinian origin, as Festini claims, since these were the languages of scientific and academic communication at the time. On the other hand, Festini poses a question about Giacomo Scotti's academic credibility who was not trained in history of philosophy.

The article "Perspektive ekološke teorije i Petrićev svjetonazor" ("Perspectives of ecological theory and Petrić's world-view") was published in *Filozofska istraživanja* in 1996. Here Festini analyzes two of Petrić's works *Ten Dialogues on History* and *New Universal Philosophy:* they "mark his world-view as a possible inclination toward the second ecological tendency" (Festini 1996: 39).

Festini's article "Platonova koncepcija o učenju / neučenju vrline – Petrić" ("Plato's concept of learning / not learning of virtue – Petrić") was published in *Prilozi* in 2003. In it Festini emphasizes a certain tension in Petrić's ethical theory. In some texts Petrić accepts Plato's aristocratic approach with the idea of the good and as the measure which is foundation of the doctrine of virtue. On the other hand, Petrić also upholds a more democratic stance: the idea of equality in community which najes Petrić a modern thinker (Festini 2003: 26).

Another article on Petrić, "Tragom utilitarizma u Petrića" ("Tracing back Petrić's utilitarism") was published in *Prilozi* a year later, in 2004. Some traces of utilitarianism can be found in Petrić's works *La città felice* (1553) and *L'Amorosa filosofia* (1577) in which the term 'filautia' (self-love) prevails. Petrić shows that self-love is the source

of all other feelings, and it is also the way to pursuit the virtue, which Festini interprets as a utilitarian position.

In 2009 the article entitled "Petrić i Acastos" ("Petrić and Acastos") was published. The main point that Festini makes is the comparison of Plato's and Petrić's ethical theory inspired by Iris Murdoch's two Platonic dialogues: "Art and Eros" and "Above the Gods", both published in her book *Acastos* from 1987. In the dialogues, Socrates and Platon make an appearance, as well as several fictional characters, one of whom is Acastos. In her text Festini argues that Petrić, although he was a declared Platonist-Pythagorean, also contributed to the disintegration of classical ethical virtue by some utilitarian interventions. More particularly, Petrić, according to Festini, came close to the idea that artistic creation contains all religion, morality, and justice. Festini does that by comparing Petrić to the fictional character Acastos.

In 2010 Festini published another article on the same topic, "Petrić i Acastos, nastavak prvi" ("Petrić and Acastos, part one"). In this text Festini, based on the previous text, analyses the fifteen selected (translated) Petrić's texts in the book by Ljerka Schiffler entitled *Frane Petrić o pjesničkom umijeću* (*Frane Petrić on Poetic Art*, Zagreb: Institut za filozofiju, 2007). According to Festini, Petrić's poetics contains some elements of Aristotelianism, which would connect Petrić with not only Baroque and Mannerism but also modernist aesthetics.

Three years later (2012) Festini published an article under the title "Frane Petrić o Empedoklu pjesniku: Petrić i Acastos, nastavak drugi" ("Frane Petrić on Empedoclus: Petrić and Acastos: part two"), again in *Prilozi*. Here Festini goes into details in explaining Petrić's critique of Aristotle's claim that Empedoclus was not a poet but a physiologist. Based on this, Festini draws two conclusions: 1. Petrić, by insiting that the essential part of poetry is form rather than its matter, inserts an element of Aristotelianism into Platonism; and 2. Petrić defends didactive poetry which makes him utilitarian.

In 2013 the article "Petrićeva *La deca semisacra* kao moguće kodificiranje morala" ("Petrić's *La deca semisacra* as a possible codification of morality") was published. Here Festini claims that Petrić's utilitarianism overcomes Plato's teaching of virtue through its two components—social and psychological. According to Festini, Petrić defends a natural path of developing virtue from exercising good laws in a just state to the experience of moral poetry.

The last article published on Petrić in *Prilozi* was "Historiografija—najslabija karika u Petrićevu lancu znanostī" ("Historiography—the weakest link in Petrić's chain of sciences") that appeared in 2016. In this article Festini proposes two fresh insights: "1. a well-grounded view of the place of mathematics and history in Petrić's science chain, 2. explanation of the terminological distinction between *cagione* and *causa* from the perspective of Petrić's *Ten Dialogues of History*. A parallel between the mentioned insights and the research into Petrić's approach

to history conducted up today contributes to a more solid interpretation of his antideterministic understanding of history" (Festini 2016: 292).

Finally, on Pavao Skalić Fesini published an article under the title "Pavao Skalić i znanost" ("Pavao Skalić and science") in *Prilozi* in 2010. The text is an analysis of Skalić's book *Epistemon* (1559, 1571). According to Festini, Skalić's uniqueness lies in his understanding of science as evidence and experience. Moreover, he was, according to Festini, an "early modern" thinker in emphasizing the usefulness of science for everyday life.

3. *Festini on some eighteenth-century Croatian philosophers: Bošković and Botturin*

Festini's article on Ruđer Bošković (Dubrovnik, 1711—Milan, 1787) was published in *Prilozi* in 2017 under the title "Što je doista indukcija u Ruđera Boškovića?" ("What is really induction for Ruđer Bošković?"). Here Festini argues that "[i]nduction in the works of Ruđer Bošković is a research topic with extensive tradition. This article aims to place Bošković's views on induction within Fermat-Pascal interpretative tradition of induction, whose protagonists were Jakob Bernoulli and Thomas Bayes, along with Wittgenstein, Carnap and Hintikka in the twentieth century" (Festini 2017: 435).

Festini also published two articles on a less known philosopher Pietro Botturin (Malcesinama, 1779—Zadar, 1861). First article entitled "Botturina koncepcija značenja i suvremena lingvistika" ("Botturin's concept of meaning and contemporary linguistics") was published in *Prilozi* in 1978. The focus of her article is Botturin's book *Ideaologia* published in 1832 which represents "a unique attempt on the philosophical foundation of human speech" (Festini 1978: 157). His goal was to interpret words which as audible-figurative signs have no meaning for themselves. The second article, "Botturina teorija jezika" ("Botturin's theory of language"), published in 1982, appeared in *Prilozi* too. In this article Festini's again analyses Botturin's book *Ideaologia*. Here Festini concludes that in Botturin's theory of language, which is concerned with the origin and the evolution of language, he synthesized empirical and illuminist tradition under the influence of Wolff, Leibniz, Condillac, Bacon, Vico and Lock.

4. *Festini on some nineteenth century Croatian philosophers: Politeo, Petrić, Nađ and Pulić*

Heda Festini also researched three less known nineteenth century Croatian philosophers: Juraj Politeo, Antun Petrić and Albin Nađ.

Juraj Politeo was very much in the focus of here interest and Festini authored two monographs on him. The first book Festini published on Politeo was in 1977 under the title *Život i djelo Splićanina Jurja Politea* (*Life and Work of Juraj Politeo from Split*), published in Zagreb.

The second, more extended version of the book with the same title was published also in Zagreb in 2003.

Furthermore, in *Filozofska istraživanja* in 2006 Festini published an article on Politeo and Albin Nađ under the following title: "Juraj Politeo i Albin Nađ, prethodnici Einsteina?!" ("Juraj Politeo and Albin Nađ, precursors of Einstein?!"). In this articule Festini concludes: "Juraj Politeo (1827—1913) was a precursor of Einstein because of his contribution to the reassessment of scientific concepts, laws, and the objects of scientific study. Albin Nađ (1866—1901) also contributed to the reassessment of scientific concepts, laws, and the objects of scientific study. Both of our thinkers gained merit with their detailed reflections about relativity, and Nađ especially in considering the relativity of space" (Festini 2006: 593).

The article "Juraj Politeo: jezik i mišljenje" ("Juraj Politeo: language and thought") was published in the journal *Filozofska istraživanja* in 1993. In this article Festini shows the way in which Politeo dealt with the problem of the relationship between language and thought. Festini argues that Politeo's concepts of this relationship is similar to the present-day discussions in philosophy of language. Politeo had "[…] original standpoint about the relation between language and thought, because they did not reduce to one another. According to Politeo, the being of the soul is the source of them, but partly with the contrastive result" (Festini 1993: 808).

Festini published the article "Politeova Plava bilježnica (1879–1880). O nacrtu neodržanog predavanja na Sveučilištu u Padovi" ("Politeo's Blue Notebook (1879–1880). On a unedited lecture at the University of Padua") in *Prilozi* in 1994. In this article Festini describes the so-called "Blue Notebook" which contained the text of a lecture Politeo prepared but never held, after having published previous nine lectures.

The article "Politeova 'smeđa bilježnica' (1860.). Moral—sloboda" (Politeo's 'Brown Notebook' (1860). Morality—Freedom") was published in *Filozofska istraživanja* in 1994. Here Festini argues that the Brown Notebook contains eleventh lecture which he began writing in 1880.

One year later another article on Politeo was published under the title "Etički naturalizam kao ekoteorija. (O natuknicama u Politeovim spisima)" ("Ethical naturalism as an ecotheory. (On the footnotes in Politeo's writings)", also in *Prilozi*. In this text Festini shows that Politeo's writings contain some footnotes which anticipate some problems typical for the ecotheoretical questions.

The article "Politeova misaona krivulja: 1845–1913 (Rani spisi: 1845–1859)" ("Polite's thought curve: 1845–1913. (Early writings: 1845–1859))" was published in *Prilozi* in 1996. The aim of this article was to show the beginning period of Politeo's opus. One year later the article "Politeova misaona krivulja: 1845–1913. (Srednje razdoblje: 1860–1889)" ("Polite's thought curve: 1845–1913. (Middle period: 1860–1889))" was published in *Prilozi,* too. The most significant characteristic of the central period of Politeo's intellectual development was

the publication of his *Genesi naturale di un' idea* and the lectures he held at the University of Padua in 1878/79. These two articles were followed by an article published in *Prilozi* in 1998 under the title "Politeova misaona krivulja: 1845–1913. (Kasno razdoblje: 1890–1913)" ("Politeo's thought curve: late period 1890–1913)". In this middle period Politeo continues developing his main idea: he tries to explain history of mankind as a progression from an anxious and unconscious mode of living towards a civilized world.

In 1999 Festini's article "Politeov temelj za Milovu logiku" ("Politeo's foundation for Mill's Logic") was published in *Prilozi*. The aim of this article is a systematic approach to Politeo's thoughts as they can be found scattered around his manuscript legacy.

The last article on Politeo was published in 2008, also in *Prilozi* "Kada analiziram Franu Petrića (1529–1597), zašto mislim na Jurja Politea (1827–1913)?" ("When I analyse Frane Petrić (1529–1597) why do I think about Juraj Politeo (1827–1913)?". In this article Festini's position is summarized in the following way: "This interpretative trajectory is possible because by adopting classical utilitarian views, both conceived of human beings as subject to high moral standards. By developing such forms of utilitarianism, they pointed in the direction of contemporary views, such as, for example, Singer's preference utilitarianism" (Festini 2008: 68).

As a summary of Festini's view on Politeo, Festini stresses that Juraj Politeo (Split, 1827—Venice, 1913) dealt with the topics that were untypical for academic philosophy in the nineteenth century Croatia: he introduced a new phenomenological method and he focused his philosophy on the concept of inner life with a special emphasis on the role of instinct and unconsciousness as a primordial basis of psychological life. Although Politeo was innovative and to a degree an original thinker his philosophy was unduly neglected until Festini published her book on Politeo's philosophy in 1977. In her book Festini concludes: "In a nutshell, this old philosopher can be read with interest even today and it does not happen very rarely that some of his attitudes or topics can still inspire new researches, i. e., bring forth new ideas" (Festini 1977: 194).

Antun Petrić (Komiža, 1829—Komiža, 1908) was a philosopher of moderate rationalism. He devoted his research to the problems of freedom and aesthetics, i.e., beauty. Although he was not always consistent, as Festini claims, he did leave a mark in the nineteenth century Croatian philosophy.

In 1976 Festini published a text: "A. Petrić, filozof umjetnosti i slobode" ("A. Petrić, philosopher of art and freedom"). A. Petrić was, according to Festini, primarily a philosopher of aesthetics. His starting point was Gioberti's work *Il bello*: "As a critic of Gioberti Petrić conformed to the old metaphysical theses (the objectivity and absoluteness of the beautiful), but sporadically he stood apart from some of those

theses (renouncing the intrusion of theology in aesthetic reasoning and rejecting the rationalisation of artistic creation)" (Festini 1976: 133).

Her book *Antun Petrić: filozof iz Komiže* (*Antun Petrić, a philosopher from Komiža*), published in Zagreb in 1992, is supplemented with Festini's very useful partial translation of Antun Petrić's Italian works. In 1992 her text "Interpretacija lijepog u Ante Petrića" ("Interpretation of beauty by Ante Petrić") was published. Festini argues that Petrić's aesthetics "[…] appears as a conglomerate of many contradictions which contains all Romanesque failures of that time but also shows a sincere effort to penetrate to the being of the beauty" (Festini 1992: 215–216).

Festini's article on Albin Nađ, "Logistika Trogiranina Albina Nađa" ("Logistics of Albin Nagy from Trogir"), was published in *Prilozi* in 1975. That was the first text ever published on this philosopher. According to Festini, Albin Nađ (Trogir, 1866—Taranto, 1901) was a very talented philosopher whose ideas on mathematical logic were surprisingly modern. Festini writes: "Especially impressive are the results of this logistic conception in the field of the philosophy of science and in the anticipation of the new methodology, arrived at by its new rationalistic orientation" (Festini 1975: 138).

In 1999 the article "Znanje o jeziku u Jure Pulića (Dubrovnik, 1816.—Rome, 1883.)" ("Knowledge of language in Jure Pulić") was published in the journal *Scopus*. In this article Festini claims that Pulić anticipated almost all three stages of scientific research which were indicated by Ch. S. Peirce (1839—1914).

A few years later, in 2005, the article "O nekim rezultatima i novim zadacima u istraživanju hrvatske filozofske baštine" ("About some results and new tasks in the research of Croatian philosophical heritage") was published. In this article Festini analyses Pulić's fascination with Botturin and the Croatian bishop and benefactor, Josip Juraj Strossmayer. Pulić developed a philosophical appreciation of morally strong personalities which were shaped by adopting the habit of thoughts, who could not claim their right if they haven't fulfilled they duties first (Festini 2005: 264).

5. *Festini on some twentieth-century Croatian philosophers: Filipović and Brida*

In 1985 Festini published an article on her friend and teacher Vladimir Filipović (1906—1984) in *Prilozi*. In this article, under the title "Vladimir Filipović—profesor zagrebačkog Filozofskog fakulteta i odsjeka za filozofiju u Zadru" ("Vladimir Filipović—Professor at the Philosophical Faculty in Zagreb and the Department of Philosophy in Zadar"). In this two-page short text Festini describes Filipović's professorship at the Department of Philosophy of the University of Zadar.

In the book *Vladimir Filipović: život i djelo (1906–1984)* (*Vladimir Filipović: Life and Work (1906–1984))*, published by the Insti-

tute of Philosophy in 2008, Festini published a chapter "Dr. Vladimir Filipović—baština za generacije" ("Dr. Vladimir Filipović—heritage for generations"). In this text she puts emphasis on three of Filipović's contributions to the Croatian philosophical heritage. The first is that Filipović paved the path for methodology of how the past Croatian philosophers should be dealt with in contemporary philosophical and societal movements. His second big contribution was the establishing the Department of Philosophy in Zadar. And thirdly and according to Festini, most importantly, he established the journal *Prilozi za istraživanje hrvatske filozofske baštine* in 1975 which is still published by the Institute of Philosophy in Zagreb.

One of Festini's last published text was "Marija Brida (1912. –1993.) o H. Bergsonu" ("Marija Brida on H. Bergson") (Boršić and Skuhala Karasman 2017: 177–184). The text is dedicated to her friend from the University of Zadar, the Croatian woman philosopher Marija Brida. In this text Festini deals with Brida's "Introduction" to Bergson's book *Ogledi o neposrednim činjenicama svesti* (*Essai sur les donnés immédiates de la conscience*) published in 1978 in Belgrade. Festini claims that in this "Introduction" Brida gave contemporary interpretation of Bergson's book *Ogledi o neposrednim činjenicama svesti* but that she also succeeds to evaluated his philosophy as "intuitivism". Furthermore, Festini notices that Brida equally praise and criticises Bergson, although she agrees with him in the perspective of mysticism.

In 1994 Festini also published a review of Brida's posthumously published book *Misaonost Janka Polića Kamova* (*Thoughtfulness of Janko Polić Kamov*). In her review, after a thorough analysis of Brida's work, Festini concludes that the book is "extraordinarily stimulating".

At the end it is necessary to say a few words about Festini's understanding of the future of Croatian philosophy. In her article "**O nekim rezultatima i novim zadacima u istraživanju hrvatske filozofske baštine**" ("About some results and new tasks in the research of Croatian philosophical heritage") published in 2005 Festini claims that Croatian philosophical heritage is not sufficiently explored, especially the nineteenth century philosophers. Furthermore, she states that Croatian philosophers are more known outside Croatia than in Croatia. Festini concludes that there is enough work for younger generations that are interested in studying Croatian philosophy.

References

Festini, H. 1975. "Logistika Trogiranina Albina Nađa." *Prilozi za istraživanje hrvatske filozofske baštine* 1: 75–138.

_____, 1976. "Petrić filozof umjetnosti i slobode." *Prilozi za istraživanje hrvatske filozofske baštine* 2: 101–134.

_____, 1977. *Život i djelo Splićanina Jurja Politea*. Zagreb: Institut za filozofiju Sveučilišta u Zagrebu.

_____, 1978. "Botturina koncepcija značenja i suvremena lingvistika." *Prilozi za istraživanje hrvatske filozofske baštine* 4: 157–180.

_____, 1979. "Frane Petrić o principima historijskog istraživanja iz perspektive problematičke povijesti." *Prilozi za istraživanje hrvatske filozofske baštine* 5: 27–32.

