5

The Metaphilosophy of Language*

Daniel Cohnitz

1. The dilemma of metaphilosophy

Within the field of philosophy, metaphilosophy has moved to the centre of attention in the past decade.[1] This development was in part provoked by the rise of experimental philosophy as an alleged alternative to standard armchair philosophy.

Of course, the fact that methodological questions in philosophy are now assessed more systematically than they were in the past[2] is a positive development. However, some of the discussions between methodological conservatives and experimental revolutionists appear to be surprisingly naïve. Typically, it seems to be assumed, on both sides of the debate, that there is such a thing as a general methodology of philosophy, which can be discussed and assessed without paying attention to the peculiarities of the different sub-disciplines of philosophy.

But it is easy to show that this is not a terribly plausible assumption. Methodological questions arise relative to the aims or goals one has set for oneself. They are questions of the type "What is the best way to do X?" But it isn't clear at all that all areas of philosophy plausibly involve the same X here at any (even minimally) interesting level of abstraction. Let's take a look why this might be so. Perhaps the best candidate for philosophy's general goal is to find the truth. After all, philosophy is a compound of φιλεῖν and σοφία and translates as Love of Wisdom, so shouldn't we conceive of philosophy's aim in general as cognitive?

First of all, it is not clear that it is even true that all philosophy aims at truth. Is practical philosophy, ethics in particular, an attempt to discover moral truth? If you are a non-cognitivist about ethics, you will dispute such an assumption. Moreover, ethics isn't the only area in which a non-cognitivist attitude can make sense. For example, all areas of philosophy that conceive of their enterprise as one of providing explications in Rudolf Carnap's (1950) sense of the word, the aim of philosophical inquiry might rather be seen as the development of normative proposals (e.g. proposals to use certain concepts in certain refined ways).

Second, even if we agreed that at least large parts of philosophy aim at truth and knowledge, it isn't clear that 'truth and knowledge' alone characterize an aim that could determine a specific methodology. Different kinds of truths might require different methods for their discovery. Is our epistemic access to metaphysical truths the same as our access to conceptual or logical truth? Those who hold on to an analytic/synthetic distinction might doubt that it is. Logical truth is perhaps plausibly accessible a priori, but at least some metaphysical truths might be only a posteriori knowable.

Rejecting the a priori/a posteriori distinction improves the situation only slightly. Sociology, linguistics, history, particle physics and astronomy are all a posteriori disciplines, but they have very different methodologies that display similarities only on a very high level of abstraction. The reason for this is that their subject matter and our epistemic access to that subject matter are very different, even if this access is in all cases a posteriori.

It seems that in order to argue for a general methodology for philosophy as a whole, one would first need to answer a number of substantial philosophical questions in specific ways. Are moral questions cognitive? Are there metaphysical truths that are only a posteriori knowable? Is logic a priori? Are there any analytic truths that are knowable on the basis of linguistic competence alone? When dealing with questions like these, one is already engaging with central topics in first-order philosophy, one is already doing philosophy. But how can one hope to formulate a methodology of philosophy, if that presupposes answers to questions like these?

In this chapter, I will use the example of philosophy of language, and in particular the recent debate concerning the role of intuitions in choosing between alternative theories of reference, to demonstrate which specific assumptions and considerations enter into methodological discussion. Though, as we will see, these assumptions can't plausibly be generalized to other areas of philosophical inquiry.[3]

In Section 2, I will begin with a short summary of the methodological discussion concerning the role of intuitions in philosophy of language, which developed in response to empirical results by experimental philosophers that seemed to indicate a certain cross-cultural variability in intuitions about the reference of proper names. One central question in this debate is which and whose intuitions should count in the first place for theory choice on the standard account that experimental philosophers intend to criticize. As we will see, the experimental philosophers as well as many of their critics seem to assume that the relevant intuitions are certain metalinguistic judgements the reliability of which depends on how correctly they inform us about an independent realm of objective semantic facts.

This realm of semantic fact that a theory of reference is supposed to describe will be the topic of the third section. How should we conceive of the subject matter of theories of reference? Are these semantic facts indeed independent of the intuitive interpretation and production of linguistic items by competent speakers? I will argue that this assumption would be very implausible and at odds with the explanatory aim of philosophical semantics, namely to make a systematic contribution to the explanation of successful linguistic communication. Far from being independent of the intuitive interpretation and production of linguistic items by competent speakers, the realm of semantic facts is instead constituted by it.

Section 4 will discuss how the judgements that seem to play a methodological role in the standard methodology of philosophy of language relate to the intuitive interpretation and production of linguistic items by competent speakers. I will argue that these judgements are best understood as reports of intuitive interpretations or productions. Thus the question of how reliable these judgements are is a question of how reliably they report what they are supposed to report. As we will see, there are good *prima facie* reasons to assume that these reports are reliable.

However, since the reliability of these reports is ultimately an empirical question, Section 5 will sketch how some methods of psycholinguistics could be used to determine their actual reliability empirically.

In the sixth and last section, the results of the discussion of this example will be summarized and the extent to which the results reached can be generalized will be discussed. I will argue that the generalizability of these results is very limited and probably doesn't go beyond philosophy of language (and even within philosophy of language, they don't generalize to all topics or questions).

