

University of Tartu
Department of Philosophy
Theoretical Philosophy

Henri Zeigo

**Testing Intuitions about Reference: What conclusions
should be drawn from the experimental data?**

Master's Thesis

Supervisor Prof. Dr. Daniel Cohnitz

Tartu 2014

Table of Contents

Introduction	3
1. Empirical Challenge for Reference	5
1.1 Intuitions and Theories of Reference.....	5
1.2 The Method of Testing Intuitions.....	7
1.3 The Empirical Study.....	8
1.4 Intuitional Sensitivity.....	11
2. Experimental Semantics	13
2.1 The Role of Hypothetical Cases.....	13
2.2 Are there any expert intuiters?.....	15
2.3 A New Empirical Study.....	16
3. Ambiguous Perspectives in Experiments	20
3.1 New Perspectives in Semantic Intuitions.....	20
3.2 Four Studies.....	21
4. Dual Nature of Reference	28
4.1 A New Ground for Reference.....	28
4.2 An Experiment for the Dualistic Account.....	29
4.3 Discussion about the Dual Nature of Reference.....	30
4.4 The Complicated Nature of Reference.....	33
Conclusion	38
References	40
<i>Abstract</i>	42

Introduction

In this thesis I am going to argue against the interpretation that experimental philosophers offered as an account of their empirical findings concerning philosophical intuitions about reference. Their empirical findings seem to exhibit that philosophers' and folk intuitions are sensitive to various factors and that these sensitivities cause diversity in the intuitive responses within and between different people and cultures (Machery, Stich, Mallon, Nichols 2004). It has been noticed that Western people are more likely to show intuitive responses that can be associated with a causal-historical view of reference, while East Asians are inclined to show intuitive responses which are associated with a descriptivist view of reference. The experimental philosophers argue that if there is such diversity in intuitions we cannot be sure that *our* intuitions provide us with reliable data that is needed for theory choice in philosophy of language. This means that our intuitions cannot serve as a foundation for theories of reference.

Roughly speaking, the basis for their analysis is intuitions of ordinary people, because the experimental philosophers believe that the intuitions of ordinary speakers are the foundation for theory-choice in philosophy of language. And those empirically tested variations in ordinary peoples' intuitions, as well as the intuitions of experts', lead the experimental philosophers to conclude that our intuitions generally cannot serve as a foundation for theory choice. In this thesis I am mostly considering the arguments by Machery, Stich, Mallon and Nichols who champion this particular empirical challenge.

I will argue that there are two main aspects which reduce the significance of their empirical study for philosophy of language. Firstly, I believe that experimental philosophers' empirical findings rest on a mistaken methodology which affects the interpretation of their findings. The design of the test methodology takes into account only two options, but the interpretation of this test allows more perspectives than the experimenters have thought so far. If the current experimental method does not consider more options for the test, then it is possible that the interpretation of the test results would have been also something else. When all perspectives are being considered equally well, then responses to hypothetical thought experiments do not differ so much at all.

Secondly, what makes this experimental method so problematic is that experimenters already assume that there are two kinds of people. People with either descriptivist intuitions or causal-

historical intuitions of reference. This distinction plays an important role in their interpretation of the test results. I am going to show that this distinction made by the experimental philosophers is rather deceptive. Peoples' use of language assumes the existence of both theories and they recognize them in quite a similar way with respect to the situation they are going to judge. Generally, I believe that experimental philosophers are forced to rethink how different perspectives and the dual nature of reference match together. If there are more options or perspectives to choose from and multiple ways to judge them, then the design of the vignettes that the experimental philosophers use is not suitable for this kind of measurement. The empirical findings of experimental philosophy lose their significance.

The first chapter is dedicated to explaining the notions that are part of this discussion and also to make a brief overview of how the experimental philosophers' method works and what their initial results tell us. The second chapter expands the topic by laying out the argumentation line between the experimental philosophers and Michael Devitt who questions their way of testing intuitions and defends the role of expertise. It is necessary to highlight this debate, because this forces the experimental philosophers to justify and improve their claims of research. Here I support the idea of experimental philosophers that hypothetical thought experiments are generally a good way to test our intuitions and there is no reason to value experts' intuitions more than lay peoples'. However, the third chapter shows that there are some methodological errors in the probes which mean that the initial empirical results are deceptive. Thus, the interpretation of empirical findings and the experimental challenge based on it are not convincing anymore. The fourth chapter extends the criticism further by arguing that the nature of reference is dual, and that, therefore our judgments can sometimes be affected by both descriptivist and causal-historical ways of thinking. When we consider the arguments in chapter three and four together, then it is a serious setback for the experimental philosophers, because the nature of reference and our referential intuitions are more complicated than they have thought so far. Nor is their methodology able to measure this kind of relation between different perspectives and the origin of the dual nature of reference. All of this reduces the significance of the initial empirical findings by the experimental philosophers for philosophy of language.

1. The Empirical Challenge for Reference

In this first chapter I am going to introduce the basic notions and theories of reference that constitute the essential part of the philosophical discussion. This conceptual framework helps to understand and follow the main argumentation line through the full inquiry. In the section 1.1 I introduce the concept of intuition and how this is related to the referential challenge. In sections 1.2 and 1.3 I talk about the design and the motivation for the experiments by Machery et al. Also, I am going to analyze their initial empirical findings. Section 1.4 summarizes the conclusion to be drawn from the initial experiments.

