

# *Referential intuitions are still problematic*

MASSIMILIANO VIGNOLO AND FILIPPO DOMANESCHI

## *1. Introduction*

The dialectic between the parties engaged in the debate about the evidential role of experiments in theory of reference is very delicate. We show that the difficulty lies in large part in the fact that linguistic performances of different kinds might be elicited in test subjects. These linguistic performances include making a judgment about what a speaker refers to, judging whether a speaker speaks truly/falsefully, understanding what a speaker refers to and using a word to refer to something. It is difficult to isolate the evidential weight of each of these linguistic performances taken separately from the others. Moreover, empirical surveys that test the correlation of different linguistic performances seem to weaken the evidential weight of the data collected in experiments that test only linguistic performances of one and the same kind.

We discuss this point focusing on the debate generated by the reply of Machery, Olivola and de Blanc (2009) – henceforth ‘MOD’ – to a criticism by Martí (2009) to the seminal paper ‘Semantics, cross-cultural style’ by Machery et al. (2004) – henceforth ‘MMNS’. In particular, we argue that so far there is no evidence that people’s referential intuitions (judgments about what a speaker refers to in actual and possible situations) are reliable guides to the correct theory of reference for proper names.

The main idea underpinning MMNS’s paper is that the methodology for assessing theories of reference consists in the method of cases, which states that the correct theory of reference for words of type T is the one that is best supported by people’s intuitions about the reference of words of type T in actual and possible situations. If a theory of reference predicts that a word E of type T refers to an entity O, the theory is confirmed if people have the intuition that E refers to O and disconfirmed if they have the intuition that E does not refer to O. MMNS presented two vignettes to Westerners (undergraduates at the University of Rutgers, USA) and Easterners (undergraduates at the University of Hong Kong) modelled on Kripke’s Gödel/Schmidt case. MMNS found out that referential intuitions vary across and within cultures. About 60% of Westerners expressed referential intuitions that are in line with the causal/historical theory, whereas about 70% of Easterners expressed referential intuitions that are in line with classical descriptivism. MMNS affirm that the demographic variation of referential intuitions casts a sceptical shadow on theories of reference, especially on the mainstream causal/historical one.

MMNS's paper has provoked a large and intense debate among philosophers about the evidential role of referential intuitions in theory of reference. Martí (2009) put forward one of the most radical objections to MMNS's experiment. Martí argued that MMNS's experiment does not elicit the kind of data that are relevant for selecting the correct theory of reference. MMNS's experiment prompts participants to think about what a hypothetical speaker refers to with a proper name and to reflect on how reference is determined. Martí said that the way test subjects think reference is determined is not relevant for confirming a theory of reference. Asking test subjects to tell whether the character of the vignette uses the name 'Gödel' to talk about the logician who really demonstrated the incompleteness theorem or the thief of the demonstration may at most reveal their inclinations towards ways of theorizing about reference, but it reveals no significant aspect of the real phenomenon in which philosophers are interested, namely, how reference is determined. Philosophers of language are not interested in what people think about how proper names refer. Rather, they want to know how proper names refer. Martí concluded that referential intuitions are evidentially irrelevant for individuating the correct theory of reference and, as a consequence, MMNS's experiment has no bearing on Kripke's refutation of classical descriptivism.

## 2. *Testing truth-value judgments*

In order to respond to Martí's objection, MOD conducted an experiment aimed at showing that people's referential intuitions are congruent with the way they use proper names. They developed two vignettes, one for testing how people use proper names and the other for testing their referential intuitions. The vignette developed for testing how people use proper names is the following:

Ivy is a high school student in Hong Kong. In her astronomy class, she was taught that Tsu Ch'ung Chih was the man who first determined the precise time of the summer and winter solstices. But, like all her classmates, this is the only thing she has heard about Tsu Ch'ung Chih. Now suppose that Tsu Ch'ung Chih did not really make this discovery. He stole it from an astronomer who died soon after making the discovery. But the theft remained entirely undetected and Tsu Ch'ung Chih became famous for the discovery of the precise times of the solstices. Everybody is like Ivy in this respect; the claim that Tsu Ch'ung Chih determined the solstice times is the only thing people have heard about him. Having read the above story and accepting that it is true, when Ivy says, 'Tsu Ch'ung Chih was a great astronomer', do you think that her claim is: (A) true or (B) false?