2. The experimentalist challenge

It is probably fair to say that the recent discussion concerning the role of intuitions in philosophy of language started with the publication of the provocative paper "Semantics, Cross-cultural Style" by Edouard Machery, Ron Mallon, Shaun Nichols and Stephen Stich (2004).[4] In the paper the authors provide reconstructions of two families of theories of reference; that is, theories that are supposed to explain how certain linguistic expressions (e.g. proper names) refer to objects in the world. The two families distinguished are descriptivist theories of reference (characterized by D1 and D2) and causal-historical theories of reference (characterized by C1 and C2):

> Descriptivist View
>
> D1. Competent speakers associate a description with every proper name. This description specifies a set of properties.
>
> D2. An object is the referent of a proper name if and only if it uniquely or best satisfies the description associated with it. An object uniquely satisfies a description when the description is true of it and only it. If no object entirely satisfies the description, many philosophers claim that the proper name refers to the unique individual that satisfies most of the description . . . If the description is not satisfied at all or if many individuals satisfy it, the name does not refer.
>
> (Machery et al. 2004, B2)

> Causal-Historical View
>
> C1. A name is introduced into a linguistic community for the purpose of referring to an individual. It continues to refer to that individual as long as its uses are linked to the individual via a causal chain of successive users: every user of the name acquired it from another user, who acquired it in turn from someone else, and so on, up to the first user who introduced the name to refer to a specific individual.
>
> C2. Speakers may associate descriptions with names. After a name is introduced, the associated description does not play any role in the fixation of the referent. The referent may entirely fail to satisfy the description.
>
> (Machery et al. 2004, B2–B3)

Machery, Mallon, Nichols and Stich claim that in philosophy theories are chosen if they accord with the intuitions of philosophers when evaluating actual and hypothetical cases in the domain of these theories and theories are rejected if their predictions are in conflict with the intuitive judgements of philosophers. Also in the choice between these two families of theories of reference, a choice was made based on such intuitive judgements. In particular, Saul Kripke had described hypothetical cases in *Naming and Necessity* (1980), the intuitive evaluation of which was compatible with the predictions of a causal-historical theory but contradicted the predictions of descriptivist theories.

Inspired by previous research on cultural variation in cognitive strategies between Westerners (Ws) and East Asians (EAs) (c.f. Nisbett et al. 2001; Nisbett 2003), as well as results about cultural variation (between the same groups) in intuitive judgements about thought experiments in epistemology (Weinberg et al. 2001), Machery et al. conjectured that a similar cultural variation should also be found for the hypothetical cases described by Kripke. Thus, their paper describes two experiments intended to test the following hypothesis:

> When presented with Kripke-style thought experiments, Ws would be more likely to respond in accordance with causal-historical accounts of reference, while EAs would be more likely to respond in accordance with descriptivist accounts of reference.
>
> (Machery et al. 2004, B5)

In order to test this hypothesis Machery et al. formulate four vignettes featuring hypothetical cases that are modelled after the cases discussed by Kripke. One of these, the so-called 'Gödel/Schmidt-case', reads as follows:

> Suppose that John has learned in college that Gödel is the man who proved an important mathematical theorem, called the incompleteness of arithmetic. John is quite good at mathematics and he can give an accurate statement of the incompleteness theorem, which he attributes to Gödel as the discoverer. But this is the only thing that he has heard about Gödel. Now suppose that Gödel was not the author of this theorem. A man called "Schmidt", whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and claimed credit for the work, which was thereafter attributed to Gödel. Thus, he has been known as the man who proved the incompleteness of arithmetic. Most people who have heard the name "Gödel" are like John; the claim that Gödel discovered the incompleteness theorem is the only thing they have ever heard about Gödel. When John uses the name "Gödel", is he talking about:
>
> (A) the person who really discovered the incompleteness of arithmetic? or
>
> (B) the person who got hold of the manuscript and claimed credit for the work?
>
> (Machery et al. 2004, B6)

Vignettes like this were presented to undergraduate students from Rutgers University and the University of Hong Kong. In response to vignettes modelled after the Gödel/Schmidt-case, about two-thirds of the Ws chose answer (B), while only about one-third of the EAs chose that answer.

Machery et al. draw far-reaching conclusions from this result. In (Machery et al. 2004) they argue that philosophers of language should reconsider their methodology. In a later paper (Mallon et al. 2009) the same authors argue that theorizing about reference should be abandoned completely, since there seem to be no viable methodological alternatives. In other places, experimental philosophers argue that these experiments show that the standard methodology of analytic philosophy (viz. the consideration of hypothetical cases in theory choice) is 'bankrupt'.[5]

These radical claims about philosophy as a whole, and the methodology of philosophy of language in particular, provoked a lively debate. In this debate we find two sets of objections against the argumentation by Machery et al. One set objects to the analysis of standard methodology by Machery et al. The other set objects to details of the experiment:

1. Objections to the characterization of standard methodology
   1.1 Intuitions don't play the special role in philosophy that experimental philosophers assign to it. In particular, intuitions aren't the foundation for philosophical arguments. Therefore the empirical investigation of the possible cultural variation of intuitions is simply irrelevant for philosophical methodology. (Cappelen 2012)
   1.2 Even if the intuitive evaluation of the Gödel/Schmidt-case played some role, Kripke presented many more arguments in *Naming and Necessity* which were much more relevant and significant and which are independent of intuitive judgements about hypothetical cases. (Deutsch 2009, 2011a; Martí 2014)[6]
   1.3 The intuitions of laymen shouldn't play a role in the evaluation of theories of reference. If intuitions should be given an evidential role, then it's the intuitions of experts, that is, professional philosophers of language and perhaps linguists. (Devitt 2011a)[7]

2. Objections to Details of the Experiment
   2.1 The result of the experiment can't inform us about variation in the relevant intuitions, since the answers (A) and (B) were formulated as metalinguistic judgements. However, that people differ in their metalinguistic judgements was obvious from the start. After all that is why there are two different families of theories of reference. (Martí 2009, 2014)[8]

2.2 The result of the experiment can't inform us about variation in the relevant intuitions, because the question at the end of the vignette contains an ambiguity, which can account for the variation in answers. "Who John is talking about" can either mean the speaker meaning (who does John intend to talk about?) or the semantic referent (what does "Gödel" refer to in John's utterance?). (Deutsch 2009; Sytsma and Livengood 2011)[9]

The dissimilitude of these objections already indicates that there is no consensus within philosophy of language on whether semantic intuitions (should) play a significant role, and if so, which and whose intuitions are of relevance. Objection 1.1 denies the relevance of intuitions completely. 1.3 denies the relevance of intuitions of laymen. 2.1 seems to allow intuitions some role but objects that the experiment tested the wrong ones.