1.1 Intuitions and Theories of Reference

Philosophers of language and linguists are interested in how to explain our linguistic behaviour. This also involves the study of our intuitions. For instance, philosophers of language investigate how people refer to something and how they manage to express claims about a certain object. In these investigations, philosophers of language invoke their philosophical intuitions about hypothetical cases and use these as evidence for theories of reference. By philosophical intuitions I mean that these are kinds of semantic evaluations about how something is judged to be. In other words, philosophical intuitions provide us with data that is used to justify our theories of reference.

Generally speaking there is no common definition of intuition among philosophers, but despite the fact I am going to maintain a position that should be plausible for the rest of the discussion. As I have noted in the previous section, I am going to take intuitions in terms of judgment. These kinds of judgments have an opaque origin, because we do not know how they arrived in our minds. For instance, Jonathan Weinberg has described the origin of intuition as something inaccessible saying that there is no reasonable inferring or perceiving or even having sense of remembering that something is so (Weinberg 2007: 318).

It seems that the origin of intuition is very mysterious and in order to say something about intuitions we have to find other measurements to do so. However, what we can say is that intuitions play a significant role in philosophy. We advance philosophical theories on the basis of our ability to explain our philosophical intuitions and justify our philosophical beliefs on the basis of their accordance with our philosophical intuitions (Alexander 2012: 1-2). This reliance has allowed many philosophers to articulate theories which try to describe linguistic reality. I focus here on the most influential and at the same time most controversial theories of reference. These are the descriptivist theory of reference and the causal-historical view of reference.

The founder of descriptivist theory was Frege who articulated the initial idea (Frege 1892). According to the descriptivist theory, proper names refer to whatever person best satisfies the description that competent users associate with a name (Searle 1958; Jackson 1998). For instance, when a competent speaker uses the name 'Barack Obama' and the only thing he believes about Barack Obama is that he is the current president of the United States of America, then by believing so, his usage of that name refers to the current president of the United States (and thus to Barack Obama).

On the other hand, there is also the causal-historical view of reference. The central idea of this view is that the proper name refers to the person the proper name was introduced to refer to (Kripke 1972). The name keeps referring through a causal chain of reference-borrowing between competent speakers. Also, it should be noted that this view does not require competent speakers to associate any identifying descriptive content with the name (Machery et. al, 2004: B2). Basically this means that when somebody intends to introduce the proper name 'Barack Obama' for someone, it is going to be refer via a causal-chain of competent speakers to the last user of 'Barack Obama' who has acquired it from the users of this particular chain. When Barack's mother introduced her son to relatives, friends etc., then she is the first introducer of the particular proper name and the other people, who have gained the use of the name 'Barack Obama', are causally linked to this chain of speakers, and eventually, to Barack Obama.

1.2 The Method for Testing Intuitions

In the last section I described the basic (and probably the most well-known) theories of reference. These referential theories are crucial background-assumptions in the hypothetical thought experiments that experimental philosophers use for testing our philosophical intuition about reference. In this section will be explained what role the descriptivist theory of reference and the causal-historical view of reference, and also intuitive judgments in theory-choice, play in experimental philosophers' methodology.

Both theories of reference are distinct and therefore competing with each other. In order to find out which theory of reference describes our referential intuitions in the best way, the experimental philosophers like Machery, Mallon, Nichols and Stich (2004) have worked out a seemingly suitable method which they use for testing intuitions.

Their experimental method of testing involves philosophers, linguists and even lay people. All of them are assumed to have intuitions about what words refer to in actual and possible situations (Stich, Machery forthcoming: 19). Creating these actual and possible situations in a test form, the experimental philosophers are able to collect empirical evidence from the participants, who have to decide how to refer in these particular cases that have posed in the experiments.

Their method is a kind of socio-psychological approach, because instead of thinking theoretically, experimental philosophers turn to a wide range of people and hope to collect empirical evidence by letting them answer questions about various thought experiments. All these results are being analysed and used to make further generalisations about what are the participants' inclinations of theory-choice in referring to the object in question. It would be important to note that the experimental philosophers do not ask directly, which theories of reference the participants would prefer. Rather the preferences will come out after the answers are being associated to the particular theory of reference. At the end of the thought experiment there are normally two answers, where one is associated with descriptivist theory and the other with the causal-historical view of reference. The next paragraph explains precisely how the empirical research looks in practice and what the initial results of this empirical study tell us.

1.3 The Empirical Study

The motivation for empirical studies by the experimental philosophers is largely based on the previous cultural and anthropological studies that were run by Richard Nisbett. Nisbett and his colleagues conducted a series of experiments which all indicate that there are large and systematic differences between East-Asians and Western peoples' cognitive processes. They have discovered that cultural background plays a significant role in shaping human cognition, including the processes how people explain events and categorize objects (Nisbett 2003 and Nisbett et al. 2001).

According to his cultural studies, Nisbett makes a clear distinction between Western and East-Asian thought. He believes that East-Asians are inclined to think holistically which makes their cognitive processes more dependable on the context and relationships between objects. At the same time, Western peoples' thought is described as being analytic which means that they analyse objects independently of the context (Nisbett et al. 2001: 291).