The vignette developed for testing referential intuitions is identical to the previous one apart from the question task, which is identical to the question task of MMNS's experiment:

Having read the above story and accepting that it is true, when Ivy uses the name 'Tsu Ch'ung Chih', who do you think she is actually talking about:

- (A) the person who (unbeknownst to Ivy) really determined the solstice times? or
- (B) the person who is widely believed to have discovered the solstice times, but actually stole this discovery and claimed credit for it?

The first vignette is supposed to test how people use proper names by means of a truth-value judgment. More precisely, in MOD's view, the answer 'true' shows congruence with descriptivist referential intuitions and the answer 'false' shows congruence with causal/historical referential intuitions. MOD conducted the experiment in three countries: India, Mongolia and France. The proportion of causal/historical answers was a bit higher with the test of truth-value judgments than with the test of referential intuitions in all three groups of test subjects, but the difference was not statistically significant. MOD found out that about a third of Indian, Mongolian and French participants chose the descriptivist answer 'true'. This result confirmed a substantial within-culture variation in truth-value judgments. MOD concluded that referential intuitions are congruent with the way people use proper names. People's intuitions about what proper names refer to in hypothetical scenarios are congruent with the way people are disposed to use those proper names. MOD claimed that their experiment provides evidence that referential intuitions are reliable since it shows that they are congruent with the way people use proper names.

In the remainder of this section, we will discuss a conceptual limitation of MOD's experiment. In the next section, we will present an empirical survey that shows that the data about truth-value judgments do not prove what MOD take them to prove.

Let us start with the conceptual limitation of MOD's experiment. Before presenting our criticism, it is useful to explain the extent to which MOD's experiment is a test of linguistic usage and not of referential intuitions. Martí (2012: 74–5) has restated her objection that MOD's experiment is not a test of linguistic usage, because it does not require participants to use proper names. MOD's experiment is based on a truth-value judgment task, which is still a metalinguistic evaluation. For instance, the question task tests participants' intuitions about the truth-value of the sentence 'Tsu Ch'ung Chih was a great astronomer' as uttered by the fictional character Ivy. The question task asks participants to reflect on Ivy's use of the proper name 'Tsu Ch'ung Chih', it does not require them to use it.

As Devitt and Porot pointed out to us,<sup>1</sup> a truth-value judgment test is a somewhat imperfect test of linguistic usage. In fact, given the logical and expressive role of the truth predicate, which is captured by the equivalence schema, the answer ‘true’ in response to MOD’s question task means that the participant asserts the proposition that she/he takes Ivy to have asserted. It is helpful to think of this process as divided into two parts. First, the participant understands the assertion that Ivy made by uttering the sentence ‘Tsu Ch’ung Chih was a great astronomer’. Second, by responding ‘true’ (‘false’) to MOD’s question task, the participant asserts (denies) the proposition that she understands Ivy to have expressed. Let X be the man that Ivy was taught to have discovered the solstice times and Y the astronomer who really made the discovery. Since participants know from the vignette that Y and not X made the discovery of the solstice times, the participants who respond ‘true’ to MOD’s question task understand Ivy as using the name ‘Tsu Ch’ung Chih’ to refer to Y and not to X. Therefore, the proposition that those participants understand and then assert is about Y and not X. This provides evidence from linguistic usage and not from referential intuitions. Indeed, there is no evidence that participants make the metalinguistic judgment that Ivy is referring to Y by using the name ‘Tsu Ch’ung Chih’. What their choice of ‘true’ evidences is that they *understand* Ivy as referring to Y not that they judge that she refers to Y with the name ‘Tsu Ch’ung Chih’. Likewise, if participants choose ‘false’, then this is evidence that they understand Ivy as referring to X and they deny the proposition that X was a great astronomer.