_____, 1982. "Botturina teorija jezika." *Prilozi za istraživanje hrvatske filozofske baštine* 8: 75–92.

_____, 1985. "Vladimir Filipović – profesor zagrebačkog Filozofskog fakulteta i odsjeka za filozofiju u Zadru." *Prilozi za istraživanje hrvatske filozofske baštine* 11: 207–208.

_____, 1992. *Antun Petrić: filozof iz Komiže*. Split: Književni krug.

_____, 1992. "Intrepretacija lijepog u Ante Petrića." In Z. Posavac (ed.). *Hrvatska filozofija u prošlosti i sadašnjosti*. Zagreb: Hrvatsko filozofsko društvo: 215–235.

_____, 1993. "Marija Brida, *Misaonost Janka Polića Kamova*"—recenzija?

_____, 1993. "Juraj Politeo jezik i mišljenje." *Filozofska istraživanja* 13: 803–808.

_____, 1994. "Politeova Plava bilježnica (1879–1880). O nacrtu neodržanog predavanja na Sveučilištu u Padovi." *Prilozi za istraživanje hrvatske filozofske baštine* 20: 243–253.

_____, 1994. "Politeova 'smeđa bilježnica' (1860.). Moral – sloboda." *Filozofska istraživanja* 14: 137–142.

_____, 1994. "M. Brida, *Misaonost Janka Polića Kamova*, Izdavački centar Rijeka , 1993, str. 125." *Prilozi za istraživanje hrvatske filozofske baštine* 20: 494–496.

_____, 1995. "Etički naturalizam kao ekoteorija. (O natuknicama o Politeovim spisima)." *Prilozi za istraživanje hrvatske filozofske baštine* 21: 291–300.

_____, 1995. "Još jedaan pokušaj talijanizacije Petrića." *Filozofska istraživanja* 15: 191–195.

_____, 1996. "Perspektive ekološke teorije i Petrićev svjetonazor." *Filozofska istraživanja* 16: 33–39.

_____, 1996. "Politeova misaona krivulja: 1845–1913 (Rani spisi: 1845–1859)." *Prilozi za istraživanje hrvatske filozofske baštine* 22: 343–359.

_____, 1997. "Politeova misaona krivulja: 1845–1913. (Srednje razdoblje: 1860–1889)." *Prilozi za istraživanje hrvatske filozofske baštine* 23: 147–181.

_____, 1998. "Politeova misaona krivulja: 1845–1913. (Kasno razdoblje: 1890–1913)." *Prilozi za istraživanje hrvatske filozofske baštine* 24: 145–183.

_____, 1999. "Politeov temelj za Millovu logiku." *Prilozi za istraživanje hrvatske filozofske baštine* 25: 157–190.

_____, 1999. "Znanje o jeziku u Jure Pulić (Dubrovnik, 1816.—Rome, 1883.)." *Scopus* 4: 79–85.

_____, 2003. "Platonova koncepcija o učenju / neučenju vrline – Petrić." *Prilozi za istraživanje hrvatske filozofske baštine* 29: 19–26.

_____, 2003. *Život i djelo Splićanina Jurja Politea*. Zagreb: Hrvatsko filozofsko društvo.

_____, 2004. "Tragom utilitarizma u F. Petrića." *Prilozi za istraživanje hrvatske filozofske baštine* 30: 59–67.

_____, 2005. "O nekim rezultatima i novim zadacima u istraživanju hrvatske filozofske baštine." *Prilozi za istraživanje hrvatske filozofske baštine* 31: 261–272.

_____, 2006. "Juraj Politeo i Albin Nađ, prethodnici Einsteina?!" *Filozofska istraživanja* 26: 585–593.

_____, 2008. "Kada analiziram Franu Petrića (1529–1597), zašto mislim na Jurja Politea (1827–1913) ?" *Prilozi za istraživanje hrvatske filozofske baštine* 34: 55–68.

_____, 2008. "Dr. Vladimir Filipović – baština za generacije." In E. Banić-Pajnić, M. Girardi-Karšulin and Lj. Schiffler (eds.). *Vladimir Filipović: život i djelo (1906–1984*). Zagreb: Institut za filozofiju: 35–41.

_____, 2009. "Grisogonov iskoračaj u novu znanost." *Filozofska istraživanja* 29: 725–732.

_____, 2009. "Petrić i Acastos." *Prilozi za istraživanje hrvatske filozofske baštine* 35: 45–53.

_____, 2010. "Pavao Skalić i znanost." *Prilozi za istraživanje hrvatske filozofske baštine* 36: 39–48.

_____, 2010. "Petrić i Acastos, nastavak prvi." *Filozofska istraživanja* 30: 451–456.

_____, 2012. "Frane Petrić o Empedoklu pjesniku: Petrić i Acastos, nastavak drugi." *Filozofska istraživanja* 38: 79–83.

_____, 2013. "Petrićeva *La deca semisacra* kao moguće kodificiranje morala." *Filozofska istraživanja* 39: 27–33.

_____, 2016. "Historiografija – najslabija karika u Petrićevu lancu znanostī." *Prilozi za istraživanje hrvatske filozofske baštine* 42: 283–292.

_____, 2016. "Što je Petrić u trećem svesku *Peripatetičkih rasprava – filozof ili teolog?*" *Prilozi za istraživanje hrvatske filozofske baštine* 42: 69–81.

_____, 2017. "Što je doista indukcija u Ruđera Boškovića?" *Prilozi za istraživanje hrvatske filozofske baštine* 43: 417–436.

_____, 2017. "Marija Brida (1912.–1993.) o H. Bergsonu". In L. Boršić and I. Skuhala Karasman (eds.). *Filozofkinje u Hrvatskoj*. Zagreb: Institut za filozofiju: 177–184.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Constructing a Happy City-State. In memoriam Heda Festini

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

The paper honors Heda Festini; it's first part contains author's personal memories of Heda. The central part of the paper addresses a favorite author of Heda Festini, Franjo Petrić, and his Utopia The Happy City-State. *It then places the utopian construction on the map of contemporary understanding of political theorizing. Utopias, like the one due to Petrić, result from thought-experimenting; in contrast to purely epistemic thought-experiments they are geared to "guidance", as Petrić puts it, namely advice giving and persuading. Political thought-experimenting can be understood to a large extent as work in ideal theorizing; a matter little noticed in the literature. Classical cases cover "ideal theory" in the sense of given, non-temporal arrangement; "ideal" either in a very limited sense of strict compliance (Rawls), or in a wider sense of normatively marked properties, not instantiated in actual political reality. Platonic tradition belongs to a third genus, "ideal" in the sense of recommended end-state; Utopias add to this theoretical quality the dimension of "guidance", so that they are motivational, time-related ideal theories. The paper depicts these relations between thought-experimenting as a wider genus, and ideal theorizing as its prominent political-philosophical sub-species. The paper is thus a tribute to Heda Festini who helped me find my way to analytic theorizing, and help analytic philosophy to start serious institutional life in our native Croatia.*

Keywords: Franjo Petrić, *The Happy City-State*, Renaissance Utopia, ideal theorizing, political thought experiments.

1. Introduction

The paper honors Heda Festini; at the same time important historian of philosophy, and the head of Zadar philosophy department, who has started the most successful line of analytic philosophy teaching in Croa-

tian (and has helped me enormously in my philosophical career). The paper follows these lines; I shall refer to her simply as "Heda". The following section is dedicated to Heda's work in the Zadar department, which I retell as part of precious personal memories. The next section briefly summarizes Petrić's *The Happy City-State,* stressing his obsession with health, individual, environmental and social-political, reminiscent of contemporary green ideologies. The fourth section turns to theory, and attempts to place Petrić on the map of ideal theorizing. In order to do this, it places ideal theories within the framework of thought experimenting and proposes a fresh taxonomy of ideal theories, stressing two elements that have been absent from the literature: the specificity of motivational ideals, characterizing Utopias, from Moore and Petrić to socialist utopias, and the functioning of dystopias as a kind of (anti-)ideal theories. The conclusion returns to Heda's reading of Petrić, stressing her original proposal to see him as an early utilitarian.

2. *Memories of Heda Festini*

So, let me start briefly with my personal memories. The encounter with Heda that has changed my life happened in spring of 1975. At that time I had worked at the Medical faculty in Rijeka, teaching "Marxism", and I was avidly looking for a university job in philosophy. So, at a conference in Ljubljana I met Heda Festini, who was in company of her colleague Saša Kron, a first-rate logician from Belgrade. I had a presentation on philosophy of Althusser, fresh from a meeting with him in Paris. After the presentation I joined Heda and Saša. I had no idea they were commenting my paper; suddenly Heda asked mi if I would come to work in Zadar, and it was obvious that Saša was very much in favor of this offer.

"When?" I asked. "Soon, in the fall, if you want." I accepted with enthusiasm, and this decision has shaped my professional life from then on. Thanks to Heda, and Saša, I got the job in philosophy, at the age of twenty five. So, I ended up in Zadar.

The local philosophy department was ruled by two lady philosophers, Heda, the younger of the two, and Marija Brida, the senior, supported by old and tired Anđelko Habazin. They both took care of me, way beyond any formal obligations. Heda tried to persuade me to do more exercise to get rid of my asthma, she loved sport and exercise in general. She was taking care of me all the time I was in Zadar. You can guess how much all this meant emotionally for me from the fact that my daughter got her name Heda after Heda Festini.

Philosophically the most important component of the story was Heda's interest and enthusiasm for analytic philosophy. She was working all her life on the history of Croatian philosophy, but the area where she left the deepest trace was the creation of analytic tradition in Zadar. In the seventies, I become disappointed with French continental philosophy fashion(s) and was looking for a new area. Heda supported

me enthusiastically, so I become converted, like Saint Paul, suddenly and totally; thus, we became the leading analytic duo on the Adria. We spent a lot of time discussing the literature we were reading. The biggest challenge was logic; I remember how we deciphered the newly published texts on recursion and similar topics.

Habazin unfortunately died in 1978, but we then got the offer, indeed the command from the ministry of education to employ several younger persons. This employing, done by Heda, became the crucial event in the history of the Department. With four young, promising assistants, we had the first analytic philosophy department in Croatia; our six-membered group was small, but clearly oriented in the analytical direction. We were getting support from Zagreb, from colleagues working in philosophy of science, and, above all from Belgrade, thanks to Heda's good relations with Kron. The young professor Vanda Božičević joined in with sympathies for analytic tradition. She was followed by Boran Berić, and from English language department by Dunja Jutronić, interested in philosophy of language. Heda also engaged several younger colleagues from Rijeka, Elvio Baccarini and Boran Berčić started as visiting teachers in Zadar; the event later turned out to be very important, for the future philosophy department in Rijeka.

The outbreak of the war at the beginning of nineties changed everything. We were staying under artillery fire in besieged Zadar. After the end of the war, the political countdowns began. As the result, Berić and Vanda Božičević left for the US, and Dunja Jutronić and myself ended with jobs in Slovenia. Boran Berčić and Elvio Baccarini got some teaching engagement in Rijeka. Right at the end of the millennium the philosophy department in Rijeka was created, to become an important center for analytic philosophy in Croatia, with four international conferences per year at the IUC in Dubrovnik, and a whole lot of local symposia.

It is quite obvious, in retrospect, that Heda, with her activity in Zadar in the seventies and eighties, has done a lot for the creation of this new analytic team. Berčić, Baccarini, Jutronić and I started in analytic philosophy in her institutional framework.

Heda thus stands at the beginning of the only institutionalized home analytic tradition; if the tradition goes on successfully, her name will be written on it in golden letters.

3. *Petrić—Heda's long time favorite philosopher. Pursuing the health of the republic*

Let me now pass to Franjo Petrić, philosopher extensively discussed by Heda; for an overview of her interest see the paper by Boršić and Skuhala Karasman in this issue. Here we shall discuss his short booklet, *La città felice* from 1553. Heda has been reading Petrić from a contemporary, even clearly analytic perspective; we shall later discuss briefly

her paper on Petrić and utilitarianism (2004). I shall here follow her inspiration and attempt to do the same for his utopianism.

Petrić's avowed goal of the treatise is to persuade his prominent readers, to whom it is dedicated, to implement it in reality. The work is dedicated to the two Della Rovere nobleman, Vigerio and Girolamo, and here is the leading metaphor of what the work is supposed to do:

> It will make the path easier and more passable for you, namely the path that leads to the top of the mountain on which the happiness has built the paradise of its enjoyments (*ha posto il paradiso delle sue delizie*. (Petrić 1553: iii)[1]

We shall take this guidance motivation as central for placing Petrić's work in relation to the rest of ideal theorizing. And here is the other guiding metaphor: the city offers the relief of our thirst, physical and spiritual, needed for our health and happiness.

The most adored city of the world

> If our city will be such as we have described, it will be able most abundantly to relieve the thirst and to be sated with the waters that will fall upon it from that blessed stream. This city in its greatest height, elevated among all the other cities of the world and placed in the sight of all, will be venerated by them, and adored, and implored to deign to dip its finger in the saving waters of its happy stream and to bathe their mouth, burned and thirsty, with a drop as a comfort to their miseries. (Petrić 1553: 16)

What about the adored city itself? One interesting, and probably central feature of the picture proposed is the importance of *health* as the ideal of a good city. He talks of the bodily health of the inhabitants and the need for healthy environment, going into details, relying probably on his two years of study of medicine in Padua. Here is a typical passage:

> So we shall chose places where there are no swamps or other stagnant and muddy waters, and places without those forests we have described, and places high and open, and exposed to the east wind and the north wind. But because health is corrupted not only due to the above described reasons, but by the style of our ongoing life and by the disorders which all bring upon themselves and which arise from innumerable accidents that come upon us, which are born neither from the cold nor from the heat nor from corrupt air, we need another sort of artisans who oppose these evils, with whose help we shall be liberated from the violence of them. Such are the physical medical experts, the surgeons and their assistants, the barbers, the assistants in the baths, and the specialists. (Petrić 1553: 6)

But then, in the sequel, health becomes the paradigm of the well-functioning of the city-state as a whole: "the health of the republic" (*la salute della republica*, Petrić 1553: section 7). He seems quite obsessed with the health ideal; in contemporary fashionable terms, closer to

[1] I shall refer to Istrianet website, with Italian original, and English translation by E. Ryan: http://www.istrianet.org/istria/illustri/patrizi/works/citta-felice.htm.

Greta Thunberg than to Slavoj Žižek. Of course, this plays a role in the imagined geography of the City:

> Now I come to the second defect, when, after the spirits are generated, they are dispersed. And this usually happens in two ways: either being pure and natural beyond the body, or being broken within the body. They are broken within the body due to too much condensation or too much rarefaction, or due to a poisonous quality contrary to their substance; or they become corrupt due to some other accident. Too great density is usually caused by the cold, internal or external. The rarifying likewise comes from heat that is either internal or external. And the poisonous quality is in the same way either internal or external. (Petrić 1553: 5)

We now pass to the political. The most important motivation is love among the citizens:

> Thus, there will not be private enmities in our city if love reigns among the citizens; and love is not generated except toward something that is known. So, the citizens must have information about one another. This is had in a medium-sized and manageable group rather than in an innumerable one; and even here it becomes more still easy if the group is not simply thrown together but distinguished by lineage. (Petrić 1553: 7)

How should this ideal of love be implemented? Here is the project:

> Our city, then, should not be filled with an infinite multitude of people, but with such a number that they will be able to know each other easily; and to bring this about better, they shall be divided on the basis of blood and lineage. And in order that this root of reciprocal love grows and comes to such perfection that it produces perfect fruit, I will that the people be fed in public dinners which will be celebrated at least once every month in accordance with the ancient custom of Italus, King of Italy, who put this practice into use before anyone else. Thus, let there be situated public rooms in public places where these dinners may be celebrated, and let one part of the territory of the city be public, the fruits of which may be destined only for this purpose. (Petrić 1553: 6)

Now, the equality among the citizens is an important factor of stability. The division between rulers and the ruled is determined by age: the old should rule, the young should obey and act in accordance. This then gives to the members of young generation the reasonable hope that they will be rulers in the future: "All civil discords and dissensions will cease, then, if the fire of youthful ambition will be extinguished by water of the certain hope of ruling." (Petrić 1553: 8).

However, we soon discover that things are not as ideal as they seem. Not all citizens are equal, and here is the division. There are "six types of men" in the polis. The first three are the following: (i) rural workers, (ii) artisans, for instance those "who produce for us carriages and carette and manage the horses and the mules," and (iii) "the merchants who by their industriousness lighten the road for us."[2]

[2] I shall neither introduce nor comment Petrić's use of metaphor in characterizing the task of the workers; we don't need it here.

The last three are the ones we expect from Platonic-Aristotelian tradition: (iv) the warriors, (v) the magistrates and (vi) the priests. Here comes the dramatic inequality. Categories (i)–(iii) involve so much effort and so many impediments that they block the perspective of happiness: "due to these impediments they (the members of the three categories-NM) cannot acquire the activities and the habits of the virtues that constitute the last step in arriving at beatitude." And here is the final picture of inequality:

> The remaining three orders, that is the warriors, the governors and the priests, can live for a long time, since necessities are provided for them by the three other orders that have already been described, so that with a quiet mind and without the anxiety of procuring food for themselves, they can devote all their souls to virtue both civil and contemplative. Therefore, since we want to institute a city that is blessed, because the three laboring orders cannot be clothed in the wedding garment nor be seated at table together with those wearing these garments, they will not be recognized among the invited. But they will serve at this banquet, some as cooks, others as food bearers, and the third as servers of the knife and the cup. (Petrić 1553: 1)

The city thus "has two parts, the one servile and miserable (in the original: *l'una servile e misera*), the other seigniorial and blessed (*l'altra signora e beata*)" (Petrić 1553: 1). Only the second one is made of citizens.