This difference in focus makes the experimental philosophers formulate a hypothesis which predicts that cultural differences can be involved in our semantic intuitions as well. This means that our theory-choice in referring is largely dependable how we cognitively understand and value the world around us. Assuming that Western people make more causation-based judgments, which is explainable by their analytic thought, they would prefer to respond according to the causal-historical view of reference. East-Asians, on the other hand, would respond according to the descriptivist account of reference. To make sure whether this hypothesis has any empirical value, the experimental philosophers made several experiments to test their claim.

Machery et al. (2004) started with an empirical study for which they invited undergraduate students from Rutgers and Hong Kong University. All students were competent English speakers and they had a basic knowledge of philosophy of language. Experimental philosophers asked participants to read the famous Gödel case and then hoped to see answers for two different responses. The Gödel case is 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: 51)

The responses to the thought experiments were quite predictable. A significant number of Western participants preferred answer B, which according to the interpretation of experimental philosophers describe that those people have intuitions which rather support the causal-historical view of reference. At the same time East Asians were more likely to chose A. Their intuitions were associated with the descriptivist view of reference. The valuation of the test results were scored binomially by involving the empirical results of the Tsu Ch’ung Chih case as well. The Tsu Ch’ung Chih case was another thought experiment which the experimental philosophers ran alongside the Gödel case, but which has not received enough attention in this thesis, mostly because the results from that case were not so important for the further discussion. The answer B was given a score of 1 and the answer A was given the score of 0. When the results of the two experiments were summed, the final score ranges from 0 to 2. The cumulative score respective to the Gödel case for Western people was 1.13 and for East-Asians it was 0.63. These results show that there are serious variations in intuitions among and within cultures.

In order to increase the relevance of empirical findings, Machery et al. ran more experiments from different perspectives. They also designed similar hypothetical situation taking into account Asian background as well. Once again participants read the story and answered questions. This time as follows:

“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. Many people are like Ivy; the claim that Tsu Ch’ung Chih determined the solstice times is the only thing they have heard about him. When Ivy uses the name “Tsu Ch’ung Chih”, is she talking about:

(A) the person who really determined the solstice times? or

(B) the person who stole the discovery of the solstice times?”

(Machery et al. 2004: B9)

The outputs of Tsu Ch’ung Chih case did not differ much from the previous case. It seems that our philosophical intuitions are not universally shared. Empirical research suggests that there is intuitional diversity among and within cultures. This cultural diversity or any other type of diversity has a significant effect on our understanding of how philosophy should be done. Looking back into the history of analytic philosophy, philosophers have believed to have expert intuitions about philosophical questions. Particularly in the philosophy of language, they have been using different theories which describe our linguistic reality. Take for example the descriptivist theory or the causal-historical view of reference. All of them can be constructed without rising up from the armchair and these well-known theories have been articulated by expert thinkers. So far there has not been an urgent need for empirical evidence in order to produce good generalisations. We have simply trusted our philosophical intuitions. But now when the experimental philosophers have demonstrated that in Western and East-Asian countries the participants’ responses to the thought experiments vary within and across cultures, then it will lead to doubts whether we can trust our intuitions in theory-choice. We cannot be sure whether intuitions provide us with reliable evidence. This means that the role of expertise is also losing its significant value among philosophers who have been doing philosophy in armchairs. All of this together has made philosophers question whether traditional ‘armchair philosophy’ can still be useful or whether we must change our current methodology.

Machery et al. do not support the traditional methodology, which according to their findings is not successful in explaining the linguistic reality as it is. One way to see which theory describes our referential thinking is to investigate how people really think. This is not just a

methodological difference in philosophy. It also causes a change in understanding of how philosophy is done.

1.4 Intuitional Sensitivity

So far philosophers have thought that intuitions can be shared universally. New empirical studies that were made across and within cultures have shaken this dogmatic view about intuitions. What philosophers have not expected is that there are various factors which have direct influence on shaping our intuitive judgments. This is something that philosophers have not anticipated before. Knowing that our philosophical intuitions can be sensitive to unknown factors, we should probably put more effort in to the investigation of these factors.

For instance, this widely discussed cultural diversity should direct our attention to various unknown factors that might have caused these results. The experimental philosophers have reason to believe that our philosophical intuitions seem to be sensitive to our own cultural background (Stich 1988, 1990), but also to sexual, educational and moral factors.

One test made by Jennifer Zamzow and Shaun Nichols (2009) has shown that when men and women are asked to consider the famous Trolley case in two different variations, they give different responses. In the first part the participants were quite unanimous. Men and women mostly agreed to save five persons rather than one stranger from a moving train. But in the second case when the stranger was replaced with a 12-year-old boy, the women were not so unanimous anymore. The number of participants, who would have directed the train toward the five persons, increased dramatically compared to the previous case. These empirical findings indicate that women think that it is morally acceptable to redirect the trolley in the second case when they can save a 12-year-old boy from the lethal accident (Alexander 2012: 75).

This and many other similar thought experiments have indicated that there might be a huge range of unknown factors that shape our intuitive responses.

And of course, if there is too much sensitivity, then philosophers cannot be sure whether we can trust intuitions as evidence. There seem to be two options how to deal with this sensitivity: philosophers either must explain why we can tolerate intuitional sensitivity as such or they just must claim that these philosophical intuitions are not evidence in our practice (Alexander 2012: 82).