Which out of these objections are valid and which should be rejected? As explained in the introduction, methodological questions depend on one's aims. Thus, in order to see what role intuitions should play in our methodology, we should first get clear on what purpose theories of reference are supposed to have. This question will be discussed in the next section. Only after that will Section 4 turn to the question of which and whose intuitions (if any) should be of relevance in choosing between different theories of reference.

3. Meta-Internalism vs. Meta-Externalism

The two families of theories of reference that were introduced in section 2 (viz. descriptivist theories and causal-historical theories) are often also classified as "internalist" or "externalist" theories of reference, respectively. A descriptive theory is internalist in this sense because what a proper name in a certain speaker's usage refers to depends partly on her internal states (because it depends on which description, or bundle of descriptions, the speaker associates with the name). A causal-historical theory, in contrast, is externalist in this sense because it depends solely on the existence of a causal chain of name-borrowing between language users that determines what a name refers to, regardless of the speaker's awareness of that chain.

In order to clarify which kind of evidence matters for deciding between these two kinds of theories, we should first inquire into which kinds of facts it could depend on, that is, whether an internalist or an externalist theory is true. When we know this, we can inquire into what is the best way (or at least a good way) to find out about these facts.

As argued in (Cohnitz and Haukioja 2013), we can make progress on the first question by also drawing a distinction between internalism and externalism at a meta-level. We should then distinguish between Meta-Internalism and Meta-Externalism:

> *Meta-Internalism*: How a linguistic expression E in an utterance U by a speaker S refers and which theory of reference is true of E is determined by individual psychological states of S at the time of U.
>
> *Meta-Externalism*: How a linguistic expression E in an utterance U by a speaker S refers and which theory of reference is true of E is independent of the individual psychological states of S at the time of U.

Many theories of reference are meta-internalist: that the referent of, say, a proper name is a matter of the history of its usage within in the linguistic community is usually (though often not explicitly) considered to be so because the speaker in using the name had the (tacit) intention to engage in that tradition of reference borrowing. Thus, on this account, reference is determined via external factors (the history and tradition of using the name) but that it is these external factors is determined by the (tacit) intentions of the speaker.

On a meta-externalist view, it could be that, even if the speaker had no such intentions, nonetheless the reference of proper names in her usage could be determined by the way the name is used in the tradition of her linguistic community, because it is external factors (independent of her intentions, or dispositions or other psychological states) that determine which first-order theory is true of linguistic expressions in her usage.

Hence in this case, semantic facts are independent of the internal facts of speakers, including their dispositions to produce and interpret expressions in certain ways and, consequently, independent of the intuitions speakers might have concerning the semantic properties of the expressions they use. Thus, if we assume that a speaker's intuitive judgements in response to Gödel/Schmidt-cases reveal her dispositions concerning how she'd use or interpret a certain type of expression (an assumption we will scrutinize below), it would be possible that her judgements just don't track the semantic facts (although they'd track her linguistic dispositions). Hence, the meta-externalist response to the found variation between Ws and EAs would be that at least one of the groups is getting the semantic facts wrong. On that account, semantic facts are as independent of intuitions as physical facts are. This view seems[10] to have obvious methodological consequences. Just as we shouldn't have much trust in our intuitions when it comes to matters of physics, we shouldn't have much trust in our intuitions when it comes to semantics.

One philosopher who seems to hold such a meta-externalist view is Michael Devitt. Devitt argues that the subject matter of semantics and linguistics is 'linguistic reality', the study of physical expression tokens and their semantic and grammatical properties (Devitt 2006). This reality should be distinguished from psychological reality. How the expression tokens get their properties seems to be a secondary question for him, and he allows that expression tokens get their semantic properties independently of the psychological states of the speaker at the time of the relevant utterance.

Other, perhaps clearer, examples of meta-externalist positions are theories of reference that assume that objective structures (for example, natural properties) can function as reference magnets for a speaker's expressions and thereby override the intentions of the speaker that pertain to the speaker's use of a term in her repertoire (including the absence of any intentions to use it for whatever the objective structures happen to be).[11]

Although meta-externalism seems to be endorsed by some radical first-order externalists, it is a rather implausible position. As argued in Cohnitz and Haukioja (2013), it leads to the possibility of 'semantic secrets'[12]; expressions might (systematically) 'refer' to objects that are irrelevant to the contents transmitted in communication. But if that can happen, then a theory of reference wouldn't systematically contribute to an explanation of successful linguistic communication. For the latter, it seems that reference must somehow be tied to certain psychological states of speakers, in particular to their dispositions to produce and interpret expressions in certain ways and to revise their usage systematically in light of new information.

Just consider a population like the EAs in the experiment reported by Machery et al. and assume that their judgements (formed in response to the probe) reveal their dispositions to use and interpret proper names and that these dispositions don't align with the semantic facts.[13] In that case, whenever the causal-historical-theory and the descriptivist theory make distinct predictions about the referent of a proper name in an utterance, the content communicated by the utterance (i.e. the content intended to be communicated and the content received) would be systematically different from the content assigned to the utterance by the true theory of reference (in this case, the causal-historical theory). It seems that the semantic properties ascribed by the theory of reference to the utterance tokens would be irrelevant to any explanation of how the content in question got communicated via the utterance. But what would be the point of such a theory?

On the basis of these considerations, the following definition of reference is suggested in (Cohnitz and Haukioja 2013):

> Reference: A token expression E refers in language L to object O iff (i) E is standing in the R-relation to O and (ii) competent speakers of L are disposed to interpret objects (of the type of O) to be the referents of expressions (of the type of E), if they believe these are connected by the R-relation.