4. *Placing Petrić in the theoretical context: Ideal theory and thought experimenting—a general overview*

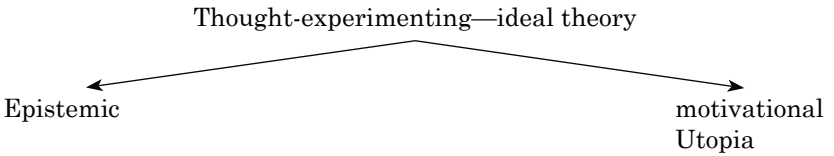
How should we classify Petrić' proposal, his ideal of *La città felice*? Fortunately, in recent times there has been an abundance of theorizing on political ideals, all under the name of "ideal theory" that has been introduced by Rawls half a century ago (for the source passage see footnote 3). However, we now have a wealth of proposals of classifications of "ideal theories"; here, we shall propose a classification that is to large extent our own, and then try to locate Petrić within it.

Let us start from initial, Rawlsian cases; they implement "ideal theory" in the sense of given, non-temporal arrangement; "ideal" either in a very limited sense of strict compliance (Rawls), or in a wider sense of normatively marked properties, no instantiated in actual political reality. Here is the description offered by Laura Valentini in her excellent overview:

> This methodological debate on the proper nature of political philosophy, and its ability to guide action in real-world circumstances, has become known as the debate on ideal and non-ideal theory. A quick glance at what falls under the heading 'ideal/non-ideal theory', however, reveals the heterogeneity of this debate. (Valentini 2012: 654)

Here I would like to introduce two proposals. The first is to situate the construction of ideal theory within the framework of thought-experi-

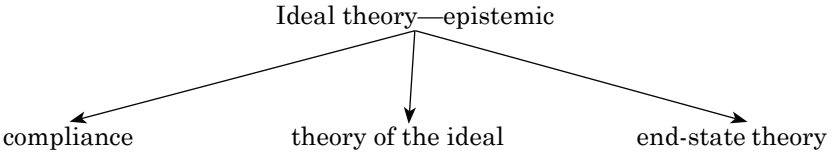
menting. Although the connection is obvious, it is largely unmentioned in the literature; one valuable exception is Rippon and Zala (2018: 55).[3] The second is to introduce a distinction, not noticed by Valentini and other authors. Many ideal theories, from Plato to Rawls, have primarily epistemic purpose: find out what is justice, most prominently. Application is secondary, and its discussion is not mandatory. Others take the opposite path: they propose the motivational function as primary, and the epistemic function as subservient to the motivational one. I shall place Petrić's work in this category. Let me reserve the term Utopia (with capital "U" for this kind):



Utopias stress the dimension of "guidance", so that they are motivational, time-related ideal theories.

The present paper depicts these relations between thought-experimenting as a wider genus, and ideal theorizing as its prominent political-philosophical sub-species.

Let me start by sketching the epistemic side: here is the main division of "ideal theorizing" taken as theorizing with the central epistemic goal: finding out the nature of the just arrangement of society. I am borrowing the main idea from Valentini:



We shall look at the three sub-kinds in the three sub-section that follow. I shall later add two more sub-kinds, the fourth one purely motivational, and a fifth one quite distinct from the rest of ideal-theoretical constructions. But now, let us look at the three kind focused on the epistemic goal.

4.1 *Compliance*

Rawls has introduced the term "ideal theory" in his *Theory of justice* in a quite modest way, as the theory of the just arrangements that relies on the assumption of full compliance of the participants in the arrange-

[3] It is also mentioned by Oana Crusmac in her (2018: 66). Amartya Sen does mention thought-experimenting (2009: 268) when talking about "transcendental" theorizing, only in the critical context and not using the term "ideal theory."

ment. He simply says at the beginning of his work that he shall "for the most part" "examine the principles of justice that would regulate a well-ordered society" (1999: 8), and then calls the resulting theory "ideal".[4] Of course, the biggest part of the discussion of this sense of "ideal theory" was dedicated to the relation with the non-ideal situation: what are we supposed to do if we know that citizens will not comply? In fact, Rawls' original suggestion is neutral in regard to the status of other characteristics of the just arrangement being discussed: we can imagine that it is a very demanding arrangement, or just a variant of existing ones. What makes it "ideal" in the first and weak sense is simple the assumption that participants comply with the rules of the arrangement.[5]

The full compliance meaning of "ideal theory" is too modest for our purpose of locating Petrić's political philosophizing, and we shall not discuss it further. Instead, we have to make the next step, as most discussants of the notion of ideal theory have.

4.2 *The theory of the ideal*

The second meaning we shall identify here is 'theory of the ideal' as opposed to 'realistic' theory.[6] Commentators and historians point to differences between projects of the theory of the ideal (and ideal theory in general). The most famous pair are Plato's "*Republic*" and his *Laws*. The project of the first takes to some extent into account the psychological and institutional possibilities (not sufficiently, Aristotle will criticize in his *Politics*, ch. 2). *Laws* are much less demanding than the *Republic*, relying on traditions and experience of various Greek polises, from very conservative ones, like Crete, to the less conservative ones. We can add a third possibility, the most radical one: the claim that factual possibility and impossibility are irrelevant for the status of the ideal. Cohen comes close to embracing this third, strongest option. We thus have three kinds of the theory of the ideal.

[4] Here is the relevant statement by Rawls:

> Thus I consider primarily what I call strict compliance as opposed to partial compliance theory (§§25, 39). The latter studies the principles that govern how we are to deal with injustice. It comprises such topics as the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and conscientious objection to militant resistance and revolution. Also included here are questions of compensatory justice and of one form of institutional injustice against another. Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. (Rawls 1999: 8)

[5] The now standard source is Simmons (2010), but see also Stemplowska (2008).

[6] Valentini notes that "/O/n this second reading of the 'ideal/non-ideal' distinction, the debate on ideal and non-ideal theory focuses on the question of whether feasibility considerations should constrain normative political theorizing and, if so, what sorts of feasibility constraints should matter" (2012: 654).

- Weak (second-best, relatively undemanding) exemplified by Plato's *Laws*.
- Moderate, exemplified by his "*Republic*" and
- Strong: exemplified by Cohen at his most radical incarnation.

The reader primarily interested in contemporary political philosophy might notice the following: Rawls, in his *Theory of justice* presents his view as a variant of the full compliance theory, nothing more. But his development suggests a different picture, reminiscent of Plato's progress. After the publication of the *Theory of justice* he came to the view that it is, as it stands, too non-realistic. And in his later work he turned to building up a more moderate theory, which was then achieved in his *Political liberalism*.[7]

The least pessimistic way is the one suggested by G. A. Cohen (2008).[8] Just a few words about this third, strongest kind. As Valentini notes, G. A. Cohen has been stressing the theoretical independence of this second meaning that she also describes as "utopian": for him it points to the value of the arrangement considered, value that can be in competition with other factors when people decide how to act politically.

There is an additional subtlety waiting in the offing. Often the proponent of an ideal arrangement proposes her scenarios as moderate, and the interlocutor sees it as strong, and almost impossible. The dialogue of Plato and Aristotle is an early example of this contrast. The history of Marxism is full of more recent examples: *The Communist Manifesto* proposes communism as a relatively normal goal; the proposal has triggered criticism that have lasted till our days.

It is hard to discuss this second, and very important meaning without briefly introducing the next one, namely the understanding of "ideal theory" as the "end-state" theory, a kind of blueprint of ideal future. The two meanings are quite connected in the practice of theorizing and writing. Our thought-experimenter imagining the idealistically valid arrangement of a community can hardly avoid seeing it also as state that would be a desirable future state of the community.

I propose that we see this second meaning as a philosophically relevant abstraction from the way actual thought-experimenting (TE) proceeds; considering it, we should stay with the imagined arrangement, and abstract from the temporal dimension that shows its relevance in the third meaning of the "ideal". The debate and our tentative systematization points to the richness of thought-experimental methodology.

4.3 *End-state theory*

The mentioning of a blueprint for a future arrangement brings us to the already mentioned third meaning of the contrast: "ideal theory"

[7] See, for instance, Weithman (2010).

[8] Discussed by Valentini in her (2012).

is "end-state theory". The non-ideal theory might concern stages of transition from the present-day arrangement(s) to the end-state one(s). This is how philosophical constructions of the ideal political world are usually read and understood in teaching philosophy and political theory: the thought-experimenting author, say Kant or J.J. Rousseau, probably had hopes that his ideal arrangement, or something recognizable close to it, will become implemented at some point in future times.

This brings in one new element: the relevance of time. If you look for a theory that is implementable here and now you will relativize, if you are into reconstructing a "great social ideal" you will stay with idealistic theory[9] (Valentini 2012: 260). To illustrate, we can imagine a theoretician, say an anarchist critic of the historical development of last five centuries or so, call her Kropotkina, who is pessimistic about the possible implementation of her anarchist ideal. The history has taken the wrong turn, she explains; six centuries ago it would still have been possible to implement it, but now, with the development of production and new, fake needs of the majority of population, this has become impossible. This is why, for Kropotkin, the ideal theory is not a temporally relevant one, and its recommendations are not recommendations for a future state.

So, the first meaning of the contrast refers directly to the conflict between the temporal structure and the modal structure in the TE: what is modally possible is not accessible in time, not feasible any more.

The discussion in the last decade has made it clear that the relevant contrast points to several different dimensions of political TE-ing. The clear given is the relative independence of the modal, dimension from the temporal one; the evaluative dimension is the third one, interacting with both in actual proposals (we can think of two non-factual aspects of the situation imagined, the axiological and the deontological one).[10]

We now pass to the second big sub-category, the motivational Utopia.

4.4 *The motivational goal: guidance*

Our fourth sub-kind is marked by motivational elements: the goal of the philosophical work is primarily to serve as guidance to political practice. Interestingly, this important sub-species of political thought-experimenting has not been understood in the literature in these terms.

[9] Valentini usefully notes the following: "If this is how we understand the ideal / non-ideal distinction, then the debate on ideal and non-ideal theory focuses on the question of whether a normative political theory should aim at identifying an ideal of societal perfection, or whether it should focus on transitional improvements without necessarily determining what the 'optimum' is" (2012: 654).

[10] Valentini mentions a third, relativisation: it relativizes all the beforementioned contrasts to the aims of the thought-experimenter and her audience. We shall leave this pragmatic aspect aside here.

Of course, if we read early modern classics as primarily motivation-focused thinkers, we can use their work as examples, with the same classification as above.

- Weak (relatively undemanding) (Bacon 1989).
- Moderate: exemplified by other early modern classical works, of authors like Campanella, More, and Petrić.
- Strong: exemplified by Cohen if we take him as defending and recommending a vison of a future society.

I shall use "Utopia" with capital "U" for this kind of motivational ideal theory (to distinguish this meaning from others, e.g. utopia as a mere unreachable dream and the like).

Most importantly for us now, Petrić clearly belongs here; remember that he recommends his work to the powerful Della Rovere politicians by telling them that "it will make the path easier and more passable" for them, namely the path that leads to a polis of perfect happiness. He describes the function of his work as guidance which is similar to the use of "model" in Bacon's *New Atlantis*. So, we have hopefully located Petrić's project within a taxonomy of ideal theorizing, and thereby taxonomy of constructional political thought experimenting.

Let me conclude the section by noting that motivational ideal theory can be, like the epistemic ones, a-temporal or temporal, relativized to time. The classical modern Utopias are motivational and not relativized to time: Petrić, More and Campanella don't tell us how their Cities would fit into the actual history of mankind. In contrast, socialist utopias, from Owen and Fourrier to Marx, Engels (see References) and Marxist utopians see their ideal societies as marking the end of history as we know it (ironically, they sometimes talk in this sense of "end of prehistory").

4.5 *Anti-ideal theory and negative utopia*

The final sub-species has not been discussed in the literature in the context of ideal vs. non-ideal theory, and it is much more present in the fiction than in philosophy. It is the negative utopia, or dystopia, like Zamyatin's community in his work *We* or Orwell's two imagined countries, one from *1984* and the other from *Animal Farm*. We can imagine a more philosophical anti-utopian theorizing, taking such dystopian construction as its starting point: for example, Chomsky has been arguing that our present-day "freedom of the press" is in fact completely "Orwellized".[11]

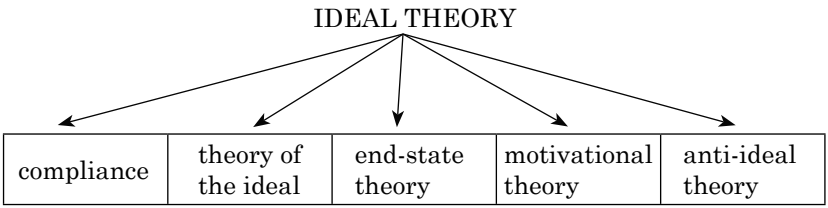
If developed, such an argument would be a symmetrical negative image of an ideal theory, and this is why I am calling it anti-ideal theorizing. Another small move in this direction is the chapter titled "1984

[11] See the summary at https://orwellsocietyblog.wordpress.com/2015/10/06/chomsky-orwell-and-the-myth-of-press-freedom/.

Is Upon Us" in Joseph E. Stiglitz's (2012) book.[12] This kind of development might be expected, given the attractiveness of dystopias for political philosophy. Again, we might apply the distinctions used for ideal theory here: distinguish a purely dystopian, "anti-idealizing" construction from a time-relativized one, presentation of dystopia as the fearsome and threatening end-state of the world. [13]

And again, we might distinguish purely epistemic function, making the possible bad things known to the reader (the way LeGuin does it), from a more usual, motivational function (present in Zamyatin and Orwell), namely warning the reader from possible threatening scenarios.

So, here is the main division repeated:



Let me conclude by pointing to the wider context of idea theory building, namely to political thought-experimenting. It can be understood to a large extent as work in ideal theorizing; a matter little noticed in the literature.

But then, what kinds of thought-experimenting yield ideal or non-ideal theories? Not imagining of some particular event (like in the Trolley problem TE), but rather a construction of a larger social and political arrangement, like a "happy city-state". Such constructive, or constructionist TEs yield ideal/non-ideal theories.

So, to reiterate, we have located Petrić's utopia into the framework of motivational ideal theory. Interestingly, this motivational component has been noticed and stressed by Vladimir Filipović, one of the best historians of philosophy in 20th century Croatia, in his Introduction to the Croatian translation of Petrić's book. He places it together with practically oriented "utopias of the late Enlightment", "works that give concrete direction about how to change norms and practices of life, so that the relations would become better and more just, and the life better for everybody" (Filipović 1975: 14).[14]

[12] See for instance the doctoral dissertation by Matthew Benjamin Cole (2017) *Dystopia and Political Imagination in the Twentieth Century*, available at ProQuest, for a reading of Habermas and Foucault along the lines of anti-ideal approach.

[13] See for more epistemic sounding approach Ursula LeGuin's *The Dispossessed*, presenting three distinct, quite negative scenarios without suggesting how close they are to actual reality.

[14] He contrasts it to "Romanesque utopias" of other Renaissance thinkers, most prominently More and Campanella. I would not go that far: I think their utopias are equally practically oriented, guidance giving works, only that Petrić is more clear in his intention that, say Campanella (whose silence on guidance might be the result

5. *Conclusion. Back to Heda Festini*

Let us close by very briefly returning to Heda's work on Petrić. What was the implicit normative framework of his construction of his motivational Utopia? We might borrow a characterization of his normative thinking from Festini (2004); she claims it is utilitarian. In her paper she starts from Petrić's appropriation of the Aristotelean "philautia", which she interprets as pursuit of what is useful to one. For Petrić, all relations to others are marked by *philautia*; friendship is grounded in the usefulness of the friend to us, and even the love for god derives from respect of ourselves, and gratitude for goods he gave to us (2004: 62). She notes that in *The happy city* the desire for well-being (*del bene essere*) plays an important role; and this well-being finds it culmination in living together. She also mentions health as the main metaphor for well-being (2004: 63). Her final diagnosis is that Petrić's utilitarianism is closer to Mill's than to Bentham's, but she also points to a possible analogies with the utilitarianism of Peter Singer (2004: 64).

So, let me summarize the main claims of the present paper. Utopias, like the one due to Petrić, also result from thought-experimenting; in contrast to purely epistemic thought-experiments they are geared to "guidance", as Petrić puts it, namely advice giving and persuading. On the opposite end of ideal theories are the ones geared to theoretical understanding; this contrast is our main contribution to classification of ideal theories.

The epistemic deal theories, the ones geared to theoretical understanding, can be either minimal, assuming only compliance with a conception of justice, or wider. The wider variant includes proposals which are not relativized to time; we called them, following proposals in the literature "theory of the ideal". The other groups are those that see the ideal situation as an "end-state" ideal. The final group are dystopias, anti-ideal theories, strong or weak.

Petrić's work can and should be understood as motivational ideal theory, a Utopia whose primary goal is guidance.

I hope that this interpretation fits well with the main line of Heda Festini's interest in Petrić, and other Croatian philosophers, trying to bring their work in connection with present day analytic efforts; it is a tribute to her work and its lasting value.

References

Bacon, F. 1989. *New Atlantis and The Great Instauration*. Wheeling: Crofts Classics.

Boršić, L. and Skuhala Karasman, I. 2019. "Heda Festini's Contribution in the Research of Croatian Philosophical Heritage." *Croatian Journal of Philosophy* 19: 575–584.

of the historical threatening context of his problems with Inquisition and various courts and his very long stay in prison).

Campanella, T. 1602. *The City of the Sun*. Radford: Wilder Publications.

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge: Harvard University Press.

Festini, H. 2004. "Tragom utilitarizma u F. Petrića." *Prilozi za istraživanje hrvatske filozofske baštine* 30 (1–2).

Crusmac, O. 2018. "Feminist Political Theory and Non-Ideal Theory as its Methodology." In B. van den Brink et al. (eds.). *Report on the workshop "Ideal and Non-Ideal theories of Justice": Towards a Non-Ideal Theory of Justice in Europe*. ETHOS: 66–79.

Filipović, V. 1975. "Uvodna napomena." In Petrić, F. *Sretan grad*. Zagreb: Fakultet političkih nauka i Libar.