2. Experimental Semantics

In this chapter I would like to direct attention to some procedural questions that might have become questionable during this paper and that other philosophers have also criticised. In the section 2.1 I argue why hypothetical cases still play a significant role in empirical research, despite the fact they have been criticised. Therefore, in sections 2.2 and 2.3 I explain why we should involve lay speakers in the test. This would be necessary to discuss, because at the end of my thesis I am going to present experiments which use similar empirical procedures in order to argue against the initial empirical findings of Machery et al.

2.1 The Role of Hypothetical Cases

The empirical study which Machery and other experimental philosophers ran has focused on hypothetical cases and to the ordinary speakers' participation in testing our philosophical intuitions about reference. Some philosophers, like Michael Devitt, have criticised the lay speakers' capacity to evaluate this kind of hypothetical thought experiments as in the Gödel case. Devitt believes that expert intuitions count more than lay speakers' intuitions and that lay folk are not able to deal with the highly abstract cases that are presented in the experiments. In this chapter I am going to argue that Devitt's criticism of experimental philosophy is mistaken. This chapter is also necessary for my further argumentation, because at the end of this thesis I am going to return to the thought experiments in question. For that purpose I would like to show that the general idea of testing our philosophical intuition in the way that experimental philosophers suggest is not faulty. The real problems seem to be somewhere else.

Michael Devitt starts his argumentation with a criticism that is addressed at the empirical findings that were initially presented in the Machery et al. paper "Semantics, Cross-Cultural Style".

Devitt believes that folk should not be involved in a test where we use hypothetical cases, because hypothetical cases are quite outlandish and our intuitive responses to them do not say much about reference in ordinary course. Therefore, intuitions about actual cases, even if sometimes fallible, are easy to test and make more sense in the case of ordinary competent

speakers. For instance, we can control whether people use the proper name 'Barack Obama' in a correct way or not. Thus, Devitt concludes that we should be interested in actual cases, where we can follow whether a person has succeeded in referring to the object in question (Devitt 2011: 425-428).

“It seems that hypothetical cases are mainly for philosophers who have a special training that helps them to solve those problems. We can more easily examine actual cases which are part of our everyday cognition. Some people have probably a better access to the data of actual cases than others and that makes them more experts. This better access could give a person more reliable intuitions in reporting his intuitive judgments about the actual case.” (Devitt 2011: 425-428)

Despite the fact that intuitions about actual cases are sometimes more easily followed than hypothetical ones, as Devitt would say, then from this claim it will not follow that the folk are unable to examine the hypothetical thought experiments. The Gödel case does not contain highly technical vocabulary that might be unknown for non-experts, who do not know anything about the philosophy of language. The aim of the test is to give an exact description what John knows about Gödel and asks how he refers with the name even when the circumstances in the story have been changed. The experimental conditions for philosophers and non-philosophers are always the same. Very limited restrictions in test design do not give much space for philosophers to reconsider other possibilities for theorising, because two given answers of how to interpret the hypothetical situation put philosophers and ordinary people into similar position, where they have to judge how John refers with the name.

Moreover, the distinction between hypothetical and actual cases is not well established. If the idea that hypothetical cases are always very outlandish and do not play any significant role in everyday cognition, then this seems just false. It is true that hypothetical situations are sometimes more complicated than actual cases, but this does not mean that these complications are not part of our everyday cognition that we are facing in life. For instance, there can be many situations where people, including the experts, associate the wrong description with a name or pick out the wrong referent and so forth. Such cases happen time to time to anyone who is ignorant about something. Therefore, the test that the experimental philosophers designed provides a good opportunity to see how people would have referred, if they had been in John's position.

Mallon et al. have provided a good example from the history of science to show that such cases also happen in real life. The story tells how a dinosaur called Triceratops was once mischaracterised. First they presented a neutral condition which has been acceptable for a long time.

“The dinosaur Triceratops has a giant bony structure, a “frill”, behind the head. As scientists have accumulated more fossils, they have learned much more about this dinosaur. For instance, as the Triceratops ages, the frill becomes much longer. The frill also develops giant holes and these holes become covered with keratin, which is a key component in human skin. Researchers have argued about different ways that the frill might become longer and develop holes. This remains an issue of discussion in dinosaur research.” (Mallon, Nichols, Pinillos 2013: 12)

Then later scientist managed to do more empirical investigations and now the description which corresponds to the Triceratops is as follows:

“The Triceratops is a large dinosaur with a giant bony structure, a “frill”, behind the head. However, our understanding of the Triceratops has changed dramatically over the last century. As scientists have accumulated more fossils, they have come to recognize that there were many mistakes made in the initial description of this dinosaur. For instance, scientists thought that the Triceratops had skin like an elephant, but it turns out that it really had scales like an alligator. They thought that Triceratops was exclusively a plant eater, but now hold that the Triceratops was at least partly a meat-eater. And most strikingly when scientists first named Triceratops, they thought it was an ancient bison. Only later did they realize that it was really a dinosaur” (Mallon, Nichols, Pinillos 2013: 12)

This story and many other similar cases in our everyday practice show that we are faced with instances of seemingly outlandish hypothetical situations all the time. When we put this real life example into a test, it would be interesting to know how we would have described our philosophical intuitions about reference, when the circumstances had changed as they can in real life as well.

2.2 Are there any expert intuiters?

From the previous criticism of Michael Devitt there also emerged the idea that there are people who have more reliable intuitions than others. When Devitt did not want involve lay speakers in the test, he presumably assumed that some peoples’ intuition count more, namely the intuitions of experts. In this section I am going to show that this claim does not have any scientific grounding for being true.