We argue that even if Devitt and Porot’s reconstruction of MOD’s experiment is correct, as we think it is, the methodology of MOD’s experiment for testing theories of reference against linguistic usage is flawed. We agree that the participants who answer ‘true’ understand Ivy as referring to Y with the name ‘Tsu Ch’ung Chih’ and are disposed to assert the proposition that Y is a great astronomer. The point is that these data give us no clue on whether those participants are disposed to use the name ‘Tsu Ch’ung Chih’ to semantically refer to Y. The mere fact that a participant understands Ivy to have expressed a proposition about Y is not evidence that she/he is disposed to use the proper name ‘Tsu Ch’ung Chih’ according to the semantic convention that the reference of ‘Tsu Ch’ung Chih’ is determined by the description ‘the person who discovered the solstice times’.

The point here is methodological and concerns the difficulty of testing a theory of reference by testing linguistic usage. A theory of reference is a theory about how reference is determined. It is not a theory about how people use proper names, at least not in the sense that one can extract from a theory of reference alone – taken separately from pragmatic assumptions – any prediction about people’s linguistic performances.

1 Devitt and Porot pointed this out to us in a talk they gave at the Warsaw Conference on Experimental Semantics in April 2016.

Granted that MOD's experiment tests a kind of linguistic performance – i.e. participants' understanding and evaluation of Ivy's utterance of the sentence 'Tsu Ch'ung Chih was a great astronomer' – there is at least one explanation of the answer 'true' that does not imply that the participant is disposed to use the name 'Tsu Ch'ung Chih' according to the descriptivist convention that the reference of 'Tsu Ch'ung Chih' is the satisfier of the description 'the man who discovered the solstice times'. Here is the sketch of the explanation. When the participant is asked to evaluate Ivy's utterance of the sentence 'Tsu Ch'ung Chih was a great astronomer', she/he imagines Ivy in the course of a conversation and tries to figure out her communicative intention. If the participant imagines (i) that Ivy's communicative intention is to speak of the man who discovered the solstice times and (ii) that Ivy is simply mistaken about his name, given that she/he learns from the vignette that Y is the man who really discovered the solstice times and, therefore, that Y is a great astronomer, the participant will judge that what Ivy said is true.

This sketched account explains a participant's 'true' answer as a typical case in which, as Kripke (1977) taught us, (i) the speaker's (Ivy) general communicative intention of using a proper name ('Tsu Ch'ung Chih') with its conventional meaning clashes with her specific communicative intention of making reference to a particular individual who is not the semantic reference of the name and (ii) the hearer privileges the speaker's specific communicative intention in order to figure out what the speaker said. One can explain a participant's 'true' answer by an account that appeals to the distinction between semantic reference and speaker's reference. Thus, in order to interpret the 'true' answer as evidence that the participant is disposed to use the name 'Tsu Ch'ung Chih' according to the descriptivist convention that its reference is the satisfier of the description 'the man who discovered the solstice times', we need some reason for ruling out the above explanation. But nothing in MOD's experiment provide such a reason.

One might reply to this kind of objection that the vignette gives no hint about any specific communicative intention of Ivy. Accordingly, when participants are asked whether Ivy's claim is true or false, they are driven to focus only on cases in which Ivy's specific communicative intention coincides with her general communicative intention. This means that participants are driven to reflect on what Ivy semantically expresses when she utters the sentence 'Tsu Ch'ung Chih is a great astronomer'.<sup>2</sup>

We believe that this is not a good reply. It is true that the vignette contains no information about any specific communicative intention of Ivy, but that is

- 2 Machery (2011), Machery and Stich (2012) and Machery et al. (2015) give voice to a similar reply to the objection that the results of MMNS's experiment are ambiguous because it is not clear whether the experiment tests intuitions about speaker's reference or semantic reference. See Ludwig 2007, Deutsch 2010 and Sytsma and Livengood 2011 for the speaker's/semantic reference objection to MMNS's experiment.