Fourier, Ch. 1971. *Design for Utopia: Selected Writings. Studies in the Libertarian and Utopian Tradition*. New York: Schocken.

Hamlin, A. and Stemplowska, Z. 2012. "Theory, Ideal Theory and the Theory of Ideals." *Political Studies Review* 10: 48–62.

LeGuin U. 1974. *The Dispossessed: An Ambiguous Utopia*. New York: Harper Collins.

Miščević, N. 2012. "Plato's Republic as a Political Thought Experiment." *Croatian Journal of Philosophy* 12: 153–165.

Marx, K. and Engels, F. 1988. *The Communist Manifesto*. Oxford: Oxford University Pres.

More, T. 1975. *Utopia*. Cambridge: Cambridge University Press.

Owen, R. 1991. *A New View of Society and Other Writings*. London and New York: Penguin Books.

Petrić, F. 1553. *La città felice*. Italian original and English translation by Eugene E. Ryan. Available at http://www.istrianet.org/istria/illustri/patrizi/works/citta-felice.htm

Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

_____, 2005. *Political Liberalism*. New York: Columbia University Press.

Rippon, S. and Zala, M. 2018. "The Importance of Ideal and Non-Ideal Theory." In B. van den Brink et al. (eds.). *Report on the workshop "Ideal and Non-Ideal theories of Justice": Towards a Non-Ideal Theory of Justice in Europe*. ETHOS: 48–65.

Simmons, J. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38, no. 1, 5–36.

Sen, A. 2009. *The Idea of Justice*. Cambridge: The Belknap Press, Harvard University Press.

Stemplowska, Z. 2008. "What's Ideal About Ideal Theory?" *Social Theory and Practice* 34: 319–340.

Stiglitz, J. E. 2012. *The Price of Inequality: How Today's Divided Society Endangers Our Future*. New York: Norton & Co.

Valentini, L. 2012. "Ideal vs. non-ideal theory: a conceptual map." *Philosophy Compass*, 7 (9): 654–664.

Weithman, P. 2010. *Why Political Liberalism: On John Rawls's Political Turn*. Oxford: Oxford University Press.

Zamyatin, E. 1924. *We*. New York: E. P. Dutton.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Identity between Semantics and Metaphysics

DUŠAN DOŽUDIĆ
Institute of Philosophy, Zagreb, Croatia

In this paper, I consider several issues related to the concept of identity—the concept that is in many ways related to Heda Festini's early philosophical interests. I specifically focus on discussion of the issues in Frege, Russell, and Wittgenstein. I contrast two competing conceptions of identity—the objectual (according to which identity is a relation in which every object stands only to itself) and the metalinguistic (according to which identity is a relation between coreferential names)—and consider reasons these authors had for accepting or discarding one or the other. In addition, I consider how issues concerning identity relate to issues concerning identity statements.

Keywords: Frege, identity, identity statements, indiscernibility of identicals, informativeness, metalinguistic conception, objectual conception, relation, Russell, Wittgenstein.

1.

In 1992 Heda Festini published a book *Uvod u čitanje Ludwiga Wittgensteina* [*An Introduction to Reading Ludwig Wittgenstein*], covering in an introductory yet novel way major themes in Wittgenstein's philosophical development, from early *Notebooks* to his late *On Certainty*. In the course of that, she payed a particular attention to issues that occupied her own thinking over the two previous decades, from mid 1970ties onward. The first of the issues concerns the connection between Wittgenstein's analysis of language-games and linguistic meaning as based on use, and Dummett's and Hintikka's semantic conceptions, as well as their antirealist inclinations (see e.g. Festini 1985, 1986/1987, 1988/1989). The second one concerns Wittgenstein's earlier semantic insights, and his implicit or explicit exploitation and exploration of Fregean sense/reference (or, more generally, intension/exten-

sion) distinction, particularly in *Tractatus Logico-Philosophicus* and writings of the middle period (see e.g. Festini 1976/1977, 1978, 1982). Festini's book was never intended to be a complete, overall exposition of Wittgenstein's ideas, of course, and a number of his ideas—even those closely related to her primary interests—she never discussed. Among them is Wittgenstein's (2001) criticism of the traditional, "objectual" conception of identity (according to which it is a relation in which every object stands to itself and no other object), and his elimination of the identity sign from conceptual notation (i.e. logic) altogether.

Naturally, the issues concerning identity were not only in Wittgenstein's focus. They were of considerable interest to his predecessors—Frege and Russell—whom early Wittgenstein identified as central figures affecting his thought (Wittgenstein 2001: 4). Indeed, Wittgenstein explicitly identified Russell's (and Whitehead's 1927: 22, 57, 168) definition of identity as the primary target of his criticism (Wittgenstein 2001: 5.5302). But most of his critical remarks concern other related conceptions as well. Russell (2001: xviii) initially thought it is "a destructive criticism from which there seems no escape", but subsequently changed his mind, seeing it instead as "invalid" and "mistaken" (Russell 1959: 115).

The definition of identity Russell and Whitehead proposed in *Principia Mathematica* clearly relates to Frege's views on identity.[1] Indeed, they all fall within the "Leibnizian" tradition that in one way or another exploits the indiscernibility of identicals principle, which Leibniz formulated as: "Things are the same as each other, of which one can be substituted for the other without loss of truth" (Frege 1980a: 76; 1984: 200).[2] In addition, Wittgenstein's more positively oriented remarks about identity in *Tractatus*—what identity would amount to if it turned out not to be eliminable—bare similarities to Frege's treatment of identity in *Conceptual Notation*. Wittgenstein, for example, writes (2001: 4.241): "When I use two signs with one and the same meaning, I express this by putting the sign '=' between them. / So '$a = b$' means that the sign 'b' can be substituted for the sign 'a'." Both of them, at the time, would say that identity is a matter of linguistic conventions, rather than a sterile objectual relation. It would be a relation between names of objects provided they are coreferential, rather than objects themselves; call this the "metalinguistic" conception.

[1] The peculiarity of Russell's and Whitehead's definition—$x = y =_{\text{def.}}$ "F ($F!x \rightarrow F!y$)—steams from their hierarchisation of functions (generally, drawn to avoid various antinomies), the definition appealing only to the predicative functions. So, they insist: "We cannot state that *every* function satisfied by x is to be satisfied by y, because x satisfies functions of various orders. And these cannot all be covered by one apparent variable" (1927: 168; see also p. 57). Wittgenstein (2001: 5.5302) pointed to an addition problem with *that* feature of Russell's and Whitehead's definition, but in what follows, I will not consider it further.

[2] In fact, the tradition would be more accurately labelled "Aristotelian"; see Kneale and Kneale (1962: 42).

The concept of identity was of considerable interest to Frege, and Frege's insights about it made a considerable impact on Russell (and Wittgenstein), as well as on Dummett and Hintikka. In turn, Wittgenstein and the latter two influenced much of Festini's thinking over the two decades. And if one adds to all that that the intensional/extensional distinction is typically defined in terms of the unrestricted substitutivity that stems from identity, the concept of identity seems to be an appropriate theme for a paper included in a collection dedicated to Festini.

2.

In *The Principles of Mathematics*, Russell summarised much of the basic worries surrounding the concept of identity of the period from Frege's *Conceptual Notation* and "On Sense and Reference" to Wittgenstein's *Tractatus*. He writes:

> The question whether identity is or is not a relation, and even whether there is such a concept at all, is not easy to answer. For, it may be said, identity cannot be a relation, since, where it is truly asserted, we have only one term, whereas two terms are required for a relation. And indeed identity, an objector may urge, cannot be anything at all: two terms plainly are not identical, and one term cannot be, for what is it identical with? Nevertheless identity must be something. (Russell 1992: 63)

Here, as in many other related passages of that period, the worry starts as a metaphysical one. Russell asks, does identity *exist*, and, if it does, what is its *nature*. Immediately, however, the discussion becomes a *semantic* one—the focus now being on "where it [identity] is truly asserted" rather than on identity itself. The reason is obvious. If one limits himself strictly to metaphysical issues, the claims with which one end up are either largely uninformative, trivial, and impotent, or plainly contradictory, or even nonsensical. Namely, all one can say is that identity is a relation in which every object stands to itself and no other object, and then specify properties of that relation, such as reflexivity, symmetry, and transitivity; or one can start by saying that two objects, a and b, are identical only if some conditions Ψ are met. As Wittgenstein (2001: 5.5303) puts it, "to say of two things that they are identical is nonsense, and to say of one thing that it is identical with itself is to say nothing at all". Wittgenstein (2001: 5.53, 5.533) himself thought that this is a sufficient reason to abandon the concept of identity altogether, and to eliminate the identity sign from conceptual notation. In the reformed language, according to him, the identity of an object would be expressed by the identity of its name rather than an identity statement. That means that no two objects would bare the same name, and no single object would bare two (or more) of them. Not too many philosophers followed Wittgenstein on that point (see Ramsey 1990 for an exception). For, even if one forms a language free of the identity sign, thus carrying no information about identity, to

square mathematical and ordinary language within it would come with a high price to pay. For most philosophers, then, the feeling remained that there is more to identity than a mere tautological description, that it goes beyond contradictory statements about it, and that it is a genuine phenomenon that needs to be explained, not eliminated. To cope with the feeling, one naturally turns to ways we *talk* about, or *express* identity in ordinary language, and then try to come up with a plausible explanation of the phenomenon based on the semantic analysis of relevant statements.

A clear example of such a strategy can be found in the opening passage of Frege's "On Sense and Reference." Frege (1960: 56) too starts with a metaphysical worry: "Equality gives rise to challenging questions which are not altogether easy to answer. Is it a relation? A relation between objects, or between names or signs of objects?" But, instead of offering a straightforward answer to these questions based on whatever considerations one would classify as metaphysical, Frege turns to considerations of identity *statements*. Afterwards, nowhere is his paper does he deals with the first question (although his second question strongly suggests the answer), and the second dilemma is settled explicitly only negatively: Neither option is acceptable to Frege because neither can explain the relevant phenomena concerning the identity statements. This Frege's point, I think, is based on the confusion of metaphysical and semantic (and epistemological) issues. The rest of Frege's paper deals exclusively with the latter issues, although the way he opens his paper, as well as the way he concludes it, suggests he deals with the former.

So, immediately after posing the questions about the nature of identity, Frege turns to consideration of identity statements. He distinguishes statements of the form "$a = a$" (e.g. "Cicero is Cicero") from statements of the form "$a = b$" (e.g. "Cicero is Tully"). The distinction, however, is not made on the ground that they have different *form* (see also Frege 1972: 124). Rather, Frege (1960: 56) insists, the distinction should be made because "$a = a$" and "$a = b$" differ in cognitive value: "$a = a$ holds *a priori* and, according to Kant, is to be labelled analytic, while statements of the form $a = b$ often contain very valuable extensions of our knowledge and cannot always be established *a priori*". This feature of identity statements—particularly statements of the form "$a = b$"—Frege suggests, supports the view that identity is a relation between *names* of objects. To say that Cicero is Tully, for example, is to say that names "Cicero" and "Tully" designate the same object.

Given the general English conventions about functioning of proper names and the verb "is" (interpreted as identity)—namely, what a competent English speaker tacitly knows when correctly using these expressions—from the fact that the sentence "Cicero is Tully" is true there follows that names "Cicero" and "Tully" designate the same object. And it certainly comes as a discovery to learn that a person bears

another name, and what that other name is—a discovery one cannot know a priori unless he stipulatively introduced that other name into discourse.[3] Accordingly, a statement expressing that circumstance cannot be analytic. I know, for example, that "Lady Gaga", "Lil' Kim", and "Nicki Minaj", are names of Lady Gaga, Lil' Kim, and Nicki Minaj, respectively, and I strongly suspect that these are not the only names of these singers. But only after some googling, I discover what names they bear in addition, and that Lady Gaga is Stefani Germanotta, that Lil' Kim is Kimberly Jones, and that Nicki Minaj is Onika Maraj. Given all that, the proposal that the difference in cognitive value be explained by appeal to metalinguistic information sounds appealing.

3.

Frege embraced the metalinguistic conception of identity, and for the similar reasons, in *Conceptual Notation* (Frege 1972: 124–126; 1960: 56). There, instead of the standard identity symbol "=", Frege introduced a novel symbol, "≡", that stands for the *identity of content* of symbols placed on the left and the right of it, and explained it as follows (Frege 1972: 124): "Identity of content differs from conditionality and negation by relating to names, not to contents. Although symbols are usually only representatives of their contents […] they at once appear *in propria persona* as soon as they are combined by the symbol for identity of content, for this signifies the circumstance that the two names have the same content."

Frege's "≡" is more general in application than "=". It can be combined with symbols with which "=", strictly taken, cannot, as long as these symbols have a (conceptual) content.[4] And, combined with the double judgment stroke, it serves to Frege as the indicator of abbreviative definition (Frege 1972: 126, 167–168).[5] Other than that, there is no difference, and it would be wrong to conclude that Frege intended to use "≡" in addition to "=" (for the latter symbol is used nowhere in the concept script). Nor should one think that Frege intended to elimi-

[3] For a discussion about the possibility of knowing a priori truths that are otherwise known *a posteriori*, see e.g. Kripke (1980: 63). Frege (1972: 167–168), and Russell and Whitehead (1927: 168), thought that such stipulative or abbreviative definitions are not identity statements on the par with "Cicero is Tully" or "Hesperus is Phosphorus".

[4] In his latter writings, Frege treated both singular terms and sentences as proper names of objects, so all such expressions could, from that perspective, flank the standard identity sign, and in his writings they do.

[5] Russell and Whitehead in *Principia Mathematica* distinguished three senses of Frege's "≡"—as identity, equivalence, and abbreviative definition—by representing them formally using different signs, namely, the identity sign "=" (1927: 22–23), the equivalence sign "≡" (1927: 7), and the definition sign "= Df", which is to be taken as a single symbol, rather than as composed of two symbols, the identity sign and "Df" (1927: 11). Wittgenstein (2001: 4.241, 5.101) followed Russell and Whitehead in that respect.

nate "=" in any significant sense; certainly not in Wittgenstein's (2001). The symbol "≡" is merely broader in application, and free of whatever unwanted burden "=" might bring for a reader into the concept script from mathematics and ordinary use, the burden which in *Conceptual Notation* Frege was eager to avoid. Therefore, wherever "=" would be used, "≡" could be used as well (but not vice versa). In its literal use, "Snow is white = Snijeg je bijel" would make little sense, but "Snow is white ≡ Snijeg je bijel" would be perfectly fine.[6] And that there is a need for such a symbol, with the intended metalinguistic interpretation, Frege demonstrates using a geometrical example where "*A*" and "*B*" ultimately name the same fix point on the circumference of a circle around which a straight line rotates, and concludes (Frege 1972: 126) "that different names for the same content are not always merely an indifferent matter of form; but rather, if they are associated with different modes of determination, they concern the very heart of the matter. In this case, the judgement as to identity of content is, in Kant's sense synthetic". So, the point is the same as in the previously quoted passage—statements of the form "$A \equiv B$", just as the earlier ones of the form "$a = b$" are always synthetic, and, at least sometimes known a posteriori.

Frege subsequently become dissatisfied with the proposed conception of identity. The problems he saw with it in "On Sense and Reference" are not fully clear, but it seems that his main point was that, if interpreted metalinguistically, "the sentence $a = b$ would no longer refer to the subject matter, but only to its mode of designation; we would express no proper knowledge by its means" (Frege 1960: 56). That most likely means that, e.g., the discovery *that Hesperus is Phosphorus* is not a linguistic discovery about the coreference of names "Hesperus" and "Phosphorus". Rather, it is an astronomical discovery about a planet that goes beyond linguistic conventions of English. And the above proposed conception of identity apparently fails to capture *that* fact. So, even if the English sentence "Hesperus is Phosphorus" in some sense implies that names "Hesperus" and "Phosphorus" designate the same thing, it is certainly not what that sentence primarily, or literally, says. Apparently, then, Frege became dissatisfied with his early conception of identity for the same reason he was dissatisfied with formalist treatments of arithmetic. Just as numerals are not a proper subject-matter of arithmetic, so names are not a proper subject-matter of identity statements (see e.g. Frege 2013: ix). Russell and Whitehead (1927: 67) gave a similar objection to metalinguistic reading of identity statements. Their complaint was not, however, that such reading changes

[6] An example of a nonstandard use of "=", the one that appeals to our pragmatic intuitions, and which is defined nowhere in the book, can be found in Russell and Whitehead (1927: 138), where, for example, they interpret the proposition "*p*" as "*p* = Socrates is a Greek", and the propositional function "*fx*" as "*fx* . = . *x* is a Greek". These are certainly not identity statements, and, as it seems, they are not worthy of being labelled definitions.

the subject matter of identity statements, but rather their truth conditions, because part of truth conditions of any such statement would be that a certain object be called a certain name—but the truth of such a statement cannot depend on that feature.[7]

Of course, one can think of a number of other problems with the proposed conception of identity. The crucial one is that no matter how the thesis is ultimately spelled out, it will always presuppose the competing objectual conception of identity. Just consider Frege's two variants of metalinguistic definition of identity, namely "the symbol *A* and the symbol *B* have the same conceptual content" (Frege 1972: 126), and "the signs or names '*a*' and '*b*' designate the same thing" (Frege 1960: 56). Both definientia contain the phrase "the same", which must be interpreted in terms of the objectual identity relation. And an alternative metalingusitic definiens, namely, "the object named '*a*' is identical to the object named '*b*'" (Frege 1960: 78), faces the problem even more obviously. It follows, then, that any such metalinguistic definition of identity presupposes the objectual identity, and so, whatever it merits would be, it could not be its alternative, but, at best, a supplement. But, in the light of the above objection, should one even consider keeping the metalinguistic definition? One reason would certainly be to keep it not as the definition of identity, but rather as the explication of the content or truth conditions of identity statements. That would certainly be compatible with Frege's (1972, 1960) reasons to consider it in the first place. Nevertheless, one would still face Frege's initial objection.