First, Devitt is forced to answer why it is so that there are cultural differences in intuitive responses. This means that Devitt must explain how words refer in situations where our intuitions of reference vary within and across cultures. How it is possible that Western people are more inclined to Kripkean intuitions and Asians to the descriptivist viewpoint? (Machery 2012: 225)

Secondly, there is also a deep disagreement among philosophers of language about what the right theory of reference is. There are philosophers who advocate the descriptivist view as the best theory that describes our linguistic reality, and oppositely, there are philosophers who think that the causal- historical view does it best. Disagreement between these two views does not let us say that one is more expert than another (Machery 2011: 13).

Thirdly, we know that empirical research has failed to find evidence that linguists' syntactic intuitions are more reliable (Machery 2011: 14). For instance, Culbertson and his team found that linguistic expertise did not improve any reliability in judgments about grammaticality. Instead they managed to show that some participants had experience with behavioural experiments, while others were unsure when they had to judge something.

All three points by Machery have raised serious questions about the role of expertise in answering hypothetical thought experiments. Theoretically there is no reason to exclude non-philosophers from empirical research because there is no clear evidence which shows the existence of superior intuitions. The next section will demonstrate that Machery et al. have good reasons to hold their hypothesis about the role of expertise.

2.3 A New Empirical Study

To show that experimental study is moving on the right path, the experimental philosophers made more tests in order to prove that their findings actually do cast some doubt on the Expertise Defence. Their new experimental study is more ambitious this time. Machery et. al invited the world's top linguists, philosophers etc., to a web study, where they were asked to answer to the Tsu Chung Chih vignette. Having read the above story and accepting that it is true, when Ivy uses the name Tsu Chung Chih, then they had answer to a traditional question: who do you think she is actually talking about?

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

Moreover, after this they were asked to answer on a scale ranging from 0 (not sure) to 100 (sure). Finally they had to specify whether they work on language and what their area of specialisation was, name some famous linguists that best exemplify their theoretical perspective on language, etc. Their experimental study was this time more widespread than what they did in 2004 (Machery 2011: 17).

The results say that the majority of participants in all groups reported Kripkean intuitions, but the proportion varied across groups from 66% to 98%. What is more, further analysis allowed Machery to conclude that there are three distinct groups of answerers. In the first group were semanticists and philosophers of language. Their average result was 96% in respect of Kripkean intuitions. In the second group were historical linguists and sociolinguists. Their average result was 68%. Finally, in the third group were comparably educated folk. Their final result was something between first and second group, namely 76% respect to the Kripkean intuitions (Machery 2011: 19-20).

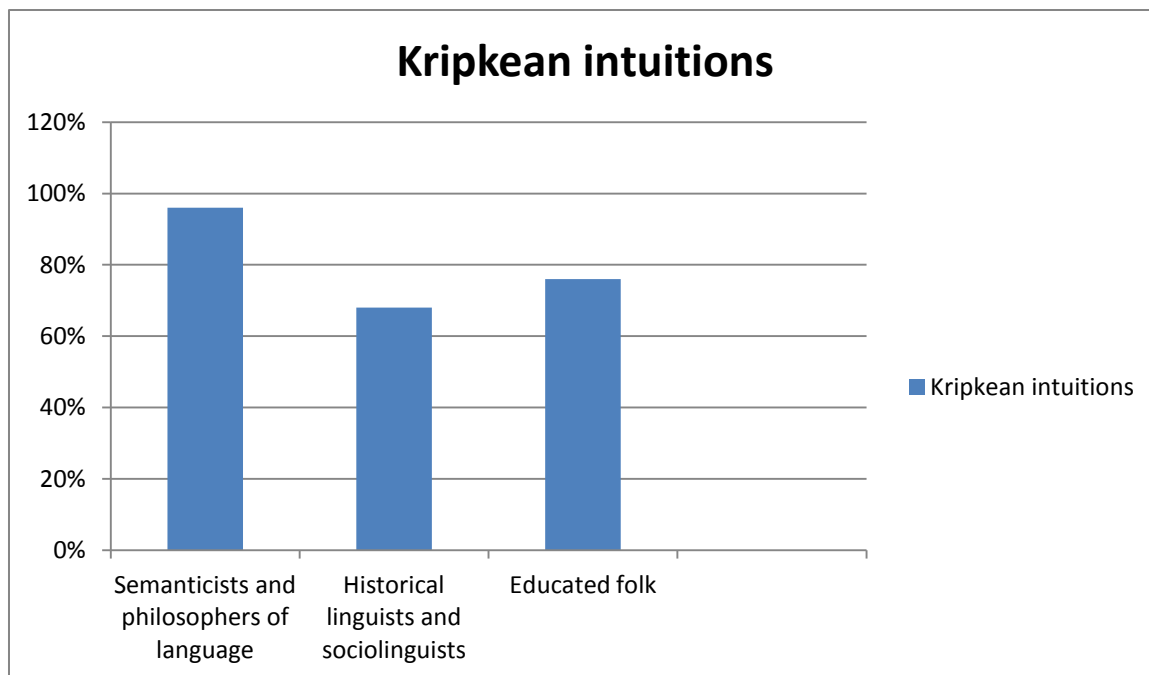


Figure 1 (Machery 2011)

These findings have led Machery et al. to critical conclusions about the Expertise Defence. First, they point out that educated folk like most of the experts had Kripkean intuitions in response to the Tsu Chung Chih vignette. If folk and experts share more or less similar intuitions, then experts will have no superior expertise over lay persons. In order to argue for Expertise Defence, Devitt must demonstrate that experts give more precise answers and their responses are highly different from folk, but in respect of the test results this does not seem evident (Machery 2011: 22).