beside the point. We need to bear in mind that the participants in MOD's experiment are laypersons, not philosophers of language. They do not master the theoretical concepts of semantic reference, speaker's reference, what is semantically said, what is pragmatically communicated, (truth of) utterances, (truth of) sentences, (truth of) propositions, type expression, token expression etc. Therefore, it is a *prima facie* plausible hypothesis that the way in which participants accomplish the task of evaluating Ivy's claim is by making an effort of imagination for picturing a hypothetical conversation in which Ivy utters the sentence 'Tsu Ch'ung Chih was a great astronomer' with some communicative intention, even if no specific communicative intention is suggested in the vignette. Of course, our hypothesis is in need of justification, but nothing in MOD's experiment rules it out.

MOD do not take into account this conceptual limitation of their experiment. They claim that asking people to assess the truth-value of a sentence relates to – though is not the same thing as – assessing their willingness to assert it. They claim that typically if one says that someone else's utterance is true, one is disposed to assert it oneself (see MOD 2009: 691).<sup>3</sup> Thus, in MOD's view, if participants answer 'true', this is evidence that they are disposed to assert 'Tsu Ch'ung Chih was a great astronomer', and given that they learn from the vignette that Y is the man who discovered the solstice times, this is evidence that they use the name 'Tsu Ch'ung Chih' to refer to Y. Since Y is the satisfier of the description 'the man who discovered the solstice time', this is evidence that they use the name 'Tsu Ch'ung Chih' with the convention that its reference is determined by that description.

We think that MOD's argument is a *non sequitur*. From the premiss that one judges someone else's utterance of a sentence S to express a true proposition, it does not follow that one is disposed to utter S to semantically express that proposition. Suppose there is an emergency and John says, 'The man drinking a martini is a doctor'. If you know that that man is a doctor and a teetotaler but as a matter of fact he is holding a martini glass among a group of people all holding a pint of beer, you might say, 'That is true. Fetch him'. This does not mean that you are disposed to use the sentence 'The man drinking a martini is a doctor' to semantically express the proposition that that man is a doctor (notice also that this in itself proves nothing about the semantics of definite descriptions).

In the next section, we will present an empirical survey that shows that the fact that a participant answers 'true' to MOD's experiment does not mean that she/he uses the name 'Tsu Ch'ung Chih' according to the convention that its reference is the satisfier of the description 'the man who discovered the solstice times'.

3 See also Machery 2011: 125.

### 3. *Empirical evidence against MOD*

We set up an experiment in which a truth-value judgment task similar to the one used in MOD's experiment is followed first by a comprehension task and then by a forced-choice task.<sup>4</sup> We used the following vignette<sup>5</sup> and tasks:<sup>6</sup>

Mary is a university student. In her logic class, Mary is told that Gödel is the author of the theorem of the incompleteness of arithmetic, an important theorem in logic. Mary is a good student and she is able to demonstrate the theorem of the incompleteness of arithmetic, whose demonstration Mary attributes to Gödel. But this is the only information about Gödel that Mary has. Yet, Gödel is not really the author of the theorem of the incompleteness of arithmetic. Gödel stole the demonstration of the theorem from Schmidt, a logician who died soon after demonstrating the theorem. Gödel got hold of the demonstration of the theorem and claimed credit for it. So Gödel became famous for the demonstration of the theorem of the incompleteness of arithmetic. Most people are like Mary in this respect; the claim that Gödel discovered the theorem of the incompleteness of arithmetic is the only thing they have heard about him.

TASK 1 (truth-value judgment)

Having read the above story and accepting that it is true, when Mary says, 'Gödel was a great logician', do you think that her claim is

(A) true

or

(B) false?

4 The experiment was run in Italian. 103 students (University of Genoa, Italy), all Italian native speakers, participated in the experiment as volunteers ( $M = 27.83$ ;  $SD = 7.71$ ; 39 males, 64 females). The experiment was administered online.

5 The vignette used in our experiment differs slightly from MOD's one in order to be suitable for Italian test subjects. Here, we present the English translation of our Italian version of the probe.