If we follow this suggestion (assuming that the causal-historical theory is the correct first-order theory), then a certain token of the name ‘Angela Merkel’ refers in English to the person Angela Merkel because Angela Merkel is at the other end of a causal-historical chain that leads up to the usage of that token, and speakers of English are disposed to interpret persons as the referents of proper names, if they believe that a name-token and a person stand in such a causal-historical relation.

Of course, speakers of English (unless they are particularly nerdy linguists or philosophers of language) do not form any beliefs about the exact causal-historical relations in which people stand to linguistic expression-tokens. What is required is that their usage, including their dispositions to correct their usage, is sensitive to information that happens to be about the causal-historical chain in question.

Thus, when confronted with a Gödel/Schmidt-case, we are disposed to interpret John’s usage of the name “Gödel” as referring to the man who stole the manuscript when we learn the facts about the causal historical chain, rather than to interpret him as referring to Schmidt, the man who in fact proved the incompleteness of arithmetic. We are sensitive to information about the causal-historical chain in our interpretation dispositions rather than sensitive to information about the beliefs a person happens to have concerning the (purported) bearer of a name. We don’t need to think at any point about any of this in terms of the R-relation, and so on.

Meta-Internalism is a theory about the subject matter of theories of reference and thus about the facts that determine which theory of reference is true. According to Meta-Internalism, these are facts about certain dispositions of competent speakers. Therefore, if we want to know which theory of reference is true, we should try to get information about those facts. That we should endorse Meta-Internalism rather than Meta-Externalism is grounded in our epistemological aim; we are interested in a theory that can make a systematic contribution to an explanation of linguistic communication. A Meta-Externalist theory doesn’t seem to be fit for the job.

However, some authors who would agree that this is the role of theories of reference and that these are the facts that determine which theory of reference is true still disagree with the idea that in testing intuitive responses to hypothetical cases in philosophy of language we are studying the relevant intuitions. As explained in Section 2, Genoveva Martí believes that thought experiments in philosophy of language elicit meta-linguistic intuitions. But for all that is said so far, such intuitions shouldn't be very relevant for determining the truth about reference. In the next section we will consider whether Martí is right: what are the relevant intuitions in philosophy of language, and which intuitions are tested with Gödel/Schmidt-type thought experiments?

4. Semantic Intuitions

Let's consider a concrete example: the Gödel/Schmidt case described above and a hypothetical utterance (U) of John:

(U) Gödel was a brilliant mathematician.

What does "Gödel" refer to in this utterance? That's the kind of question that is typically raised in thought experiments in philosophy of language. A typical answer would be

(A) In John's utterance (U), "Gödel" refers to the man who stole the manuscript and claimed credit for the work.

In this case, what is the 'intuition'? An intuitive interpretation of the hypothetical utterance by John that we arrived at thanks to our linguistic competence? Or is it instead a spontaneous meta-linguistic judgement concerning the expression 'Gödel' that we arrived at on the basis of our everyday experience with the usage of proper names by competent speakers of English? Michael Devitt and Genoveva Martí claim that it is the latter (and thus that it is of little value for semantic theorizing). For example, in response to the experimental work by Machery et al., Genoveva Martí offers the following response:

> I think if we focus on the type of data that [the probes used by Machery et al.] are collecting, the 'semantic intuitions' that they elicit, we can see that the responses are not the kind of data that constitutes the input, the raw data that the semanticist relies on in order to start theorizing. Participants in the probes are told a fictional story about a community of speakers . . . Participants are then asked to hand down a judgement as to what the referent of a use of a

> name by a hypothetical speaker member of the fictional community is. So, the participants are asked to tell us how they think the hypothetical speaker in question, and the rest of his community, uses names. Is that the evidence that we should rely on to construct a semantic theory? I think the answer is no. (Martí 2014, p. 22)

According to Martí, these intuitive judgements that are elicited by thought experiments like the Gödel case only inform us about how people think that they use language, but not about how they actually use language, and it is only the latter that matters for the semanticist. Therefore, the only real evidence that matters in linguistics and philosophy of language is data about use, collected by observing the theoretician's own linguistic behaviour and that of the linguistic community around her. A judgement like (A) at best informs us about the theoretical preferences of the test subjects:

> The Gödel story invites a reflection on use, it does not collect data on use; it is, hence, a theoretical tool, and Kripke uses it as such. And the responses of subjects to the Gödel story will, at best, tell us what theory they are disposed to find more natural as an explanation of how the hypothetical speaker, or they themselves, use language. But what theory people are more disposed to accept is not the input of the theory itself. (Martí 2014, p. 23)

But how do we know that a judgement like (A) is a theoretical judgement rather than evidence of a relevant bit of language use? How do we know that a response like (A) informs us at best about which theory of reference a speaker is disposed to favour rather than about how she is disposed to interpret an utterance like (U)?

Martí seems to think that this is obvious from the way the probe is phrased (at least she doesn't seem to offer any other evidence). Doesn't the question who John is talking about or the question to whom "Gödel" refers in (U) require reflections on use because, after all, these are questions about how another person uses a name?

This would be right only if interpretation wasn't part of use. But it obviously is. Linguistic competence with proper names isn't only a matter of producing sentences with proper names in the right way. It's also a matter of interpreting the use of proper names in the sentences of others. Presumably production and interpretation are just two sides of the same coin. Therefore the fact that the question asked in the probe concerned the utterance of a third person does not itself establish that the probes were collecting anything other than the relevant 'raw data'.

However, there seem to be two other considerations that speak in favour of the view held by Martí and Devitt. First of all, (A) is a meta-linguistic sentence, since it obviously speaks about language. The sentence contains the word "Gödel" in quotation marks, thus the sentence is (at least in part) about the word "Gödel". Furthermore, the sentence contains an expression from semantics ("refers to"). Therefore it seems plausible to think this sentence, inasmuch as it expresses a judgement, expresses a meta-linguistic judgement.