The problem of presupposing the objectual identity could be avoided if the concept of identity occurring in definiens would be interpreted metalinguistically as well, but only at the cost of either the circularity of the definition or leading into the infinite regress. Russell most likely had that in mind when he noted that Frege's early take on identity is "a definition which, verbally at least, suffers from circularity" (1992: 502). Later, Frege made a related point when he wrote: "Since any definition is an identification, identity itself cannot be defined" (Frege 1984: 200). So, the problem of circularity would be double here: Not only does the definiens appeal to the very concept it should define, but the very definition of identity—whatever form it may take—is itself a case of identity statement, and as such it presupposes the concept. It is far from clear, however, that Frege's definition would be circular in the second sense. Namely, by defining the concept of identity of content, Frege is in fact not describing a previously established concept and determining the meaning of its familiar symbol. Rather, he introduces a novel symbol and stipulates its meaning, thus bringing a new concept. And, given the way he understood such stipulative definitions, it is far from clear that they are the case of identity statements. Also, to what

[7] They originally made that point for cases with definite descriptions, but the point goes for other singular terms as well. For a similar objection see Kripke (1980: 108).

degree Frege's concept overlaps with the familiar concept of identity might—with respect to the problem of circularity—be irrelevant. I will return to that issue in section 5.

If identity in general cannot be properly defined, since every definition of identity has the form of identity statement, and if, in addition, identity cannot be a relation between names, since that would commit us either to accept the objectual identity, or it would lead us into circularity and infinite regress, it seems that one has no choice but to grant that identity is an indefinable relation between objects. But that option Frege found equally unsatisfied. For him, the same thing that supports the metalinguistic conception—namely, the informativeness of identity statements of the form "$a = b$"—undermines the objectual conception. If identity would merely be a relation in which every object stands to itself and no other object, "$a = a$" and "$a = b$" would say the same thing, and would thus differ only in form. But that is obviously not the case. For Frege, there is more to identity than that.

4.

The conclusion of the opening passage of "On Sense and Reference" is that neither of the two mentioned options is the acceptable one. And, as far as identity goes, the concept is further discussed nowhere in the paper. The ultimate conclusion of the passage is only that "a difference [between statements '$a = a$' and '$a = b$'] can arise only if the difference between the signs ['a' and 'b'] corresponds to a difference in the mode of presentation of that which is designated". *This* tells us nothing about identity itself. And the rest of the paper is merely an elaboration and extension of this conclusion. Indeed, Frege's closing passage in the paper seems misleading on that matter. He writes:

> Let us return to our starting point. / When we found '$a = a$' and '$a = b$' to have different cognitive values, the explanation is that for the purpose of knowledge, the sense of the sentence, viz., the thought expressed by it, is no less relevant than its reference, i.e. its truth value. If now $a = b$, then indeed the reference of 'b' is the same as that of 'a,' and hence the truth value of '$a = b$' is the same as that of '$a = a$.' In spite of this, the sense of 'b' may differ from that of 'a', and thereby the thought expressed in '$a = b$' differs from that of '$a = a$.' In that case the two sentences do not have the same cognitive value. (Frege 1960: 78)

The passage is misleading because Frege's starting point was the question about *identity*, not identity *statements*, and these are two different, although related thing. Frege does not provide any metaphysical view about identity in spite of his initial metaphysical question. Instead, he offers a semantic analysis of identity statements, based on the sense/reference distinction.

To make an analogy: It is one thing to ask, for example, do propositional attitudes exist, and, if so, are they relations, what (if anything) they relate, etc. These are metaphysical issues. It is quite another thing

to ask what propositional attitude *reports*, namely sentences reporting subject's particular attitude, typically say, what are their truth conditions, etc. The metaphysics of propositional attitudes one embraces at the outset might help in forming the semantic analysis of attitude reports with which one will ultimately end up. Just as it might turn out that the semantic analysis of attitude reports one embraces will ultimately determine the way one understands attitudes themselves. Nevertheless, these are two different issues that should not be conflated. The same goes for identity and identity statements (and virtually any other metaphysical issue that finds its counterpart in semantic discussions concerning the accompanying vocabular; just think of universals or time).

In addition, the last quoted passage contains another problematic point. Recall, in the opening passage of the paper, we are left only with the negative answer to the question whether identity is a relation between objects or between names of objects. And now, given the intonation, it seems that Frege is at least hinting which of the two options he accepts when he writes "[i]f now $a = b$, then indeed the reference of 'b' is the same as that of 'a'". But what option is that? On the closer inspection, one finds that this formulation is ambiguous, and that it is compatible with either of the two options, since, on the par with Frege's (1980a: 69) transformation of numerical statements of the form "x has N ps" into identity statements of the form "the number of x's ps is (identical to) N", one could transform Frege's two formulations of the metalinguistic definiens, namely, "the symbol A and the symbol B have the same conceptual content" (Frege 1972: 126), and "the signs or names 'a' and 'b' designate the same thing" (Frege 1960: 56), into formulations resembling the above ones, namely: "the conceptual content of the symbol A is the same as that of B" and "the thing designated by the sign of name 'a' is the same as that of 'b'". In fact, Frege in *Conceptual Notation* at one point, reflecting on his geometrical example demonstrating the informativeness of identity statements, writes that "the name B has the same content as the name A" (1972: 125).

Nevertheless, given the way Frege appeals to identity in *The Foundations of Arithmetic* and his other writings after *Conceptual Notation*, one could have little doubt about which concept of identity he embraces. It is the plain objectual concept according to which identity is the relation in which an object stands to itself and no other object.[8] For him to

[8] Frege slipped into the metalinguistic interpretation even after *Conceptual Notation*: "[…] 'the number of Jupiter's moons is the number four, or 4'. Here 'is' has the sense of 'is identical with' or 'is the same as'. So that what we have is an identity, stating that the expression 'the number of Jupiter's moons' signifies the same object as the word 'four'." (Frege 1980a: 69). Apparently, such an interpretation comes naturally. Similarly, Kripke, a clear opponent of the metalinguistic interpretation (see Kripke 1980: 107–108), at one point in his book writes that "sometimes we may discover that two names have the same referent, and *express this* by an identity statement" (Kripke 1980: 28, my italics). Pace Kripke, we would more likely express

embrace the alternative formulation would be redundant, since in his later writings the phenomenon of informativeness is explained by appealing to senses, rather than modes of designation that he introduced with his early conception of identity. So, given the way the argumentation in the initial passage of "On Sense and Reference" is set, Frege should not have objected to the objectual view of identity that it cannot account for the alleged difference in cognitive value, because the view was never intended to be such an explanation. Instead, he should have said that although *identity* is a relation "in which each thing stands to itself but to no other thing" (Frege 1960: 56), identity *statements*, at least those of the form "$a = b$", convey information that goes beyond that metaphysical dictum; hence the difference in thoughts expressed by "$a = a$" and "$a = b$". Keeping that in mind, the issue Frege is particularly concerned with is whether the information identity statements involve is: (*a*) information about names flanking the identity sign, and their semantic conventions—ways of designation; (*b*) information about the object in question that is given in different ways, independently of the way it is designated; or (*c*) merely the information about the self-identity of an object.

All three options, and not just (*c*), plainly presuppose the objectual view of identity. Indeed, one would think, it seems impossible to sidestep the objectual view since it is incorporated into the very way we think about objects and the way they are related. So, is there more to be said about identity?

5.

Frege and Russell in principle agreed on many points concerning the concept of identity. For one thing, both of them appealed to identity statements and their informativeness to point out the need for a semantic analysis that goes beyond mere reference of relevant expressions. Frege thought that it strongly supports his sense/reference distinction, Russell (1992: 63) took it as crucial for the semantics of descriptive phrases (see also Russell and Whitehead 1927: 23). Also, both of them accepted Leibniz's indiscernibility of identicals principle as the fundamental law governing identity, perhaps even its definition. Thus, in *The Foundations of Arithmetic*, Frege (1980a: 76) writes: "Now Leibniz's definition is as follows: 'Things are the same as each other, of which one can be substituted for the other without loss of truth'. This I propose to adopt as my own definition of identity". In *Conceptual Notation* (Frege 1972: 161–162), and later in *Basic Laws of Arithmetic* (Frege 2013: 36), Frege introduced variants of the principle as one of the axioms (or basic laws) of the logical system, and Russell and Whitehead (1927: 23) write: "If x and y are identical, either can

this with the statement "'*a*' and '*b*' are coreferential". "*a*" and "*b*" would not be used, but only mentioned.

replace the other in any proposition without altering the truth-value of the proposition."

Taken at face value, the principle of indiscernibility of identicals, be it a definition or merely "a principle that brings out the nature of the relation of identity" (Frege 1984: 200), or "a fundamental property of identity, from which remining properties mostly follow" (Russell and Whitehead 1927: 23), it makes no sense. Firstly, the plural "things", the phrase "each other", and similar, are in obvious conflict with the very idea of identity, for no *two* (or more) things could ever be identical with each other. Secondly, even if this awkward wording is ignored, the idea that *things* are substituted is just as bad: Where exactly would we substitute a thing, and, for any given thing, for what other thing should it be substituted, and truth of what could be lost? A way to make some sense from at least a part of that formulation would be to say that propositions, rather than sentences, are primary truth bearers, that objects are their constituents, and so that one substitutes objects within propositions. The problem with that would obviously be that whenever identity holds for whatever object—and, by definition, it always holds for every object—no *other* object could be substituted for it on the ground of identity. Thus, if anything is substituted in such cases, it is certainly not an object entering the identity relation.

If one is to make any sense of Leibniz's indiscernibility of identicals principle, at least a fundamental revision of its formulation is needed. Frege did not address this issue explicitly, but the way he immediately utilised it, suggests that he most likely was aware that in its original form it makes no sense. Thus, in the same paragraph, Frege writes:

> Now, it is actually the case that in universal substitutivity all the laws of identity are contained. / In order, therefore, to justify our proposed definition of the direction of a line, we should have to show that it is possible, if line a is parallel to line b, to substitute "the direction of b" everywhere for "the direction of a". (Frege 1980a: 77)

Similarly, Russell and Whitehead (1927: 23) start with the formulation: "[i]f x and y are identical, either can replace the other […]", but just a passage below continue: "[identity] can only hold between x and y if x and y are different symbols for the same object". They clearly talk first about identity as an objectual relation—since, taken as objects in their own right, "x" and "y" are definitively not identical—and symbols "x" and "y" are *used* to represent objects.[9] But then they switch to metalinguistic mode, *mentioning* these symbols in the passage that follows. Taken in conjunction, the quoted section makes little sense. And they obviously did not intend to accept the metalinguistic conception, since, a bit further, they explicitly criticise it (Russell and Whitehead 1927: 67).

[9] Throughout *Principia Mathematica* one finds a number of places that support that reading, e.g., in their phrase "the objects which are identical with x". Here, "x" stands for an object, and it is not a disguised name of the symbol "x".

If Leibniz's definition is to be interpreted (or rephrased) in the light of the quoted passages, it is clear that it is not a thing that is substituted, but rather its name, that it is substituted for another name it bears, and that the substitution takes place in a sentence. In that case, the truth of a sentence in which it occurs is preserved.[10] So the definition would now be: *a* is identical to *b* only if *a*'s name "*a*" can be substituted for (its other) name "*b*" in a sentence without the loss of its truth. But this could hardly be taken as a definition of *identity*; if it were, what would it tell us about it? Rather, we already have to possess the concept of identity to make sense of such a formulation. Subsequently, Frege become dissatisfied with Leibniz's definition of identity because it, or any other definition of identity, would be circular (Frege 1984: 200): "Leibniz's explanation […] does not deserve to be called a definition […] Since any definition is an identification, identity itself cannot be defined".

6.

Now, if we consider this Frege's remark in the light of his earlier distinction between plain identity statements and abbreviative definitions, he obviously thinks that identity is a concept we already possess (see also Frege 1980a: 74), and can subsequently only describe. It is not a concept we introduced by a definition. Thus, a definition of identity would itself be an identity statement on the par with "Cicero is Tully". Russell and Whitehead (1927: 11, 57) took a different course, writing as if identity—at least in the context of their formal system—is introduced stipulatively, and thus that its very definition is not an identity statement. Its definition, just as any other definition in *Principia Mathematica*, according to them, would be normative—"the expression of a volition"—rather than descriptive; it "is concerned wholly with the symbols, not with what they symbolise"; and it is neither true nor false, since it is not asserted (1927: 11). But, as far as the definition of identity goes, their view seems problematic, since it is far from clear that the definition if identity is such a definition. If that is so, one should side with Frege's ultimate conclusion, namely, accept that no definition of identity is possible. At best, one could end up with its "informative analysis" that would explicate its features. Pace Wittgenstein (2001), it should be observed that, even if one could build a formal language devoid of the identity sign, one could hardly square ordinary and mathematical language within it. And if the latter ones are the phenomena one should explain, rather than explain away, a sufficiently strong formal system for that purpose should certainly keep the identity sign with its preestablished use, and the concept of identity lurking behind it.

[10] For the sake of simplicity, I ignore here the issue of intensional sentences, i.e., the problem some such sentences pose for unrestricted substitutivity.

References

Festini, H. 1976/1977. "La rilevanza semantica del' Tractatus di Wittgenstein", *Annali* 19/20: 421–429.

_____, 1978. "One Semantical Consequence of Wittgenstein's Concept of Proposition as Logical Picture." In E. Leinfellner et al. (eds.). *Wittgenstein and His Impact on Contemporary Thought*. Vienna: Hölder-Pichler-Tempsky: 139–142.

_____, 1982. "Is Wittgenstein's Concept of Proposition in *Philosophical Grammar* Intensional, Extensional or Something Else?" in W. Leinfellner, E. Kraemer, and J. Schank (eds.). *Language and Ontology*. Vienna: Hölder-Pichler-Tempsky: 498–500.

_____, 1985. "Dummett's Conception as Theory of Meaning for Hintikka's Type of Game-Theoretical Semantics (I) ('Use' and 'Language-Game' in Wittgenstein and Dummett)." In G. Dorn and P. Weingartner (eds.). *Foundations of Logic and Language*. (New York: Plenum): 639–664.

_____, 1986/1987. "Antirealism/Realism of Hintikka's Game-Theoretical Semantics." *Radovi* 26: 13–21.

_____, 1988/1989. "Antirealism/Realism of Wittgenstein's Language-Game Idea." 27: 5–20.

_____, 1992. *Uvod u čitanje Ludwiga Wittgensteina*. Zagreb: Hrvatsko filozofsko društvo.

Frege, G. 1960. "On Sense and Reference." In P. Geach and M. Black (eds.). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell: 56–78.

_____, 1972. *Conceptual Notation: A Formula Language of Pure Thought Modelled upon the Formula Language of Arithmetic*, in *Conceptual Notation and Related Articles*. Oxford: Clarendon Press.

_____, 1980a. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Oxford: Basil Blackwell.

_____, 1980b. "Frege to Husserl 24.5.1891." In G. Gabriel et al. (eds.). *Philosophical and Mathematical Correspondence*. Oxford: Basil Blackwell: 61–64.

_____, 1984. "Review of E. G. Husserl, *Philosophie der Arithmetik* I." In B. McGuinness (ed.). *Collected Paper on Mathematics, Logic, and Philosophy*. Oxford and New York: Basil Blackwell: 195–209.

_____, 2013. *Basic Laws of Arithmetic*. Oxford: Oxford University Press.

Kneale, W. and Kneale, M. 1962. *The Development of Logic*. Oxford: Clarendon Press.

Kripke, S. A. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Ramsey, F. P. 1990. "The Foundations of Mathematics." In D. H. Mellor (ed.). *Philosophical Papers*. Cambridge and New York: Cambridge University Press: 164–224.

Russell, B. 1959. *My Philosophical Development*. London: George Allen and Unwin.

_____, 1992. *The Principles of Mathematics*. London: Routledge.

_____, 2001. "Introduction." In L. Wittgenstein. *Tractatus Logico-Philosophicus*. London and New York: Routledge: ix–xxv.

Russell, B. and Whitehead, A. N. 1927. *Principia Mathematica*. Vol.1, 2nd edn. Cambridge: Cambridge University Press.

Wittgenstein, L. 2001. *Tractatus Logico-Philosophicus*. London and New York: Routledge.

Croatian Journal of Philosophy
Vol. XIX, No. 57, 2019

Book Reviews

Nicholas Shea, Representation in Cognitive Science*, Oxford: Oxford University Press, 2018, 304 pp.*

Nicholas Shea's excellent book *Representation in Cognitive Science* is the most recent attempt to provide a naturalized theory of representational content, that is, an attempt to explain how representations, understood as content-baring physical particulars, acquire their content, using non-semantic, nonmental and non-normative descriptions (11). The account offered in his book is a continuation of the two most influential naturalistic approaches to mental content—teleosemantic approach (Millikan, Papineau) and informational approach (Dretske, Neander). His account relies on standard resources of these theories—most important of those being the notions of function, information and correspondence, but develops an original understanding of how these notions converge in the metaphysical determination of representational content. In doing so it also relies on the work by Peter Godfrey-Smith. Shea's book is abundant with case studies ranging from studies of simple artificial and animal systems to those more complex but sufficiently understood by cognitive neuroscience, such as the spatial-navigation system in the rat's hippocampus. He uses them to develop his account, but also to test it in relation to standard objections made against teleosemantic approaches, such as the problem of indeterminacy of function or the infamous swampman objection.

The book is divided in three parts. The first one is introductory and offers a framework for the account developed in the rest of the book (chapters 1, 2). Part two presents his account of representational content, centering on its three main ingredients—task function, correlational information and structural correspondence (chapters 3, 4, 5). Part three answers aforementioned indeterminacy and swampman objections, offers an account of the distinction between descriptive and directive representation and concludes with several considerations concerning the explanatory role of content and content of higher personal, conscious states (chapters 6, 7, 8). Lack of space makes it impossible to present every chapter separately. It also makes it impossible to present the numerous case studies and important subdiscussions which comprise Shea's book. While admitting its insufficiency, the review will be focused on presenting several key aspects of Shea's book.