Secondly, all groups had some expertise about language, but when we compare them with each other then we can see some differences between the responses they gave. For instance, sociolinguists and philosophers of language are not sharing a similar result in a report of Kripkean intuitions. This indicates that even within experts there are disagreements which do not allow them to call one group to be better than another (Machery 2011: 23-24).

Their final conclusion emphasises the results between folk and experts in general. They say that neither the answers of historical linguists, anthropological linguists, and sociolinguists nor the answers of philosophers of language and semanticists differ significantly from the

answers given by comparably educated lay people. If this is so, then we do not have reason to talk about expertise at all (Machery 2011: 25).

So far I have argued that the problem is not so much in the hypothetical thought experiments. And probably there is no reason to make a difference among the participants, because we have not proved the significant superiority of experts' intuitions.

The criticism has not reduced the value of empirical study so far. The general idea of testing referential intuitions is quite promising. I believe that it would be more fruitful to examine the methodological design of the empirical study instead of considering the legitimate value of the test. We know the possible consequences of the empirical results, but what we do not know is how the design of the test methodology affects the final results, which have led us to this referential challenge. For this purpose it would be more reasonable to examine the outputs of methodological design separately, without criticising the basic idea of testing intuitions. The next sections will explain whether the design of methodology has an influence on the empirical results of experimental philosophy.

3. Ambiguous Perspectives in Experiments

In this chapter I am going to argue against the methodology used by the experimental philosophers. In section 3.1 I introduce arguments against the methodological design that Machery et al. use. In the section 3.2 I present the empirical study by Sytsma and Livengood in order to show that the methodological design can have an effect for the initial empirical results of Machery et al.

3.1 New Perspectives on Semantic Intuitions

I am using arguments and experiments that Sytsma and Livengood have shown in their paper “A New Perspective Concerning Experiments on Semantic Intuitions”. In their paper they criticise the empirical findings of Machery et al., which according to their point of view rest on a mistaken interpretation. Namely, they think that the probe of Machery et al. consists of an epistemic ambiguity and there is empirical evidence for that. Moreover, if this is shown by their new experiments, then Machery’s argument about the variations in intuitions is no longer convincing. Secondly, the new thought experiments made by Sytsma and Livengood demonstrates that experts’ and lay intuitions do not differ so much as Expertise Defenders have hoped, and thereby, the role of lay people can be really promising in this kind of empirical research.

There are mainly two ambiguities that Sytsma and Livengood mention. First of all, they point out that Gödel’s case does not clearly indicate whether the (A) and (B) answer choices are to be read from the narrator’s epistemic perspective or from John’s epistemic perspective. Secondly, they say that the test question is ambiguous with regard to speaker’s reference and semantic reference (Sytsma and Livengood 2011: 319).

Starting from the first type of ambiguity, Sytsma and Livengood say that one problem with epistemic perspective ambiguity is that participants are left unsure about whose epistemic perspective should be adopted in deciding who these descriptions refer to. This raises the possibility that different people might associate the same description with different people from the story (Sytsma and Livengood 2011: 319).

What they particularly mean is that there are several distinctive viewpoints for a thought experiment. The narrator's point of view assumes that 'the person who really discovered the incompleteness of arithmetic' denotes Schmidt and 'the person who got hold of the manuscript and claimed credit for the work' denotes Gödel. However, John's perspective is something very different. John only knows that Gödel discovered the incompleteness of arithmetic; he has never heard anything about Schmidt before (Sytsma and Livengood 2011: 320).

If this choice would be up to me, then I think there is nothing wrong when I agree with both of them to some extent, but it seems that experimenters do not assume this option. When I say A, that John talks about the person who really discovered the incompleteness of arithmetic, then I may think that John, according to the information he has, cannot talk otherwise. But when I say B, John talks about the person who got hold of the manuscript and claimed credit for the work, then I do not talk so much about John's limited beliefs anymore. Both answers make sense to some extent. Thus, it remains unclear how these options demonstrate something about my philosophical intuitions about the referent.

Moreover, there is also a second ambiguity that Sytsma and Livengood mention. This one considers the speaker's referent ambiguity. In this case authors are observing what John's intentions might be. For a participant there will be two options to consider: either to treat John as intending to talk about Schmidt or to treat John as intending to talk about Gödel. Here they make a distinction between semantic and speaker's referent. One could assume that 'Gödel' actually refers to Gödel, but nevertheless, he could still consider John to be talking about Schmidt. For instance, when a participant answers A, despite sharing Kripkean intuitions about the semantic reference of the name 'Gödel', then they do so because they believe that John intends to talk about Schmidt (Sytsma and Livengood 2011: 321).

3.2 Four Studies

Next Sytsma and Livengood test their assumptions in practice. They developed four studies which demonstrate the influence of ambiguity for the findings of Machery et al.

In the first study they test intuitive responses from three different perspectives. These are the original Gödel case, John's perspective and narrator's perspective.

“Original: 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?”