However, according to Devitt (2006), meta-linguistic judgements are independent of our linguistic competence. They are ordinary judgements that we arrive at on the basis of experience and background-knowledge (in this case, our experience with the way in which proper names are typically used by speakers of English and background knowledge we might perhaps have about linguistics and theories of reference). The only difference between these judgements and others is that the former are made with much greater spontaneity than the latter.

In (Cohnitz and Haukioja forthcoming) this view is discussed in some detail. There it is argued that we need to distinguish, first of all, between the results of dispositional competences and their reports. When we interpret an utterance in a context of utterance, we interpret the utterance and its component expressions intuitively. We are basically doing the same when interpreting (U) in the hypothetical context described above. What enables us to arrive at such interpretations, is our linguistic competence in English. Under certain conditions, we might also be able to report the results of such interpretations. What is required is that (a) the result is available to our consciousness, and (b) we master the concepts and expressions required for making such a report. If both, (a) and (b) are in place, we are able to report our intuitive interpretation of (U) and, in particular, our interpretation of the expression "Gödel" in (U), using the sentence (A).

In the case of semantic interpretations, i.e. our understanding of utterances and their parts, it seems plausible to assume that the results of these interpretations are available to consciousness. After all, these results, that is, our interpretations of what our interlocutors say in their utterances, enter into our inferences about what our interlocutors think and plan.

It also seems plausible that we master the relevant concepts and terms required to report our interpretations. Conversation about what other people have said and about whom they have said things forms a huge part of our everyday communication. Thus requirements (a) and (b) seem to be met for semantic interpretations.

This is already an important result. What seemed to speak in favour of the view championed by Devitt and Martí was that (A) was a metalinguistic sentence because (A) was about the word "Gödel" and (A) included semantic vocabulary. As we have seen now, whether (A) expresses a metalinguistic judgement in their sense is not so much a question of what the sentence is about but rather a matter of what cognitive process leads to the judgement expressed. As we have seen now, (A) could simply be a report of our intuitive interpretation of (U).

(A) could also be the result of a different cognitive process; namely the one described by Devitt and Martí. In this case we'd have a dispositional competence to make generalizations about the reference of expressions based on our observations of the usage of proper names in English. This acquired competence would allow us to make judgements about (U) that we'd report with sentences like (A).

In both cases, in reporting the intuitive interpretation or the metalinguistc judgement, it is possible to make mistakes. In the case of metalinguistic judgements, even if the judgement was reported without mistake, the methodological value and relevance of such judgements for the philosophy of language and linguistics depends on how good we are at making generalizations about linguistic usage based on our everyday experience. One can probably follow Devitt and Martí in thinking that this value is not very high.

For reports of intuitive interpretations, things look different. If the report is accurate and there are no other reasons to assume that our interpretation was not produced by our linguistic competence, then (A) is not only relevant for philosophy of language, but it reports the kind of facts that constitute the very subject matter of theories of reference.

But besides the observation that (A) is a meta-linguistic sentence, Devitt (2011) has a second argument for the claim that (A) expresses a meta-linguistic judgement that doesn't report the output of linguistic competence but instead something independent of it. Devitt cites empirical evidence from developmental psychology that we acquire the relevant meta-linguistic concepts relatively late in our cognitive development and that we are able to make judgements like (A) only a considerable time after we have already acquired (otherwise) complete linguistic competence.

But here Devitt overlooks the fact that this result is entirely compatible with the Meta-Internalist view. Of course, it might well be the case that it is only relatively late in our linguistic and cognitive development that we are able to report our semantic interpretations or parts of them with sentences like (A). But this doesn't provide us with any reason to doubt the view that these reports are reports of the outputs of our linguistic competence rather than reports of an independent observation of our linguistic environment. For example, the ability to visually recognize certain objects in our environment precedes our ability to report what we have recognized. The same could easily hold for our ability to report outputs of our linguistic competence.

Thus, there remains nothing that would speak in favour of Devitt's and Martí's view that the intuitive judgements that serve as evidence in philosophy of language are spontaneous judgements based on linguistic experience. Instead, these judgements are plausibly reports of the interpretation-outputs of linguistic competence.

However, the fact that Devitt and Martí misunderstand these judgements and their proper content is still cause for concern. If there are two distinct processes that can both lead to utterances of (A), and if philosophy of language should really only be interested in one of them, then there is at least the danger that the data (judgements of the form of [A]) are ambiguous. How are we supposed to know that test-subjects would judge that (A) is true on the basis of their linguistic competence, if even experts like Devitt and Martí misunderstand the task?

Moreover, it was speculated above that the reliability with which sentences like (A) report interpretations of sentences like (U) depends in part on how entrenched the relevant metalinguistic concepts and terms are in the idiolect of the test-subject. Perhaps philosophers of language are better at reporting such interpretations than the ordinary folk that Machery et al. tested.

As was said in the beginning, there is also the further problem that (A) can be ambiguous in a second way when it comes to the evaluation of the Gödel/Schmidt-case. It was argued that there might be an ambiguity between the semantic referent of "Gödel" and the speaker referent of John's utterance. In the next section we will look at ways to get around these two problems by refining the standard methodology of philosophy of language.

5. How philosophy could learn from psycholinguistics

Let us first consider possible solutions to the problem of how we could figure out which cognitive process has in fact lead to a certain judgement. *Prima facie* one might think that we are faced with a methodological dilemma. We want to know how competent speakers understand a hypothetical utterance. But we can't just look into their heads, so we need to ask them how they understood it. However, as we have just seen, those questions and their answers seem to be meta-linguistic in the sense that these questions or judgements are about linguistic expressions and in addition make use of semantic vocabulary. Doesn't that lead unavoidably to our methodological problem; namely, that we don't know whether the test person informed us in her answer about her semantic interpretation (which is what we are interested in), or rather about what her lay-theory of proper name reference predicts for this case (which we don't care much about)?