The ingenuity of Shea's account consists in its ability to reconcile resources of different, competing theories of mental content. The main idea behind his theory of representational content is that content arises out of the convergence of three elements—functions, in his account called *task*

functions, *exploitable relations* and *internal mechanism*—each of which is necessary for content. The core idea is that representational vehicles acquire content by bearing certain relations to the environment, relations that are exploited by an internal mechanism in order to perform a certain function. What is immediately noticable from this, somewhat crude, definition is that on Shea's account, content is partially determined by the relational properties of vehicles, that is, externalist. This is justified by the fact that behaviour, as an explanandum, involves responding to distal environmental features in order to bring distal environmental effects. Secondly, Shea's account is committed to there being real vehicles of representation. His account is thus a version of a representational theory of mind. However, he intends his theory of content to apply first and most to subpersonal, unconscious representations, since these figure prominently in cognitive science and cognitive neuroscience. Considerations about first-person, conscious mental states are given in the final chapter of the book. Finally, there is no reliance on the notion of a *representational consumer* characteristic of teleosemantics, since Shea finds it problematic to apply a consumer-based analysis on complex, "multi-layered" and feedback-involving systems. Even though his account shares with teleosemantics the view that the content of representations is determined by their use, it is their use in downstream processing by an internal mechanism that determines content, not their being used by a dedicated consumer system.

One important point of departure between Shea's account and its predecessors is his pluralistic framework, which he calls *varitel semantics*. Content, according to Shea, can be determined by more than one sufficient condition. The source of this pluralism is a disjunctive account of functions and exploitable relations, which makes content determination different depending on the type of function and exploitable relation present in each case.

The proper way to explain functions (in his account task functions) possessed by an organism, according to Shea, is by giving a *consequence etiology*. This is a point which Shea shares with other teleosemantic approaches. Dispositions for behavioral outcomes produced by an organism are explained by the fact that the same behavioral outcomes produced certain distal effects which were beneficial for the organism in the past. Because of their beneficial consequences, the production of these distal effect became a *stabilized function* of the behavioral outcome produced by an organism. However, while teleosemantics admits of natural selection as the only process that stabilizes functions (learning being a derivative of natural selection), Shea allows for three more types of stabilizing processes. In living organisms, these are *learning with feedback* and *persistence of organisms* (contribution to survival of an individual organism); while in artificial systems this process is deliberate design. On Shea's account each of these four types of stabilizing processes tend to converge with another property which Shea finds important in describing function, and that is *robustness*. Stabilizing processes tend to stabilize those behavioural outcomes which are sufficiently robust, that is, which can perform their function in a range of different conditions. One way for an organism to possess this feature is for its internal mechanisms to be sensitive to a wide range of different inputs. Representational explanation offers a solution of how organisms manage

to achieve this robustness. By having vehicles that bear relations to varying environmental features, an internal mechanism is able to exploit these relations in order to successfully perform a function in different conditions.

This brings us to the second part of Shea's theory—*exploitable relations*. Remember the central idea that a system performs its task function by having an internal mechanism whose processing is able to exploit the relations its vehicles bear to distal features of environment. These exploitable relations can be of two types. The first type of relation is *correlational information* (Chapter 4). Correlational information is usually conceived as a relation of probability raising. Obtaining of a certain state A raises the probability that another state B is also obtaining. However, a certain state usually correlates with a number of different states (and properties) and correlates so in varying strengths. The appropriate way to identify the correlation that is *actually* exploited by an internal mechanism, Shea believes, is to see which correlations have an *unmediated role* in explaining how a system performs its function. In turn, to explain how a system performs a function is to explain how that function became stabilized and robustly produced. Focusing on the explanatory role of exploitable relations ties content-determination tightly with explanatory considerations (a fact which does not imply a dependence on an intentional observer). To use a famous example, in explaining how a frog is able to catch a fly, we can identify different correlations which exist between the frog's retinal ganglion cells and the properties attributable to fly and thus a number of candidates for content-determination. The cells correlate with there being a *little black thing, a fly, a nutritious object, a fly I saw two minutes ago* etc. This is one aspect of the notorious problem of indeterminacy of content. However, Shea argues, a correlation with *little black thing* explains why the frog is able to perform its function of catching a fly only in a *mediated* way, that is, by also correlating with *nutritious object* or *fly* and *nutritious object* or *fly* is what unmediatedly explains why the function of catching flies came to be stabilized and robustly produced. If *little black thing* had not correlated with *nutritious object* or *fly,* an explanation of how the function of catching flies came to be (historically) stabilized by the frog would miss out on an important explanatory pattern. Let us now turn to the second type of exploitable relation that figures in Shea's varitel semantics—*structural correspondence* (Chapter 5).

Shea defines structural correspondence in the following way: a structural correspondence exists between a relation V on vehicles v_m and a relation H on entities x_n iff there is a function f which maps the v_m onto the x_n and $\forall_{i,j}\ V(v_{i,}v_j) \leftrightarrow H(f(v_i),f(v_j))$ (117). It is a well-known point that structural correspondence is too liberal for fixing content. This is because a certain relation V on v_m can correspond to many different relations $H_1....H_n$ on worldly entities x_n, given that v_m and x_n are of the same cardinality. Shea proposes to constrain this liberality by requiring that the correspondence is actually exploited by a system in order to perform a function. This involves a double restriction on the candidates for content-constituting structural correspondences. First, the mechanism performing a certain function has to be sensitive to the relation which exists on the vehicles (the relation has to be used in downstream processing) and the correspondence has to play an unmediated role in explaining how a system performs a function (it has to

be of *significance* to the system). This way the notion of structural correspondence becomes sufficiently constrained and also tractable by empirical investigations. In chapter 5 Shea gives a number of wonderful demonstrations of how his account can be applied to cases of representing spatial relations, similarity and causal structure.

Next, in chapter 7 Shea offers an account of the distinction between directive an descriptive content based on his varitel framework and gives a comparison with existing accounts (Millikan, Price, Artiga, Sterelny etc.) At first approximation, directive content is concerned with producing a certain condition, while descriptive content is concerned with reflecting a certain condition. Shea draws the distinction by relying on the resources of his varitel semantics. A vehicle R standing in an exploitable relation (correlational information or structural correspondence) with a condition C has *directive content* if the production of C by a vehicle R plays an unmediated role in explaining how a system performs a task function. On the other hand, a vehicle R standing in an exploitable relation with a condition C has *descriptive content* if C's obtaining when R is tokened plays an unmediated role in explaining the system's performance of a function, but not via R's producing C (pp. 180–181). Shea then demonstrates how his way of drawing the distinction applies to different cases presented in his book.

A final aspect we will examine is the application of Shea's account to the problems of indeterminacy of content (the problems of distality, specificity and disjunction) which is presented in Chapter 6. In presenting his solution, Shea makes several comparisons between his and other solutions present in the literature. He aligns his theory with the so-called "high church" teleosemantics which ties content to explanations of successful behaviour prompted by a representation, as opposed to "low church" teleosemantics which ties content with discriminative abilities of the organism. This is so because on Shea's account the determinacy of content is grounded in the determinacy of task functions and the exploitable relations which play a role in explaining these functions. Since explanations of stabilization and robustness of task functions are causal explanations they make a restriction on the type of properties that are adequate in explaining task functions, on both the explanandum and the explanans side. This makes the proper identification of the task function involve properties which actually led to stabilization and robustness. Returning to the frog example, this means that the correct description of its task function is *catching flies* and not, for example, *catching little black things*. On the side of the explanandum, similarly, the adequate properties will again be those that figure in a causal explanation of stabilization and robustness. On Shea's account those will include, of course, the exploitable relations of correlational information and structural correspondence. Given Shea's requirement that these relations have to be those relations that unmediately explain stabilization and robustness, they will turn out to be properties such as *fly* or *nutritious object* rather than *little black thing* or *my favourite fly*. It is noticeable that this still leaves a certain degree of indeterminacy of content. However, Shea argues that this consequence is due to the nature of simple systems such as the frog. Even in the simple cases, Shea argues, determinacy of task functions and exploitable relations provides a considerable degree of determinacy on content.

It is safe to conclude that Shea's *Representation in Cognitive Science* will become an essential reading in the literature concerning mental representation. Apart from offering an ingenious account of representational content, the book provides a clear identification of the standard problems which surround theorizing about representational content. It also makes numerous comparisons to existing accounts and provides discussions about other themes, such as the notion of biological function or explanation. In spite of its complexity and extensive use of results from cognitive science and neuroscience, which demand a more specialized reader, these facts make the book sufficiently accessible to graduate level students and other, less informed, scientific and philosophical audiences that are interested in exploring the nature of mental representation.

MARKO DELIĆ
University of Split, Split, Croatia

Rui Costa and Paola Pittia (eds.), Food Ethics Education, *New York: Springer Publishing, 2018, 239 pp.*

Food Ethics Education, edited by Rui Costa and Paola Pittia, is the selection of texts in which authors approach the topic of implementation of ethical principles in the food value chain. The book consists of three parts: Food Ethics Issues (I); Ethics for Food Professionals (II) and Food Ethics Case Studies (III). To begin with the question—how are food and ethics connected in the first place? As Rui Costa in the introduction writes, while in the past public interest focused primarily on the nutritional aspect of food, nowadays it focuses also on ethical aspects of food production—fair trade, novel foods, animal welfare, climate change and the sustainability of the natural resources (3). Consumers as well as professionals everyday face number of ethical judgments regarding food.

In the first part of the book "Food Ethics Issues", authors emphasize the main societal issues that influence the food value chain and that bring into light the importance of teaching food ethics. Harris N. Lazarides and Athanasia M. Goula in their text "Sustainability and Ethics Along the Food Supply Chain" alarms us with the fact that "the world population is increasing much faster than our capacity to increase food production" (41). The prediction is that by 2030 the growth of global population and the climate change effects will increase food production needs by 50%, energy demand by 45%, and water demand by 30% (41). This alarming fact is a result of climate changes, soil erosions, and the lack of water which leads to loss of farmland; infrastructure expansions such as large-scale recreations projects which also lead to decreasing access to farmland; uses of farmland for production of nonfood products such as oil also results in loss of farmland, as well as unsustainable production, handling, and distribution techniques which result in decreasing food production capacities. "Sustainable development", a socially and politically constructed, "a slippery and broad-ranging term", as Lazarides and Goula describe it, is defined by the Department for Environment Food and Rural Affairs as "a better quality of life for everyone, now and for future generations" (43). Precisely the lack of sustainability within the food supply chain which includes an extensive number of hu-

man activities covering all areas from farming to consumer table is a major cause of ambiguous production and consumption ethics with serious socio-economical consequences (58). Lazarides and Goula show how every part of food supply chain from raw material production, food processing, handling/distribution, food consumption, water use, food waste, and use of energy should incorporate sustainable practices. Lack of sustainability opens room for number of ethical issues and democracy, as well as human rights which come to be at stake here.

Judith Schrempf-Stirling in her text "Ethical Issues in the Food Supply Chain" focuses on frequent ethical issues within the domain of production and consumption of food. Above mentioned "food supply chain" consists of broad range of activities including agriculture and farming, food processing and manufacturing, food engineering, food transportation and distribution, food marketing, retailing and restaurants (85). Ethical question regarding the workers who wither work on the farm, in the factory or grocery store or in a restaurant have a common denominator—they are subject to low wages, health and safety risks and demanding working conditions (86). Another ethical issue concerns the food industry where many farming practices such as use of pesticides have negative impacts on eco-systems. Storage and transportation in our globalized era is responsible for high CO2 footprint; food packaging is also a practice with big environmental impact and here the role of consumer and his/her behavior treating the product waste also holds strong ethical dimension. Marketing, labeling and price promotions are practices immersed with ethics—Schrempf-Stirling asks is it ethical to advertise products high in sugar, fat and sodium content that lead to diabetes and other negative health consequences; is it ethical to target the vulnerable consumers such as children; is it ethical not to provide information needed for consumer to make a fair decision?

Anna McElhatton in her interesting text "The Ethics of Consumption" writes how food is a complex issue stated in social and cultural context and in our, Western environment, ongoing interest in "healthy food" is connected with increased consumer concern in healthy food and beverages, where "healthy" refers to essential material nutrients (63). Contemporary food chain, due to globalization, is perplexed. How McElhatton explains, on the one side there are large corporations which provide the food; at the other side, there are consumers who often consume in hedonistic way, as well as people who do not have access to quality food due to finances or time. As consequences, obesity and diabetes became epidemics of our contemporary age. According to McElhatton, healthcare, food science and associated professions have an ethical obligation to promote quality, healthy food. Consumerism is based on the paradigm of free choice which is good for consumer as well as for economy, but emphasizing that "choice without information is not real choice" (81). She singles out GM food—industries have an aversion towards positive labeling and without a mandatory labeling policy, consumers cannot make an informed choice (81).

Second part of the book "Ethics for Food Professionals" includes texts focused on ethical issues of the food production chain which can be used by teachers in order to prepare training materials as well as students, as future professionals to be aware of food ethics issues. Paola Pittia in her text

"Codes of Ethics of Food Professionals: Principles and Examples" focuses on issues concerning quality and safety. Namely, as Pittia states, increasing issues regarding food security, animal welfare, environmental occurrences, and climate change, together with labor conditions within the whole food production chain "led to the definition of a complex set of requirements to meet the expectation of the consumers and the civil society in terms of quality, healthiness, and safety of products" (107). Expanding complexity of contemporary global food system from agricultural production to consumers table requires the implementation of ethical practices. According to Pittia, first-level approach correlates with enterprises, farms, distributors, services, organizations and associations, while second-level approach is related to individuals at various levels from scientists and researchers to entrepreneurs, employees and workers. Apart from skills and knowledge, Pittia advocates that every individual is obliged with ethical responsibilities (107, 108). "Code of ethics" refers to formal documents issued for technologies and professionals by organizations and public entities and it "confirms the importance of the ethical behavior, morality, and integrity in all the activities of the food value chain" (118). Code of ethics can contribute to the work, profession and everyday life with its "guarantee of wholeness, safety, and quality of food to the consumers as well as to promote the sustainability and innovation of the food value chain" (118).

Louise Manning in his text "Corporate Social Responsibility" is also focused on ethics and he writes about corporate social responsibility—a concept gaining popularity in recent years while grasping social responsibility of individuals of different organizations as well as governments (121).

Yasmine Monterjemi in text "Whistleblowing: Food Safety and Fraud" engages with whistleblowing as a civil action. Whistleblower is defined by the Council of Europe as "any person who reports or discloses information on a threat or harm to the public interest in the context of their work-based relationship, whether public or private" (147). Monterjemi gives an example of a famous whistleblower in the domain of public health—Ignaz Semmelweis (1818–1865), a physician working in Vienna who linked high mortality rate in Viennese hospitals due to puerperal fewer with the lack of hand washing by doctors who had earlier performed autopsies. However, as it often is the case with whistleblowers, his observation was ignored, as Motarjami states, perhaps because his colleagues were not willing to change or they disapproved criticism (148). Whistleblowing is, as Motarjami indicates, often negatively perceived due to different reason: idea that information is obtained through illegal means; or that information can sabotage national security or interests; or that whistleblowers are motivated by some sort of revenge; or perhaps the idea that whistleblowing evokes some kind of denunciation or collaboration with repressive states. Despite the reason why some people perceive whistleblowing as a negative action, the reality is that whistleblowers are doing a great favor to society, and it usually comes with great personal sacrifice. Whistleblowing should be seen as a civil action, as Motarjami appeals, especially in today's globalized food supply chain where illegal actions, imprudent risk taking or negligence by purpose can have huge consequences on health and trade, as we witnessed with melamine adulteration of milk powder and the horse meat scandal (149).

As Merve Yavuz-Duzgun, Umit Altunatas, Mine Gultekin-Ozguven, and Beraat Ozcelik argue in their text "Communicating Food Safety: Ethical Issues in Risk Communication", ethics is inseparable from communication domain within food sector. They strongly argue how food safety risk communication is essential due to uncertainty of consumers about food quality and safety (165). They call upon cooperation between media, food scientists, and food industry in order to clarify uncertainties regarding food quality and safety and finally, to give consumers clear information in understandable language. In labeling and media, two important communication tools, ethical issues should not be neglected.

Also one important area where ethics is unavoidable relates to publication, as Luis Adriano Oliveira in his text "Publication Ethics" writes. Namely, as he claims, research is driving force leading to social progress, but nevertheless, the competition and pressure to publish is best suggested in phrase "publish or perish" (168). Imperative to publish might lure some researches to "shortcuts" in order to achieve a high publication rate and those "shortcuts" confronts ethical standards. As Oliviera points out, the growing number of exposed cases which testify unethical behavior illustrates the frequency of this kind of practice.

The final part of the book "Food Ethics Case Studies" consists of three case studies on ethical issue and together with underlined critical points they are excellent material for a broader discussion in classrooms. In contemporary world shaped by globalization, food becomes a burning issue. Confronted with loss of farmlands, soil erosions, lack of water, unsustainable practices within food production chain and increase in world population, food becomes major concern in global arena. As this book greatly shows, ethics should be incorporated throughout the food supply chain—from raw material production to final consumer who buys the goods. As the second part of the book shows, it also refers to scientist and researches as well as corporations and entrepreneurs. Precisely because food travels globally these days, handling it with practices containing ethics can assure quality and safety of the products and that is the reason why ethics is unavoidable when thinking about food in contemporary world.