6 It is worth noticing that the original claim of MMNS concerns the demographic variation of referential intuitions to the end of showing that Kripke's refutation of classical descriptivism is not supported by empirical evidence. This is why both MMNS's experiment and MOD's experiment were conducted over subjects belonging to different cultural groups. Our aim is not to deny the demographic variation of referential intuitions. Our goal is not to take a stand in favour of Kripke's refutation of classical descriptivism either. Rather, our aim is to show that truth-value judgments are not reliable guides for studying the way people use proper names to make reference to individuals. That is why we did not need to conduct our empirical survey over different cultural groups of people. It is enough that we prove our claim with respect to a single group of people.

## TASK 2 (comprehension)

If Mary said to you, ‘Write Gödel’s biography’, whose biography would you write?

(i) The person who really demonstrated the theorem of the incompleteness of arithmetic;

or

(ii) The person who stole the demonstration of the theorem.

## TASK 3 (forced-choice)

Suppose you wrote the biography and you send it to the Editor for printing. Which title would you choose for the biography?

(iii) ‘The biography of Gödel’;

or

(iv) ‘The biography of Schmidt’.

We used task 1 in order to reproduce MOD’s experiment. The data we collected from task 1 replicate the data of MOD’s experiment: 34.95% of participants chose the answer ‘true’ and 65.05% chose the answer ‘false’. This pattern of data is very close to the one collected in MOD’s experiment.

In order to investigate whether MOD are right that the answer ‘true’ to task 1 means that the participant is disposed to use the proper name ‘Gödel’ according to the convention that its reference is the satisfier of the description ‘the man who demonstrated the incompleteness theorem’, we added task 2 and task 3. The rationale behind task 2 and task 3 is the following. Task 2 drives participants to hypothesize a concrete and not convoluted linguistic exchange with Mary and tests how they understand Mary’s use of the name ‘Gödel’. The purpose of task 3 is to check how participants are disposed to use the name ‘Gödel’ independently of the hypothetical linguistic interaction with Mary.<sup>7</sup> We take it to be a safe assumption that the choice of the title of

7 One might be worried that the relevant usage of a proper name to be tested ought to be that of someone who is in a position like John/Ivy/Mary, i.e. people who do not believe anything else about the bearer of the name besides the information expressed by the description they associate to it. The test subjects of our survey, instead, are atypical in this respect because the vignette gives them much richer information about the bearer of the name. We agree that in order to confirm a theory of reference, the relevant usage to be tested ought to be that of subjects who are in the same informational position as John/Ivy/Mary. However, our goal is not to get empirical data about how people use proper names in order to confirm a theory of reference. Our goal is to show that MOD are wrong in assuming that if a test subject has the intuition that John/Ivy/Mary speak truly when they say, ‘Gödel was a great logician’, this is evidence that the test subject is disposed to use the name ‘Gödel’ according to the descriptivist convention that ‘Gödel’ refers to the man who proved the incompleteness of arithmetic. MOD make this assumption because their aim is to show that the test subjects who have referential intuitions in line with classical descriptivism are disposed to use proper names in accord with



the biography results from the coincidence between (a) the participant's general communicative intention of using the proper name that occurs in the title with its conventional meaning and (b) the participant's specific communicative intention of referring to the man who is the subject of the biography. Therefore if, for instance, a participant chooses the title 'The biography of Gödel', this is evidence that she/he uses the name 'Gödel' to semantically refer to the man who is the subject of the biography.

With respect to task 2 and task 3, there are four possible combinations of responses:

- (ii)/(iii): this combination shows that the participant understands Mary to refer to the thief of the demonstration with the name 'Gödel' and that the participant is disposed to use the name 'Gödel' to semantically refer to the thief.
- (ii)/(iv): this combination shows that the participant does not correctly understand the story of the vignette that says that Schmidt is the man who really demonstrated the incompleteness theorem.
- (i)/(iii): this combination shows that the participant understands Mary to refer to the man who really discovered the incompleteness theorem with the name 'Gödel' and that the participant is disposed to use the name 'Gödel' to semantically refer to that man.
- (i)/(iv): this combination shows that the participant understands Mary to refer to the man who discovered the incompleteness theorem and that the participant is disposed to use the name 'Schmidt' to semantically refer to that man.