Luckily, this isn't a real dilemma. First of all, we could investigate the interpretations we are interested in without asking the test subjects. For example, we could investigate them indirectly by observing the later behaviour of the test subjects, in particular in situations in which relevant differences in interpretation would lead to observable differences in behaviour. From that, one might be able to infer, under appropriate circumstances, the information (if any) that the test subject extracted from an utterance – either about the world or the speaker.

Of course, it would be even better if we didn't need to use such an indirect methodology (i.e. one that might introduce new ambiguities). In fact, it is indeed possible, at least when it comes to reference, to study the interpretation of linguistic expressions directly. To see how this could be done, we need to look at psycholinguistics and the methods used therein to investigate the resolution of referring expressions.

When psycholinguists investigate, for example, the resolution of anaphoric reference, they often use eye-trackers, that is, instruments that allow experimenters to track the eye-movement of a test subject. Initially this method was used to see how eyes move in reading-comprehension tasks. For our purposes, the more interesting work in psycholinguistics is that performed in studies described by Karabanov et al. (2007). In these experiments, test-subjects are confronted with two stimuli: the auditory stimulus is a spoken text (which contains in this case anaphoric pronouns) and a visual stimulus, in our case a picture of a pseudo-natural market-day situation built up with Playmobil^{TM} toy characters.

In the experiment described by Karabanov et al. (2007), the linguistic stimulus was the following German sentence:

> Heute ist Markt im Dorf. Die Marktfrau streitet mit dem Arbeiter. Sie sagt jetzt gerade, daß er kein' Ärger machen und das neue Fahrrad zurückgeben soll, das er sich geliehen hat. [It's market day in the village. The market woman is quibbling with the worker. She's just saying that he should not make any trouble and should give the new bike back that he borrowed.] (Karabanov et al 2007, p. 211)

There are several theories about the resolution of pronouns that differ with respect to the postulated cognitive processes involved in their interpretation. For example, the theory of Morton Gernsbacher (1989) assumes the interpretation of anaphoric pronouns to be a two-step process. First the antecedent of the pronoun is identified and then, in a second step, the connection to the referent of the antecedent is established. The theory by Lorraine Tyler and William Marslen-Wilson (1982), however, holds that pronouns are immediately interpreted referentially, just like full lexical NPs or proper names. This difference in postulated cognitive processes should (or, in any case, might) lead to empirically testable differences: on the view suggested by Gernsbacher the interpretation of an anaphoric pronoun would seem to require more time than the interpretation of a full lexical NP, while on Tyler and Marslen-Wilson's theory, the resolution of an anaphoric pronoun should take as much time as that of a full lexical NP.

In the eye-tracking experiment by Karabanov et al., it is assumed that eye-movement (or more specifically, the probability by which the eyes of the test-subject will fixate on a certain point in the visual scene) is directly correlated with the interpretation of a heard linguistic item. That makes it possible to compare the times it takes for the fixation probabilities for a certain item in the visual field to increase and relate these times to the type of expression interpreted. Does it take longer for the fixation probability for the referent of an anaphoric pronoun to increase than it does for the referent of a full lexical NP? Karabanov et al. found no time difference (which they took to speak in favour of the theory by Tyler and Marslen-Wilson).

This empirical result itself is (for our purposes) not so interesting. What is interesting about this experiment is the fact that we can use eye-trackers to measure the interpretation of referential expressions without having possible problematic metalinguistic considerations taint the data. Our eye-movements are involuntary and automatic. The causally relevant cognitive process is that of interpreting the sentence. Other potential influences (like the relative salience of an object in the visual field) can be controlled for in the experiment.

That way we possess at least one instrument for measuring the interpretation of referential expressions by competent speakers that is immune to Devitt's methodological worries. Philosophers of language could make use of this tool in two ways. First, one could test theories of reference directly on test subjects, given that these theories make different predictions about the referent of an expression (under certain contextual circumstances). This would presuppose that different theories of reference always make different predictions that can be turned into measurable tasks. This, presumably, is not always the case. However, we could use this methodology also for the calibration of our ordinary armchair methodology. Devitt's worries aren't automatically relevant when cognitive processes other than linguistic interpretation lead to the metalinguistic judgements in question. They are only relevant if these other cognitive processes would lead to different judgements. If it could be shown that our judgements about the reference of a term in a hypothetical utterance and our interpretations as measured by the eye-tracker coincide with high reliability, then Devitt's worries should be simply irrelevant.

The method described could also be helpful in a second sense. We said above (when discussing objection 2.2 against the experiment by Machery et al.) that the responses to the Gödel/Schmidt-probe suffered from a second ambiguity. Consider again the relevant test questions:

> When John uses the name "Gödel", is he talking about: (A) the person who really discovered the incompleteness of arithmetic? or (B) the person who got hold of the manuscript and claimed credit for the work?

It was ambiguous whether the question is about who John intends to refer to with his usage of 'Gödel', or instead about who John in fact and objectively refers to with that term.

Indeed, in a later experiment, Justin Sytsma and Jonathan Livengood showed that intra-cultural[14] variation disappears when the final question is changed to the following:

> Clarified Narrator's Perspective: Having read the above story and accepting that it is true, when John uses the name "Gödel", would you take him to actually be talking about: (A) the person who (unbeknownst to John) really discovered the incompleteness of arithmetic? Or, (B) the person who is widely believed to have discovered the incompleteness of arithmetic, but actually got hold of the manuscript and claimed credit for the work? (Sytsma and Livengood 2011, p. 324)

The amount of B-answers increased from 39.4 per cent (for the original question that was used in the study by Machery et al. 2004) to over 73 per cent for the 'clarified narrator's perspective'. This suggests that the variation detected by Machery et al. (at least the intra-cultural one) was due to the ambiguity in perspective in the final question.

However, even the clarified narrator's perspective contains a residual ambiguity. If I know that John believes that Gödel proved the incompleteness of arithmetic and know that John is good at mathematics and knows nothing else about Gödel, then I might think that John intends to (and does) talk about the guy who proved the incompleteness of arithmetic. In order to eliminate this ambiguity one could rephrase the question, using precise semantic vocabulary, thereby again increasing the probability for tainting the data with metalinguistic considerations.