ANA SMOKROVIĆ
University of Rijeka, Rijeka, Croatia

Ian James Kidd, Jose Medina, and Gaile Pohlhaus Jr. (eds.), The Routledge Handbook of Epistemic Injustice*, New York: Routledge, 438 pp.*

What is epistemic injustice? Who is vulnerable to it, and whom does it affect? What forms does it assume? What are its political and social consequences? And finally, how can we counter it? In a colossal volume extending over forty chapters, Ian James Kidd, Jose Medina, and Gaile Pohlhaus, Jr. have collected a rich philosophical resource on epistemic injustice. Although epistemic injustice is roughly outlined to include those cases where a person is harmed as an epistemic subject, it is, according to the authors, best understood by reference to the sheer plurality of its forms. The volume pro-

gresses in a linear fashion: after opening with a section on central theoretical concepts, it elaborates on the philosophical and political ramifications of epistemic injustice and closes with case studies of localized injustices. As the editors stress in the introduction, our social setting of incessant communication calls for special attention to the power dynamics immanent in those interactions. The authors, bridging the analytical-continental divide, draw from a diverse pool of intellectual sources. Amy Allen, for one, lauds Foucault's analysis of the role of power in the production of knowledge (187), and Lisa Guenther expands epistemological debates to Merleau-Ponty and the phenomenological tradition (195).

The volume is structured into five thematic clusters. The first, titled Core concepts, introduces the reader to the vocabulary used in discussions about epistemic injustice, pointing at potential interpretative difficulties and points of conflict. Gaile Pohlhaus, Jr., for instance, underlines the difficulty of defining epistemic injustice without inadvertently excluding those experiences of marginalization obscured by our limited social perspective (14). Although the focus is chiefly on those debates about epistemic injustice that had followed Miranda Fricker's eponymous work, the contributors acknowledge prior mentions of silencing and marginalization in feminist and intersectional discourse. With chapters defining the notions of testimonial and hermeneutical injustice, the first cluster functions as a toolbox for navigating the rest of the volume, and literature on epistemic injustice in general. The second section, Liberatory epistemologies and axes of oppression, explores how discussions about epistemic injustice interact with political currents in feminism, racial theory, post-colonial movements, and disability studies. The third thematic unit, Schools of thought and subfields within epistemology, examines different philosophical toolkits that can aid us in thinking about epistemic injustice. The mentioned sources range from continental thought, such as that of Foucault and Merleau-Ponty, to the pragmatist tradition (205) and the nascent branch of vice epistemology (223). In the fourth section, Socio-political, ethical, and psychological dimensions of knowing, the authors consider non-epistemological approaches to the epistemic injustice. While some authors inquire about the psychological phenomena of implicit bias and stereotype threat that often underlie unjust epistemic interactions, others analyze epistemic wrongs from a political perspective. The fifth and final thematic unit, Case studies of epistemic injustice, analyses the distinctive epistemic injustices that arise in specific political, scientific, professional, and social domains. Here, the authors link the unique epistemic configuration of each domain to different manifestations of epistemic injustice. For the sake of simplicity, I will follow the volume's structure in offering brief comments on some chapters of interest. Reviewing a volume that encompasses more than forty leading theorists in their field is no small feat. I will do what I can.

Gaile Pohlhaus, Jr. opens the volume with a chapter on the general phenomenon of epistemic injustice, and instantly recognizes the difficulty of defining such a broad field without omitting some of its subtler implications. Striving, then, to define epistemic injustice without unwittingly perpetrating it, Pohlhaus Jr. offers four lenses – or explanatory frameworks - for approaching the concept. The first lens approaches epistemic injustice by assessing

relationships of domination and oppression, and then explores how epistemic marginalization fits into these broader patterns. The second, drawing from the feminist tradition, focuses on intersubjectivity, or the shared epistemic institutions and practices that rear us into mature epistemic agents, and inquires about exclusions and breaches of trust. The third lens explores changes in epistemic systems, such as the systematic exclusion of specific perspectives that generates hermeneutical injustice. The fourth and final lens considers epistemic labor and knowledge production, analyzing those cases where agents are barred from contributing, where their contributions are invalidated, or where they are expected to produce excessive testimony about their social position, so that their epistemic labor is exploited (22). To prevent overly narrow definitions of epistemic injustice, Pohlhaus Jr. advises against limiting our analytical toolkit to only one explanatory lens.

Continuing with a chapter on testimonial injustice, Jeremy Wanderer defines it as a form of injustice that is categorically connected with the social practice of testimony as an interaction between a speaker proffering knowledge and a hearer in need of information (27). Although he remains true to Fricker's original account, inheriting most of her examples, Wanderer extends the analysis of testimonial injustice by considering its structural forms. Wanderer, thus, identifies three main varieties of testimonial injustice. The first is Fricker's preferred notion of injustice as transactional, wherein a hearer attributes the speaker less credibility than she deserves because they harbor prejudice towards her social group. Echoing Elizabeth Anderson, Wanderer expands upon this strictly interpersonal account and introduces the second, distributive dimension of testimonial injustice. In such cases, speakers genuinely lack the required markers of credibility – such as a refined vocabulary or a firm grasp on grammar – due to structural inequalities in access to education. Wanderer then goes even further by proposing a third variety of testimonial injustice, testimonial betrayal, an emotionally saturated phenomenon that emerges between individuals otherwise involved in intimate relationships. When we are denied trust by someone we have come to depend on, testimonial injustice assumes a distinctive weight, the experience of "humiliating rejection" (38). It remains unclear whether the patterns of identity prejudice present in testimonial betrayal at all differ from those in ordinary cases of transactional injustice.

In the third chapter on the varieties of hermeneutical injustice, Jose Medina adopts Fricker's early definition of the phenomenon. Hermeneutical injustice, then, occurs when an individual or an entire community cannot render their experiences meaningful to others due to gaps in collective interpretative resources, or, simply put, because their perspective is not accounted for in the public sphere. Yet, unlike Fricker, who depicted hermeneutical injustice as a structural occurrence without identifiable perpetrators, Medina stresses our individual hermeneutical responsibility in treating eccentric statements and expressive styles with maximum charity. Medina distinguishes between different varieties of hermeneutical injustice by referring to their source, dynamics, breadth, and depth (45). The most extreme form of hermeneutical injustice he terms hermeneutical death, and defines it as the complete loss of one's voice and one's interpretative capacities, resulting in the inability to socially situate oneself as a complete sub-

ject (41). Finally, Medina pleas for individual acts of hermeneutical resistance and insurrection. To embolden those vulnerable to injustice, we must be especially charitable in interpreting their claims, which we, due to differences in perspectives and expressive styles, might initially struggle to understand. Medina stresses that oppressed subjects, as an act of resistance, can strategically refuse to adapt to dominant conversational practices and work on building alternative rhetorical spaces.

Miranda Fricker's brief chapter on evolving concepts of epistemic injustice functions both as a retrospective review of her early work and a glance into the future of the discipline. In an effort to define the scope of discussion, she notes that, when speaking of epistemic injustice, she referred primarily to discriminatory cases of it, rather than distributive, and that the focus was on unintentional – yet culpable – displays of prejudice. Fricker then pleads for an enlivened and humane philosophy that begins its inquiries with lived experiences of marginalization, and, in a normative twist, seeks to rectify dysfunctions in present epistemic practices (57). Finally, looking to promising developments in social moral epistemology, Fricker points to case studies of epistemic injustice in the domains of healthcare and psychiatry.

Proceeding with a chapter on distributive epistemic injustice, David Coady argues that both testimonial and hermeneutical injustice can be fruitfully understood as instances of unequal distribution. In the case of testimonial injustice, we are dealing with an unequal distribution of credibility: the fact that marginalized groups are, due to prejudice, awarded less trust, entails the fact that privileged groups are given too much trust in return. If a black defendant is distrusted by an all-white jury, it is because the jury is attributing too much credibility to his white plaintiffs. Attributions of credibility, in Coady's view, sometimes function as a zero-sum game. Regarding hermeneutical injustice, different groups can be said to compete for hermeneutical power. Hermeneutical injustice can thus be portrayed as the unequal distribution of meaning-making capacities, which is unfairly tilted towards privileged social groups. Coady then inquires whether certain groups, such as Neo-Nazis, can be justifiably deprived of hermeneutical power, and calls for a more careful analysis of whether unequal distributions of credibility are always unjust (65). However, Coady's account of hermeneutical injustice might be too broad, as he seems to conflate influence on the public opinion with hermeneutical power. In other words, although Neo-Nazis might struggle to make their opinions widely known, they are not systematically prevented from attaining self-understanding and forming a vocabulary for their experiences, which are the central facets of hermeneutical injustice.

In a brisk chapter on trust, distrust, and epistemic injustice, Katherine Hawley proposes a normative account of trustworthiness in interpersonal interactions. She surveys whether trust is an appropriate attitude in different relationships and inquires about the connection between trust and social power (71). Expanding on Wanderer's account, Hawley explores the role of trust in accepting testimony. She then closes the chapter by inquiring whether a lack of trust can give rise to epistemic injustice in otherwise non-culpable attributions of credibility.

In a chapter on forms of knowing and epistemic resources, Alexis Shotwell argues that a stern focus on propositional knowledge is in itself a form

of epistemic injustice that fetters oppressed groups in improving their social position (87). She then calls for a broader account of other epistemic resources, such as emotions, skills, tacit knowledge, social position, and embodiment. Shotwell, criticizing traditional thought experiments which endorse a distinction between knowing that and knowing how, claims that we base our identities on a more vibrant array of epistemic resources, and that the lived experiences of disability and bodily change cannot be grasped by reference to propositional knowledge alone. Epistemic justice, according to Shotwell, should account for the epistemic systems that oversee social relationships, emotions, and skills, rather than mere propositional transactions.

Lorraine Code, reflecting upon her concept of epistemic responsibility, inquires why analytical epistemology had long lacked the vocabulary to form a coherent account of responsible epistemic behavior. Due to its restrictive individualism, inherited from logical positivism, analytical epistemology was reluctant to place its subject within society, as a knower who deliberates, feels, learns from others, and engages in interpersonal interactions (91). This self-imposed limitation to an abstract and isolated subject hampered it in recognizing the salient social aspects of being a responsible agent. As social epistemology expanded to include ethical and political concerns, prominent in discussions about epistemic injustice, talk of epistemic responsibility gained an additional normative dimension. What, then, are the requirements of responsible epistemic conduct? While epistemic responsibility cannot be reduced to a universal set of rules, Code argues that we should always approach our agency as situated within a particular "epistemic imaginary," an intellectual system akin to a Kuhnian paradigm or a Foucauldian episteme, which defines all epistemic practices in our social context. Code concludes the article by underlining the relevance of epistemic responsibility in the era of social networking and climate change denial, proposing fruitful topics for further debate.

Charles W. Mills closes the first section by rehabilitating the Marxist concept of ideology. Mills starts by noting that progressive academics have abandoned the notion of ideology in favor of postmodern conceptual tools, rendering debates about false consciousness either outmoded or seemingly conspiratorial (100). He explains ideology by noting that power differentials entail harmful epistemic consequences for all involved social groups, in that privileged groups actually cannot comprehend the social experience of oppressed factions. Central to the notion of ideology, according to Mills, is its materialism, or the fact that privileged groups have a vested socioeconomic interest in depicting extant inequalities as necessary. Mills illustrates this with the example of modern racism and explains how anti-black ideology attempts to depict socially generated inequalities as natural. Connecting ideology with contemporary discussions about epistemic injustice, he then argues that marginalized groups, albeit vulnerable to hermeneutical injustice, enjoy unique epistemic access to their social experience, and can use this advantage to form alternative rhetorical spaces.

Patricia Hill Collins opens the second section with a chapter on intersectionality, defining it as the project of connecting resistant forms of knowledge and using this unity to subvert oppressive social structures (115). As intersectionality recognized that the experiences of belonging to a specific

gender, race, class, ethnicity, and sexuality overlap, it sought to create a platform for marginalized groups to voice their problems and demand social justice. Yet, according to Collins, its entrance into an academic context was met with persistent ignorance. Since intersectionality's focus on lived experiences clashed with the dominant epistemological paradigm of asocial objectivity, its pioneers struggled to connect the political project of attaining social justice with academic agency. Drawing from the history of black feminist thought, Collins shows how intersectionality was, within academia, sanitized and stripped of its emancipatory potential. Collins then points at those academic practices, such as peer reviews and keynote lectures, which silence more radical intersectional endeavors, and calls for resistance to epistemic injustice.

In her chapter on feminist epistemology, Nancy Tuana trails how standpoint theory aimed to unearth the interests implicit in professedly neutral scientific practices. Instead of starting with Fricker's work, Tuana reverses the process, showing how debates on epistemic injustice resumed the ethical and political project launched by feminist epistemology. Feminist epistemologists, in Tuana's recounting, focused on the subject of knowledge as a socially situated agent at the crossing of different identities, and explored how power differentials mold our ability to participate in intellectual exchanges. More specifically, they sought to disclose just what kind of person traditional epistemology presupposed by its asocial knower. Once this universal subject was revealed to be white, male, educated, able-bodied, and economically privileged (126), liberatory epistemologies strived to acknowledge alternative perspectives and to oppose the institutional silencing of ostensibly strange or overly subjective voices. When writing about the subject of knowledge, Tuana explores which social features we must possess to be recognized as a credible epistemic agent. Much like Mills, she stresses that vulnerable groups have unique epistemic access to their social experience, and that privileged groups have a vested interest in remaining ignorant to the fact of their unjust opportunities. Tuana closes the article by recognizing the limits of her perspective and appealing for further opposition to epistemic violence.

With the chapter "Knowing disability, differently," Shelley Tremain concludes the second section by arguing that debates on epistemic injustice have failed to acknowledge disability. According to Tremain, this omission, evident in the usage of ableist metaphors, such as "epistemic blindness" and "epistemic deafness," renders social epistemology short of a fully intersectional approach (175). Tremain first claims that disabled individuals are, due to social stigma, particularly vulnerable to unjust hermeneutical exclusions that cannot be disregarded as mere epistemic bad luck. To further substantiate her point, Tremain shows that Fricker's prized example of testimonial injustice, the rigged trial against a black man, Tom Robinson, from Harper Lee's novel *To Kill a Mockingbird*, does not account for the fact Robinson was disabled. This fact, along with his race, class, and gender, played a crucial role in shaping his identity as an emasculated "conceptual impossibility" in the eyes of his prosecutors (181). Since whites usually equate black men with virility, physical force, and callousness, they struggled to make sense of Robinson, a disabled black man who showed empathy for his

professed victim. Inheriting Foucault's concept of *apparatus*, the system of discourses, institutions, laws, administrative measures, moral norms and scientific statements that define some phenomenon in a given historical moment, she then shows that, by including the apparatus of disability in our analyses, we construct a philosophically and politically more complete, and thus more emancipatory, account of epistemic injustice. Tremain ends by urging for a more attentive approach to disability in debates on epistemic injustice, and in social epistemology at large.

In her chapter on Foucault, Amy Allen dispels some common misconceptions about his attitude towards truth and argues that his thought is a fruitful resource for social epistemology. She focuses on three aspects of Foucault's work. First, Allen explores his dual theory of power as both constitutive and agential. Power, in Foucault's rendition, both structures us as social subjects and takes places between subjects who, on a quotidian level, internalize and reproduce social power relations. Allen links this distinction to Fricker's concepts of testimonial and hermeneutical injustice, showing how it can inform a richer understanding of epistemic harm. Second, she uses Foucault's analysis of knowledge regimes to offer an alternative account of hermeneutic injustice. Foucault's analysis of the historical processes by which knowledge is justified, institutionalized, and, finally, legitimized as credible knowledge, can help us understand the epistemic exclusions that generate hermeneutical injustice. Third, Allen rehabilitates his notion of genealogy, a "counter-memory that articulates subjugated knowledges," as a model of resistance against epistemic injustice (187). By coupling marginalized experiences with historical erudition, we can place them within the appropriate context and attempt to counter them. Allen, wondering why Foucault is not more readily cited by scholars studying epistemic injustice, attributes this oversight to the animosity between analytical and continental philosophy, and to the widespread perception of Foucault as an epistemic reductionist. She concludes the chapter by underlining the emancipatory potential of Foucault's thought.

Sandorf Goldberg proceeds by analyzing epistemic injustice from the perspective of social epistemology. He broadly defines social epistemology as a philosophical branch concerned with the epistemic relevance of other minds, one focused on the way we acquire, store, and communicate information in a social setting. Goldberg introduces his brand of social epistemology as a middle way between Steve Fuller's relativistic project and Alvin Goldman's more normative approach: he acknowledges that knowledge is produced in a social setting, but, like Goldman, retains objective standards for its justification. According to Goldberg, knowledge communities, formal and informal alike, manage their epistemic practices by imposing certain normative expectations upon other people. When approaching someone as a knower, regardless of whether they are an expert or a family member with whom we share our daily chores, we will expect them to substantiate their knowledge with a certain degree of evidence, or to display a certain degree of epistemic responsibility (215). If these expectations are illegitimate, they can generate epistemic injustices. First, injustice occurs when certain individuals are excluded from participating in epistemic practices, or when their contributions are invalidated, such as in male-dominated scientific

communities. Second, social practices can warrant normative expectations that treat people unjustly. Goldberg illustrates this with the example of low-income schools that, due to structural limitations, have lower expectations of its students, and thus fail to rear them into fully functioning epistemic agents. Third, seemingly legitimate social practices can be enforced in a way that treats certain groups unjustly. Goldberg brings this point home by describing teachers who only interact with more successful students, thus effectively excluding struggling pupils from disadvantaged backgrounds, and notes cases of referees overlooking ethnic-sounding job applications. He closes the article by appealing for the utility of social epistemology in thinking about epistemic injustice.