If MOD were right that the answer 'true' to task 1 is evidence that the participant is disposed to use the name 'Gödel' according to the convention that its reference is the satisfier of the description 'the man who demonstrated the incompleteness theorem', then we would expect the answer 'true' to task 1 to be mainly followed by the combination (i)/(iii). In fact, this would be evidence not only that the participant understands Mary as referring to the man who demonstrated the incompleteness theorem with the name 'Gödel' but also that the participant is disposed to use that name to semantically refer to him. The data collected in our experiment falsify this prediction. We found out that the answer 'true' is mainly followed by the combination (ii)/(iii).

There are eight possible combinations of responses to the three tasks: A: True-(i)-(iii), B: True-(i)-(iv), C: True-(ii)-(iv), D: True-(ii)-(iii), E: False-(ii)-(iv), F: False-(ii)-(iii), G: False-(i)-(iii), H: False-(i)-(iv). We found

---

descriptivist conventions. Notice that if the above worry is sound, this adds more criticism of the reliability of MOD's experiment.

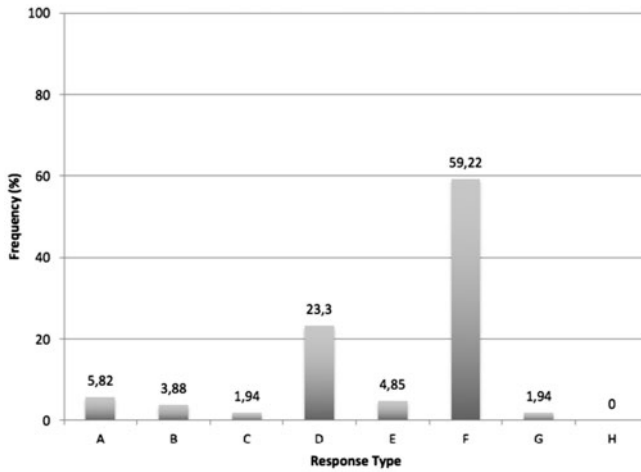


Figure 1. Comparison of the percentage of choices.

out that<sup>8</sup> 59.22% of participants chose the combination F, 23.3% the combination D, 5.82% the combination A, 4.85% the combination E, 3.88% the combination B, 1.94% the combination C, and 1.94% the combination G, 0% the combination H (see Figure 1).

Among the combinations with the answer ‘true’ to task 1 (A, B, C and D) a large majority of participants opted for D, i.e. true-(ii)-(iii). This result shows that MOD’s interpretation of the answer ‘true’ to the truth-value judgment task is mistaken. A large majority of the participants who answer that Mary’s claim is true, when she says, ‘Gödel was a great logician’, are not disposed to use the name ‘Gödel’ to semantically refer to the satisfier of the description ‘the man who discovered the incompleteness theorem’.

We conclude that MOD face a dilemma: if they are right that there is congruence between referential intuitions and truth-value judgments, then referential intuitions are not congruent with the way people use proper names, because truth-value judgments are not congruent with the way people use proper names. If they are wrong and there is no congruence between referential intuitions and truth-value judgments, then MOD’s experiment does not show