John’s Perspective: When John uses the name ‘Gödel’, does John think he is talking about: (A) the person who the story says really discovered the incompleteness of arithmetic? Or, (B) the person who the story says got hold of the manuscript and claimed credit for the work?”

Narrator’s Perspective: When John uses the name ‘Gödel’, is he actually talking about: (A) the person who the story says really discovered the incompleteness of arithmetic? Or, (B) the person who the story says got hold of the manuscript and claimed credit for the work?” (Sytsma and Livengood 2011: 322)

They summed up B answers and what was found is quite expectable from Sytma and Livengood’s point of view. Our intuitive responses are affected by various other perspectives and those may influence the results of experimental study. The results are demonstrated in Figure2 (Sytsma and Livengood 2011: 323).

It should also be noted that what makes this experiment more mysterious is that Westerns’ responses to the original case differed largely from the results that were made in 2004 by Machery et al. Sytsma and Livengood did not get similar results or even the proportions of answers that ascribed Westerners’ responses compared to East-Asians. If that was already ambiguous, then it will cast more doubt on the whole experimental study what the experimental philosophers have run for so long time. These empirical findings what Machery et al. reported in their paper “Semantics, Cross-Cultural Style” provided the basis for cultural diversity. Now it seems that the whole ground of discussion has started to shake.

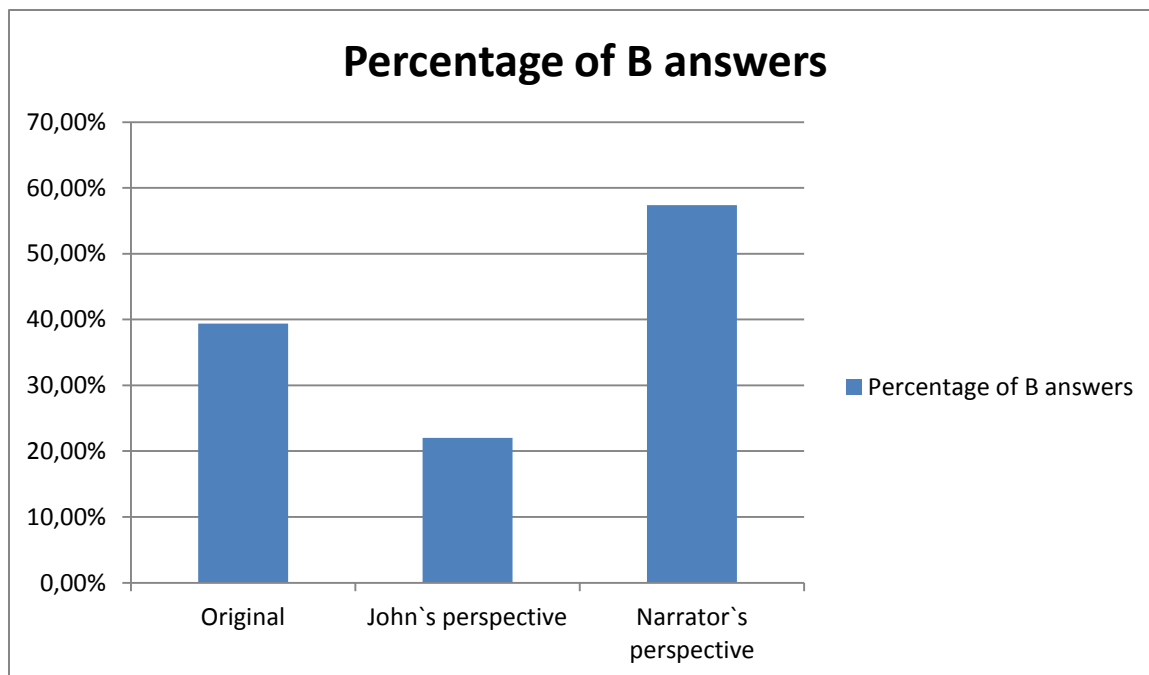


Figure2 (Sytsma and Livengood 2011: 323)

The following studies adhere to the same line of thought. Their second study expanded the question and also involved philosophers in the test. This time participants had to respond to the clarified Narrator's perspective experiment which is as follows:

“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: 324)”

They found that 73.8% of non-philosophers answered B compared to 75.9% of philosophers. The difference was not statistically significant. So it would be difficult to say that there is a very large distinction between expert and lay opinion. The second empirical study has a serious effect for Machery et al. and Devitt's account in expertise question.

After the revised empirical study Machery et al. cannot claim that philosophers and non-philosophers' intuitions are somehow sensitive to the various factors that influence participants' decision-making process. It is because all participants give relevantly constant answers despite the educational background they have.

These results are also interesting for the discussion in the previous chapter about the ability of lay people to evaluate hypothetical thought experiments. Devitt assumed that lay people are not good at making judgments about hypothetical situations, because these are very outlandish and need some special training and education that only philosophers have. However, the results of the revised test are not so pessimistic at all. These findings should convince philosophers to trust folk more than they have done so far.