Writing from the perspective of virtue epistemology and its nascent branch of vice epistemology, Heather Battaly examines whether testimonial injustice can be understood as an epistemic vice. Battaly starts by broadly defining epistemic vices as bad cognitive dispositions that impede us in attaining knowledge, and then distinguishes among three notions of epistemic vice. First, there is effects-vice, the general stance that vices are dispositions, both constitutive of our characters and entirely impersonal, that result in adverse epistemic effects. Second, the notion of responsibilist-vice implies that we have a bad character trait for which we are responsible, such as the motivated tendency to side with the easier solution, or to uncritically uphold the status quo. Third, as a middle way, Battaly introduces personalist-vice, the stance that epistemic vices are intrinsically bad cognitive traits which are not entirely under our control (228). She then argues that testimonial injustice usually takes the form of a personalist-vice, as we are partially exonerated for inheriting prejudiced beliefs from our social context, but still display bad cognitive traits. Battaly closes the article by encouraging further debate about whether we can be blamed for implicit epistemic vices that, due to social conditioning, slither beneath our conscious control.

In the opening chapter, Jennifer Saul examines the concepts of implicit bias and stereotype threat, defining them, in the above order, as the automatic tendency to identify a social group with certain features, and the fear that stereotypes might affect the way we are perceived by other people (235). Saul then denies that they should be treated as cases of epistemic injustice. First, according to Saul, someone can harbor implicit biases inherited from their social context without ever committing testimonial injustice. Simply put, while implicit bias is strictly a psychological disposition, testimonial injustice requires interaction between a speaker and a biased hearer. Second, she argues that not all implicit biases are related to credibility. It is unclear, though, whether Fricker herself, once she had defined testimonial injustice, indeed limits it to deflated attributions of credibility, or whether she allows for broader judgments of character. Third, implicit biases are wider than epistemic injustice in that they can also be positive, such as when whites automatically associate other whites with positive features. This claim depends on whether we treat testimonial excess as a form of epistemic injustice, and whether we, as Coady does, consider testimonial excess a distributive epistemic failure. The link between testimonial injustice and stereotype threat is, in Saul's recounting, that of a self-fulfilling

prophecy: members of stigmatized groups, fearing their testimony will not be received well, indeed deliver a shiftier and less convincing performance. Saul illustrates this with the examples of female mathematicians who underperform due to pressure, and Aboriginal rape victims offering clumsy responses to hostile questions in court (238). Speaking of hermeneutical injustice, Saul notes that the concepts of implicit bias and stereotype threat had filled critical hermeneutical gaps, and that implicit biases often potentiate hermeneutical marginalization. Finally, Saul inquires whether individuals can be trained to overcome their implicit biases. She concludes that, given these cognitive constrains, appeals to individual virtue must be supplemented with institutional measures for countering epistemic injustice.

Lorenzo C. Simpson proceeds with a hermeneutical approach to political agency. He draws a distinction between first-order agency, or the ability to act, and second-order agency, or the epistemic preconditions of choosing a particular action. Simpson argues, albeit obliquely, that individuals who cannot fully understand their social experience and thus fail to make good choices are both epistemically and politically harmed. A correct understanding of our present state is, then, a precondition of just political agency. By asking "how things appear from the first-person perspective from which these choices were made," we can learn whether someone was epistemically hampered from making a better and more just decision (254). This approach, which he terms "narrative representability," assesses the socially available courses of action for members of particular social groups. It also demotivates us from fallaciously "psychologizing the structural," or, simply put, from making the false assumption that disadvantaged groups fail to thrive because of innate personal deficiencies, rather than because of structural constraints. Simpson closes the article by stressing that the inability to articulate our social experience and the absence of democratic deliberative platforms are in themselves epistemically unjust.

Sally Haslanger closes the fourth section by analyzing the relationship between objectivity, epistemic objectification, and oppression. What Haslanger wants to explore is how the notion of objectivity sustains oppression by portraying the socially conditioned epistemic weaknesses of disadvantaged groups as inherent to their nature. Haslanger first detects three ways of thinking about objectivity: objective reality, objective discourse, and objective knowledge. While objective reality pertains to the world as it is, regardless of how we conceptualize it, objective discourse refers to discourses for expressing facts, and objective knowledge encompasses claims accessible to any rational agent (279). Objectivity is, according to Haslanger, closely linked to certain forms of essentialism, the idea that observed regularities express a thing's nature. Essentialism often entails normative assumptions, in that what is statistically "normal" of a thing becomes desirable, or representative of its ideal form. The failure to recognize that certain features are conditioned by social circumstances "leads us to attribute the regularities to something intrinsic to the agents" (284). In Haslanger's example, if women are structurally barred from attaining decent education, their seeming inability to participate in the public sphere may be fallaciously attributed to innate domesticity. Similarly, a social structure that unloads the burden of childbearing on women sustains the essentialist claim that women are

inherently more nurturing. This kind of status quo reasoning, when coupled with unjust social institutions, results in the looping effect: members of vulnerable groups are conditioned to attain the unseemly characteristics that are then considered part of their nature. Once the social origin of present inequalities becomes invisible, status quo reasoning justifies these inequities by naturalizing them. Haslanger exemplifies this with the case of black people receiving inadequate education, which denies them the relevant markers of credibility and confines them to poorly paid menial labor. She then identifies three distinct forms of objectification that lead to epistemic injustice: ideological, projective, and Kantian objectification (285). It is ideological objectification that conceals the contingent social roots of our unjust epistemic practices and portrays artificial inequalities as natural. Haslanger ends by stressing that a focus on individual rather than structural solutions and a bias towards stability contribute to epistemic and social injustice, and that epistemic justice will require us to dismantle unjust social structures.

To sum up, Kidd, Medina, and Pohlhaus, Jr. have compiled a fruitful collection of topics that warrants philosophical attention and will surely inspire further inquiry. The volume, however, harbors a general tendency that is worth noting: its authors, aiming for maximum inclusiveness, almost unanimously overlook the question of epistemic quality. There is no mention of whether distrusting underprivileged individuals who lack the relevant markers of credibility, albeit it entrenches inequality, can sometimes be epistemically justified. This trend of disregarding epistemic quality, or its lack, actually makes the authors less attentive to the systemic barriers that prevent vulnerable groups from attaining a decent education. The desire to attain social justice thus results in less social justice, as we end up with an incomplete understanding of the social institutions which, through inequitably distributed education and inaccessible deliberative platforms, reproduce unjust epistemic asymmetries. As this insight was fully present in Elizabeth Anderson's much earlier article on the structural causes of epistemic injustice, we may wonder whether social moral epistemology should want to revisit a more grounded approach. Nevertheless, anyone interested in epistemic injustice is well advised to expand their analytical vocabulary with the tools here offered, and certain topics, such as Carel and Kidd's analysis of epistemic wrongs in healthcare, promise fecund practical applications. There is certainly more philosophical work to be done, as most authors diagnose social maladies, leaving their solutions open for future discussions.

HANA SAMARŽIJA
University of Zagreb, Zagreb, Croatia

Maria Paola Ferretti, The Public Perspective. Public Justification and the Ethics of Belief*, London: Rowman and Littlefield, 2018, 196 pp.*

Ever since John Rawls published *Political Liberalism*, political justification has been one of the central topics in political philosophy. How can citizens, endorsing substantively different and often incompatible yet reasonable comprehensive doctrines acknowledge the same laws and political decisions as legitimate? In other words, how can citizens recognize the authority of some laws or decisions when they simultaneously see them as morally wrong or epistemically incorrect? Almost all scholars, following Rawls, solve this problem by ascribing some form of legitimacy-generating potential to the decision-making procedures that have produced these contestable results. The procedure, they claim, has some moral or epistemic qualities that all qualified (or reasonable) citizens can recognize and affirm, and it is because of these qualities that citizens can endorse laws and decisions even though they find them substantively wrong or incorrect. Maria Paola Ferretti's *The Public Perspective. Public Justification and the Ethics of Belief* follows this line of thought but introduces an innovative and original approach. Namely, Ferretti claims that the practice of political justification is possible only where people endorse a common ethics of belief, a cluster of epistemic and moral norms that guide formation and reformation of the beliefs that inform our public perspective (1). Her position thus departs from many existing accounts of public justification (particularly those presented by Rawls and Gaus) and focuses on (i) common epistemic rules and (ii) a shared commitment to a regulative, non-dogmatic idea of truth as necessary components of the process of public justification.

The book introduces or brings into focus many important and underdiscussed ideas. For example, most authors assume there is an inherent gap between our factual beliefs and our values and normative claims. Ferretti challenges this sharp division and asserts that our factual beliefs often shape our normative claims—some of the worst failures in citizens' normative deliberation (e.g. The Holocaust and genocide of indigenous peoples) had to be supported by corrupt science and pseudoscience (e.g. Nazi eugenics). Furthermore, most authors, following Rawls and Gaus, endorse the idea that we have a moral (and not epistemic) reason to abide by the constraints of public reason, i.e. to abstain from introducing the arguments that other (qualified) citizens cannot affirm or recognize as intelligible in the public deliberation. Ferretti, on the other hand, argues that there is a strong epistemic reason not to introduce some contestable claims in the public deliberation, and differentiates between epistemic commitments we have when justifying some belief to ourselves and when justifying the same belief to the public. These thought-provoking ideas, paired with imaginative and resourceful argumentation, are alone a good reason to give the book a thoughtful consideration.

However, apart from addressing some of these interesting questions, Ferretti's book undertakes a far more demanding task—it aims to establish a link between our moral requirements and epistemic commitments, thus offering a mixed account of political legitimacy. The book draws on

a tradition that goes back to Locke and his ethics of belief as developed in *An Essay on Human Understanding*. Locke postulated that we have an 'alethic obligation' to regulate our beliefs so that they track truth (or what is most probably true) and recommends rationally revising all beliefs that are sources of conflict or debate. Ferretti's goal is to draw on Locke's lesson in order to answer some contemporary questions about public justification. The ethics of belief for citizens of a liberal democratic society is based on logic, factual evidence and the state of the art in sciences. These epistemic rules do not ensure that citizens reach consensus in all situations, but that they can meaningfully talk to each other in a way that is adequately public (3). The ethics of belief thus represents an integral part of the ethics of citizens living together in a plural society as free and equal moral and epistemic agents.

The book is divided in seven chapters. The introduction presents the main aims of the book, but also displays the central motives that urged Ferretti to complete the manuscript. Namely, the declining trust in experts and the rapid increase of fake news in the media have started our transition to "post-truth societies" (2, 170), where ethics of belief and the aspiration towards right or correct laws and decisions has been disregarded, and the only hope for public justification rests in purely-procedural (and non-epistemic) qualities of a decision-making procedure. This shift can have disastrous impact on the quality of our political decisions but can also distort the democratic process and turn it into a simple majority rule characterized by domination of one group over the other. Ferretti sees the ethics of belief as a regulating principle that can improve our decision-making process, but also define the proper role of science in a democratic society.

Second chapter frames the discussion by setting the idea of public reason as a regulative ideal that determines the kind of reasons that can be introduced in the process of public justification. Political authority thus has to be justified by reasons available to all reasonable (qualified) citizens—we respect others as free and equal moral agents by justifying coercive laws and policies through public (and only public) reasons. Ferretti distinguishes her own position from the two dominant alternatives: consensus view defended by John Rawls and convergence view defended by Gerald Gaus. Both positions, Ferretti claims, have serious flaws. Rawls' conception of public reason, based on *shared agreement* on the premises in the justification process, is too conservative (20) since it precludes new considerations and inputs (e.g. new scientific discoveries, insights from the perspectives of minority groups) from entering the public deliberation and challenging the commonly accepted premises. Gaus' view, based on *joint agreement* on laws and policies (rather than reasons supporting them), lacks publicity (28): citizens are unable to see the agreement as a *joint* endeavor since they cannot critically evaluate laws and policies from the perspective of all others. Ferretti believes her position, based on Lockean social epistemology, can successfully avoid these objections.

After a somewhat unnecessary sketch of a debate between foundationalism and coherentism, where the author ends up endorsing a moderate version of foundationalism, Ferretti introduces *alethic obligation*, Locke's claim that each epistemic agent should strive to believe what is true. Strength

of our beliefs should be proportionate to the degree of probability that the proposition in question is true (44). Furthermore, the required degree of probability depends on what is at stake: if we are going to make coercive laws that will affect others around us, we need to ground them in well-regulated and very probable beliefs. Ferretti's view proceeds to embrace a conclusion similar to that suggested by Robert Talisse, Cheryl Misak and other proponents of pragmatist account of epistemic democracy—very probable beliefs are those produced by an epistemically reliable procedure. If we see other people as free and equal moral and epistemic agents, we have both moral and epistemic obligation to show respect for their autonomy by adhering to such an epistemically reliable procedure when we make collective decisions.

The fourth chapter brings a comprehensive overview of the use of the term 'reasonable' in contemporary liberal philosophy. Ferretti rejects distinction between reasonable and non-reasonable people, as well as counting only the former as participants in public justification (75). We should focus on reasonable beliefs rather than on people as reasonable. As noted in the third chapter, only very probable beliefs—those that can be publicly justified—should be used to ground laws and public policies. One might thus have reason to hold onto her belief that cannot be publicly justified, but she cannot use such a belief in the collective decision-making process. For example, Galileo had good reasons to personally believe that the Earth moves, yet the available evidence was insufficient to present a public justification for such a claim (it become available in the 19th century). Therefore, founding laws and policies (e.g. calendar reform) on heliocentric thesis could not be done publicly, though Galileo was justified in following Copernicus' view. Following Locke, Ferretti claims that we have freedom to believe what appears true to us, and we have a duty to justify those beliefs when we want them to have impact on decisions that have public relevance (88, 92).

Some might remain unpersuaded regarding the Locke's method and its ability to solve complex disagreements and conflicts in a world characterized by reasonable value pluralism. However, Ferretti holds that many value disagreements are fueled by disagreements on facts, often caused by epistemological and political asymmetries (100). Citizens often overestimate their own expertise or the credibility of evidence in support of their favorite (descriptive) beliefs (e.g. debates on dioxins, GM food, hormone-containing beef, nuclear waste storage, the greenhouse effect and cloning), which in turn shape their normative attitudes. Locke's ethics of belief can help us resolve disputes on many of these (descriptive) issues, and can thus prevent some value disagreement from emerging. Furthermore, author claims, Locke's method for belief reformation can help us in ranking the desirability of political aims (104) we want to achieve, thus resolving some of the value conflicts. Of course, Ferretti is well-aware that, even when appropriately applying an ethics of beliefs, logic and consistency, people do not always reach conclusive agreement.

The sixth chapter discusses the limits in the application of the ethics of belief. Ferretti has already argued, in fourth chapter, that demanding requirements of the ethics of belief should not be applied on beliefs one does not use in the process of public justification. We are not required to justify

to the others why we hold a specific belief, unless we want to ground some coercive law or public policy on such a belief. Ferretti discusses the value of freedom of thought, rejecting some instrumental justifications (e.g. J. S. Mill) and endorsing the idea that protection of a sphere of personal freedom enjoys a certain priority in our political reasoning (138). Her justification of the priority of (equal) freedom follows Rawls (and Quong) and is based on citizens' equal moral status.

The final chapter introduces several challenges to Ferretti's position. Our beliefs are formed within a culture and are often influenced by a cultural tradition. When different cultures coexist within a single state, their members might find it impossible to collectively practice ethics of belief since they have substantively different assessments of probability of some key beliefs. Ferretti claims that, in some instances, there are good reasons not to press with too demanding constraints of the ethics of belief—some cultural communities should be left to arrange some aspects of their public life. The reason for this, however, is not in the value of particular cultures, but in the equal moral status of citizens endorsing different moral doctrines and cultural practices (164).

Ferretti's book is undoubtedly an important contribution to the ongoing debate on public justification. Her focus on Locke's *alethic obligation* and the ethics of belief represents a novel and underexplored approach that tries to unite moral and epistemic considerations in the process of collective justification of laws and policies. However, I would like to stress two minor difficulties that need to be addressed. First, some might argue that Ferretti misinterprets Rawls when she criticizes his consensus approach to public reason as too conservative. Emancipatory ideas, but also new scientific discoveries, challenge the commonly accepted ideas and rules that they support. Rawls' account is criticized to be too rigid to appreciate these new perspectives. Namely, it seems that Rawls addresses idealized citizens in idealized conditions and is thus unable to accommodate new discoveries or shifts in public perspective that happen in the real world. However, Rawls' four stage sequence can be used to tackle this worry. He clearly states that the political decision-making process consists of four stages: first we determine principles of justice (in idealized conditions, behind the veil of ignorance), and then we proceed to draft a constitution, form appropriate legislation, and finally, we implement this legislation on particular cases, through public administration and courts. Only the first stage takes place in idealized conditions—constitutional changes can be motivated by the electorate, as was the case in "the three most innovative periods in American constitutional history: the founding of 1787–91, Reconstruction and New Deal" (PL, 406). Rawls indicates that the purpose of an amendment is to adjust basic constitutional values to changing political and social circumstances, or to incorporate into constitution a broader and more inclusive understating of those values (PL, 238). The three amendments related to the Civil War all do this (abolition of slavery), as does the Nineteenth Amendment granting women the vote. These changes were, at least in part, conducted after widespread false factual beliefs (regarding the inferiority of women or African Americans) were disputed. It seems that Rawls' public reason is not so conservative. Except for the first stage (principles of justice), when we

consider idealized citizens behind the veil of ignorance, all other are (to a certain degree) performed by real citizens in a real world and can take into account new scientific discoveries and shifts in the public perspective.

Second, it is important to emphasize that Ferretti and scholars she addresses (e.g. John Rawls) often write about substantively different things. Rawls refers to public justification through shared reasons as a precondition for political legitimacy. Liberal principle of legitimacy specifies the minimum that has to be met in order for the exercise of coercive political power to be fully proper—this power has to be exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason. (PL, 137). Ferretti, however, does not address the question of political legitimacy. She focuses on the ethics of belief as political ethics, thus setting a more demanding set of constrains than Rawls does. Considering they are writing about different things (what makes a procedure legitimate / what makes a procedure morally justified), it seems that Ferretti's contribution does not represent an alternative to Rawls' account, but a completely new contribution in a separate discussion.

IVAN CEROVAC
University of Rijeka, Rijeka, Croatia

Table of Contents of Vol. XIX

Articles

Book Reviews