8 The Chi-square statistics with Yates’s continuity correction confirmed that the frequency of option D significantly differed from all other options (D vs. A:  $\chi^2(1) = 11.27, p < 0.0007$ ; D vs. B:  $\chi^2(1) = 14.92, p < 0.0001$ ; D vs. C:  $\chi^2(1) = 19.41, p < 0.0001$ ; D vs. E:  $\chi^2(1) = 13, p < 0.0003$ ; D vs. G:  $\chi^2(1) = 19.41, p < 0.0001$ ; D vs. H:  $\chi^2(1) = 24.94, p < 0.0001$ ). Similarly, the frequency of option F significantly differed from all other options (F vs. A:  $\chi^2(1) = 64.50, p < 0.0001$ ; F vs. B:  $\chi^2(1) = 70.48, p < 0.0001$ ; F vs. C:  $\chi^2(1) = 76.92, p < 0.0001$ ; F vs. E:  $\chi^2(1) = 67.44, p < 0.0001$ ; F vs. G:  $\chi^2(1) = 39.39, p < 0.0001$ ; F vs. H:  $\chi^2(1) = 83.84, p < 0.0001$ ). The difference between option D and option F was also significant ( $\chi^2(1) = 25.95, p < 0.0001$ ), as well as the difference between option A and option H ( $\chi^2(1) = 4.29, p = 0.03$ ). All other comparisons were not significant (all  $p_s = \text{n.s.}$ ).

that referential intuitions are a reliable source of evidence. Both horns of the dilemma leave referential intuitions still problematic.

Overall, the participants' preference for the combinations F and D suggests that they are not disposed to use the name 'Gödel' to refer semantically to the satisfier of the description 'the man who demonstrated the theorem of the incompleteness of arithmetic'. Whether and to what extent this result confirms the causal/historical theory of reference must be left for further investigation.

*University of Genoa*  
*Via Balbi 4, 16126 Genova*  
*Italy*  
*maxi@nous.unige.it*

*Psychology Unit*  
*DISFOR – Department of Educational Sciences*  
*University of Genoa, Corso Podestà, 2, Room 3C3, 16128 Genova*  
*Italy*  
*filippo.domaneschi@unige.it*

### *References*

- Deutsch, M. 2010. Experimental philosophy, intuitions, and counter-examples. *Review of Philosophy and Psychology* 1: 447–60.
- Kripke, S. 1977. Speaker reference and semantic reference. In *Contemporary Perspectives in the Philosophy of Language*, eds. P.A. French, T.E. Uehling Jr. and H. Wettstein, 6–27. Minneapolis, MN: University of Minnesota Press.
- Ludwig, K. 2007. The epistemology of thought experiments: first-person approach vs. third-person approach. *Midwest Studies in Philosophy* 31: 128–59.
- Machery, E., R. Mallon, S. Nichols and S.P. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92: B1–12.
- Machery, E., C.Y. Olivola and M. de Blanc. 2009. Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis* 69: 689–94.
- Machery, E. 2011. Variation in intuitions about reference and ontological disagreements. In *A Companion to Relativism*, ed. S.D. Hales, 118–136. Oxford: Wiley-Blackwell.
- Machery, E. and S. Stich. 2012. The role of experiment. In *Routledge Companion to Philosophy of Language*, eds. G. Russell and D.G. Fara, 494–512. London: Routledge.
- Machery, E., J. Sytsma and M. Deutsch. 2015. Speaker's reference and cross-cultural semantics. In *On Reference*, ed. A. Bianchi, 62–76. Oxford: Oxford University Press
- Martí, G. 2009. Against semantic multi-culturalism. *Analysis* 69: 42–8.
- Martí, G. 2012. Empirical data and the theory of reference. In *Topics in Contemporary Philosophy, Volume 10: Reference and Referring*, eds. W.P. Kabasenche, M. O'Rourke and M.H. Slater, 63–82. Cambridge, MA: MIT Press.
- Sytsma, J. and J. Livengood. 2011. A new perspective concerning experiments on semantic intuitions. *Australasian Journal of Philosophy* 89: 315–32.

**Abstract**

In order to uphold the claim that referential intuitions are a reliable source of evidence for theories of reference, Machery *et al.* (2009) conducted an empirical research by testing truth-value judgments. First, we discuss a conceptual limitation of Machery *et al.*'s experiment on truth-value judgments. Then, we present the data of an empirical survey that shows that people's truth-value judgments are not congruent with their use of proper names. We explain why the results of our empirical research refute the conclusions of Machery *et al.*'s experiment on truth-value judgments. We conclude that referential intuitions are still problematic.

*Keywords:* Semantics; reference; theory of reference; experimental philosophy