Perhaps the eye-tracking methodology described above could help us also here.[15] As Keysar, Lin and Barr (2003) report, the pragmatic interpretation of an utterance, which takes into account the perspective and knowledge of the speaker/interlocutor, is delayed in comparison to the literal interpretation of an utterance. In one of their experiments, a test situation was arranged such that the test subject (the participant) and a second person (the confederate) were sitting at opposite sides of a table. The test subject could see more items on the table than the other subject and was aware that she could see those other things and that the other person didn't even know these other things were on the table. Nevertheless, when tracking the eye-movements of the participant, it turned out that the fixation probability for those hidden objects increased first when they were better candidates for being the semantic referents of the expressions used by the confederate, while the fixation probabilities for the speaker referents was, by comparison, delayed.

For example, in one instance of such an experiment the participant was asked to secretly hide a roll of tape in a paper bag and to store it at a location visible to her but invisible to the confederate. However, there was a cassette visible to both the confederate and the participant. While monitoring the eye-movements of the participant, Keysar, Lin and Barr found that in many cases in which the confederate instructed the participant to 'move the tape', the eye-movements of the participant revealed that she had first interpreted the referent of "the tape" to be the best semantic candidate (the hidden roll of tape in the box) rather than the only possible speaker referent (the mutually visible cassette). The interpretation of "the tape" as having the intended (speaker) referent occurred with a time delay.

These results suggest that eye-tracking can be used to discriminate between speaker reference and semantic reference, again without involving any problematic meta-linguistic questions or judgements. This interpretation also seems to be consistent with more recent experiments reported in Barr (2008), which show that although listeners in a conversation expect speakers to refer to objects in the common ground (accessible or visible to both), they are unable to reduce interference from 'privileged competitors', that is from better semantic referents that are not in the common ground (Barr 2008). Again, this suggests that eye movement is primarily responsive to semantic interpretation (and only with a delay, is it responsive to the pragmatic integration of knowledge of common ground).[16] To be clear, this interpretation of the experimental results, although consistent with them, hasn't been tested yet (as far as I know). But if this hypothesis holds up, then eye tracking would provide us with a further method to study the relevant 'raw data' (to use Martí's expression) at the level of interpretation in a more reliable way than the usual method wherein test subjects are asked to report their interpretations of hypothetical utterances.

6. Conclusions

We have seen that, if certain assumptions are made about the explanatory aims of philosophy of language and about the nature of reference as it should feature in those explanatory aims, we can investigate philosophical methodology and arrive at recommendations for an improved methodology. We then know what we need to examine, how we could check the reliability of our methods, and consider improvements.

However, the assumptions we started from, (viz. the idea that theories of reference should systematically contribute to explanations of successful linguistic communication) and the consequences these had for Meta-Internalism, are specific to this particular sub-area of philosophy. Perhaps they are even specific to this area of philosophy of language. There might well be another inquiry into 'reference', which doesn't start with a primary interest in communication but perhaps with a primary interest in representation, or perhaps information, and such a project might encumber different methodological commitments.[17]

However that may be, it should be clear that one can't just generalize the methodological insights here described to all other areas of philosophy without first proving that the same assumptions we made about theories of reference also hold for these other areas. As I tried to explain in the introduction, *prima facie*, there is currently no reason to think that they do.

Notes

[*] I would like to thank Alex Davies, Jussi Haukioja and Edouard Machery for helpful comments on an earlier version of this text. Research for this chapter was supported by the following research grants: SFLFI11085E, ETF9083, SF0180110S08 and IUT20–5.

Some of the ideas presented here are discussed in somewhat more detail in Cohnitz 2014, and a lot of it is based on work that was carried out with Jussi Haukioja, as with our (2013) and (forthcoming). However, the present chapter contains a more careful discussion of Genoveva Martí's recent criticism of experimental philosophy (Martí 2014) and further suggestions pertaining to how empirical methods could be put to use in philosophy of language. Thus readers familiar with the arguments in this earlier work can fast-forward to Sections 4 and 5.

1 Although the notion itself seems to be a rather recent invention, and the discussion has developed systematically only in the past decade, PhilPapers, the most comprehensive index and bibliography of philosophy, already has a top-level category "Metaphilosophy".

2 Of course, there have always been metaphilosophical contributions and discussions. Another unfortunate aspect of the discussion in the past decade is that it appears not to be informed by what has been written on the subject before that decade.

3 Even if we restrict – as I do throughout the chapter – philosophy to analytic philosophy.

4 In this chapter, the focus will be on the discussion of the methodological value of intuitions that resulted from this chapter. For an overview of the development of experimental philosophy of language in general, see Genone (2012) and Hansen (forthcoming).

5 Cf. Stich (in preparation).

6 Cf. Machery et al. (2013) for a response.

7 Cf. Machery (2012) for a response.

8 Cf. Machery et al. (2009) for a response.

9 Cf. Machery et al. (forthcoming) for a response.

10 Of course, the Meta-Externalist view might also be compatible with the reliability of our intuitions, if we assume that we are sufficiently attuned to the linguistic facts. However, in light of the empirical results, the Meta-Externalist would at least have reason to doubt that we (or, in any case, lay speakers) are sufficiently attuned. Thanks to Edouard Machery for pressing me on this.

11 See Schwarz (2013) for a discussion of such views.

12 The term was introduced in Schwarz (2013).

13 On that account, what we should say about the cultural variation found in the study by Machery et al. (2004) is that the Ws get it largely right how proper names work, while the EAs get it largely wrong.

14 However, the inter-cultural variation could be replicated. Cf. Sytsma et al. (forthcoming).

15 I owe the idea that eye-tracking could be used for distinguishing speaker and semantic referent to Manuel Garcia-Carpintero.

16 A curious finding by Wu and Keysar (2007) is that, apparently, listeners from a Chinese background (when listening to Chinese) show less interference from privileged referents than listeners from an English-speaking American background. This is curious, because it would be consistent with the original findings in Machery et al. 2004 and could perhaps (partly) explain the found cultural variation.

17 I don't think that that's in fact the case for reference. But it seems to me that the fact that Devitt's thinking about linguistic meaning departs from considerations about how we can extract information about the world from linguistic items, rather than from considerations about how we manage to communicate with language, explains to some extent why he believes that we can study linguistic reality in ignorance of the psychological basis of reference (cf. Devitt and Sterelny 1999).

References

Barr, D. J. (2008), 'Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground'. *Cognition,* 109, 18–40.

Bealer, G. (1996), 'A priori knowledge and the scope of philosophy'. *Philosophical Studies,* 81, 121–142.

Cappelen, H. (2012), *Philosophy without Intuitions*. Oxford: Oxford University Press.

Cappelen, H. and Winblad, D. G. (1999) '"Reference" externalized and the role of intuitions in semantic theory'. *American Philosophical Quarterly,* 36, 337–350.

Carnap, R. (1950), *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.

Deutsch, M. (2009), 'Experimental philosophy and the theory of reference'. *Mind and Language,* 24, 445–466.

Cohnitz, D. (2014), 'Experimentelle sprachphilosophie', in T. Grundmann, J. Horvath and J. Kipper (eds), *Die Experimentelle Philosophie in der Diskussion*. Berlin: Suhrkamp, pp. 235–258.

Cohnitz, D. and Haukioja, J. (2013), 'Meta-externalism vs meta-internalism in the study of reference'. *Australasian Journal of Philosophy,* 91, 475–500.

Cohnitz, D. and Haukioja, J. (forthcoming), 'Intuitions in philosophical semantics'. *Erkenntnis*.

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

Devitt, M. (2006), *Ignorance of Language*. Oxford: Oxford University Press.

Devitt, M. (2011), 'Whither experimental semantics', *Theoria,* 72, 5–36.

Devitt, M. (2011a), 'Experimental semantics', *Philosophy and Phenomenological Research,* LXXXII, 418–435.

Devitt, M. and Sterelny, K. (1999), *Language and Reality: An Introduction to the Philosophy of Language* (2nd edn). Oxford: Blackwell Publishers.

Genone, J. (2012), 'Theories of reference and experimental philosophy'. *Philosophy Compass*, 7, 152–163.

Gernsbacher, M. A. (1989), 'Mechanisms that improve referential access'. *Cognition,* 32, 99–156.

Hansen, N. (forthcoming), 'Experimental philosophy of language'. *Oxford Handbooks Online*.

Karabanov, A. et al. (2007), 'Eye tracking as a tool to investigate the comprehension of referential expressions', in S. Featherston and W. Sternefeld (eds), *Roots: Linguistics in Search of its Evidential Base*. Berlin: DeGruyter, pp. 207–226.

Keysar, B. et al. (2003), 'Limits on theory of mind use in adults'. *Cognition,* 89, 25–41.

Kripke, S. (1980), *Naming and Necessity*. Boston, MA: Harvard University Press.

Machery, E. (2012), 'Expertise and intuitions about reference'. *Theoria*, 73, 37–54.

Machery, E. (2014), 'What is the significance of the demographic variation in semantic intuitions?', in E. O'Neill and E. Machery (eds), *Current Controversies in Experimental Philosophy*. New York and London: Routledge, pp. 3–16.

Machery, E. et al. (2004), 'Semantics, cross-cultural style'. *Cognition,* 92, B1–B12.

Machery, E. et al. (2009), 'Linguistic and metalinguistic intuitions in the philosophy of language'. *Analysis*, 69, 689–694.

Machery, E. et al. (2013), 'If intuitions vary, then so what?'. *Philosophy and Phenomenological Research*, 86, 618–635.

Machery, E., et al. (forthcoming), 'Speaker's reference and cross-cultural semantics', in A. Bianchi (ed.), *On Reference*. Oxford: Oxford University Press.

Mallon, R. et al. (2009), 'Against arguments from reference'. *Philosophy and Phenomenological Research,* LXXIX, 332–356.

Marslen-Wilson, W. and Tyler. L. K. (1987), 'Against modularity', in J. L. Garfield (ed.), *Modularity in Knowledge Representation and Natural Language Understanding*, Boston, MA: MIT Press, pp. 37–62.

Martí, G. (2014), 'Reference and experimental semantics', in E. O'Neill and E. Machery (eds), *Current Controversies in Experimental Philosophy*. New York and London: Routledge, pp. 17–26.

Martí, G. (2009), 'Against semantic multi-culturalism'. *Analysis,* 69, 42–48.

Nisbett, R. (2003), *The Geography of Thought: How Asians and Westerners Think Differently . . . and Why*. New York: Free Press.

Nisbett, R. et al. (2001), 'Culture and systems of thought: Holistic vs. analytic cognition'. *Psychological Review,* 108, 291–310.

Schwarz, W. (2013), 'Against magnetism'. *Australasian Journal of Philosophy,* 92, 17–36.

Stich, S. (in preparation), 'Experimental philosophy and the bankruptcy of the great tradition'.

Sytsma, J. and Livengood, J. (2011), 'A new perspective concerning experiments on semantic intuitions'. *Australasian Journal of Philosophy,* 89, 315–332.

Sytsma, J. M. et al. (forthcoming), 'Gödel in the land of the rising sun'. *Review of Philosophy and Psychology*.

Weinberg, J. (2007), 'How to challenge intuitions empirically without risking skepticism'. *Midwest Studies in Philosophy,* XXXI, 318–343.

Williamson, T. (2007), *The Philosophy of Philosophy*. Oxford: Oxford University Press.

Wu, S. and Keysar, B. (2007), 'The effect of culture on perspective taking'. *Psychological Science,* 18, 600–606.