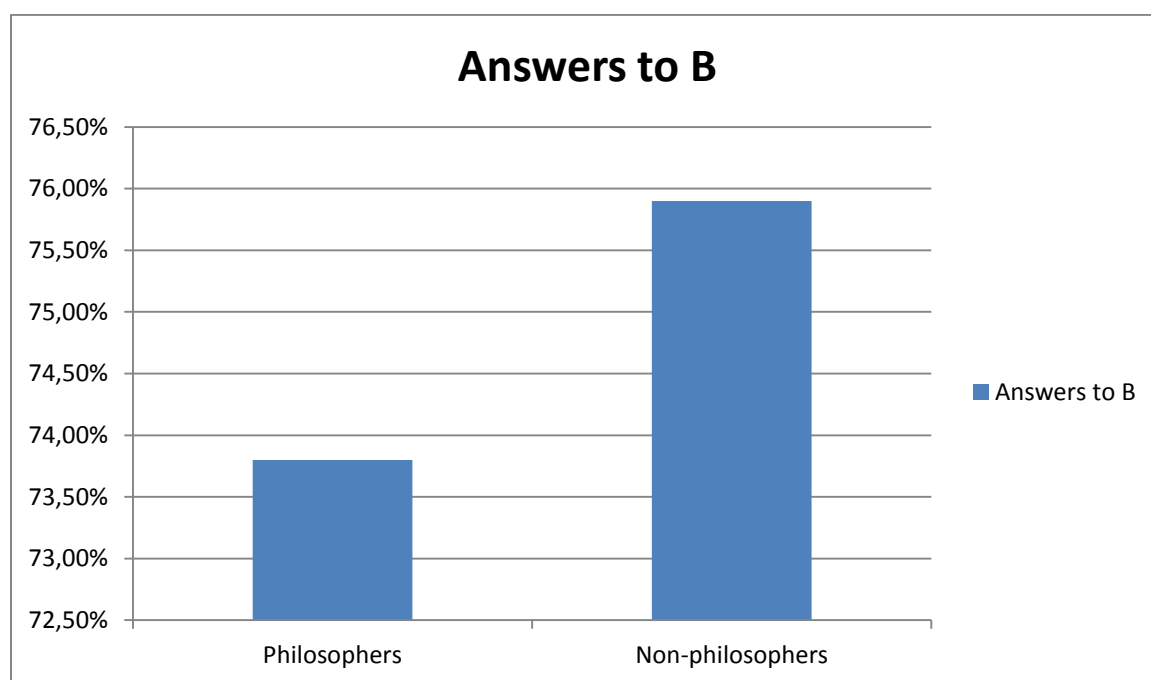


Figure3. (Sytsma and Livengood 2011:323)

The third and fourth studies focused on subjects and question-related factors, but the general conclusion remains despite the viewpoint we are going to choose. That is, the different perspectives influence our semantic intuitions. The third study was made face-to-face and descriptions that were given for A and B in all four probes were counterbalanced for a specific order. We can see the combined results of first two studies in Figure 3 and the third study in Figure 4, where results appear quite similar to the previous ones.

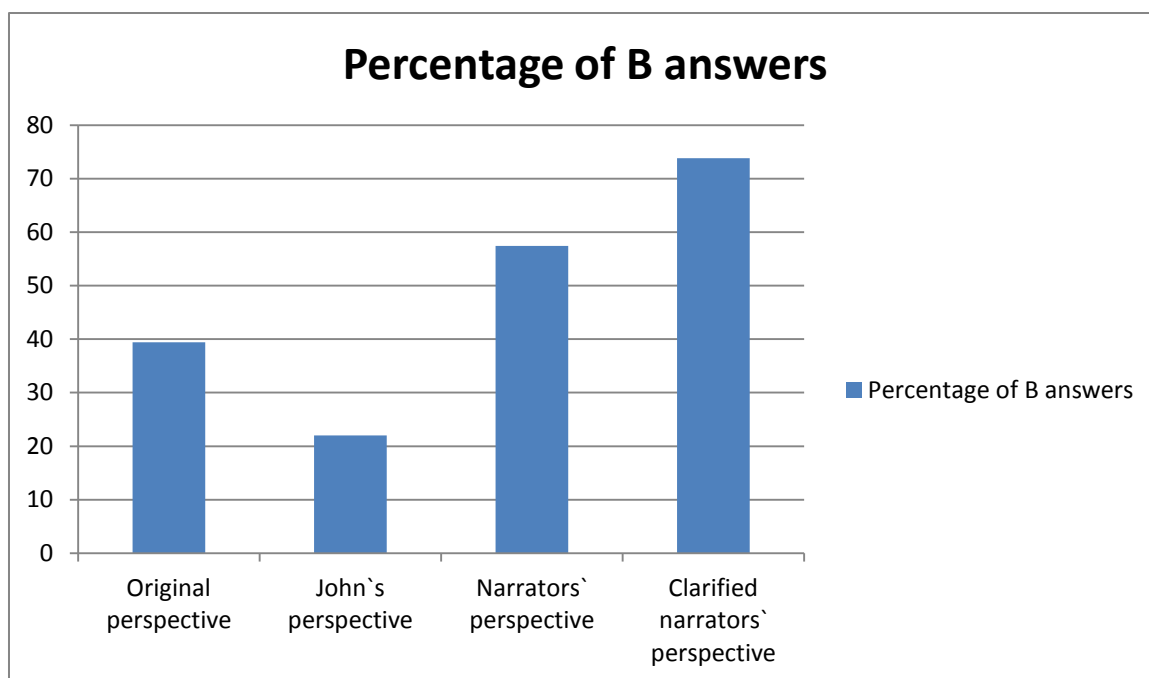


Figure 4. The combination of study 1 and 2 (Sytsma and Livengood 2011: 326).

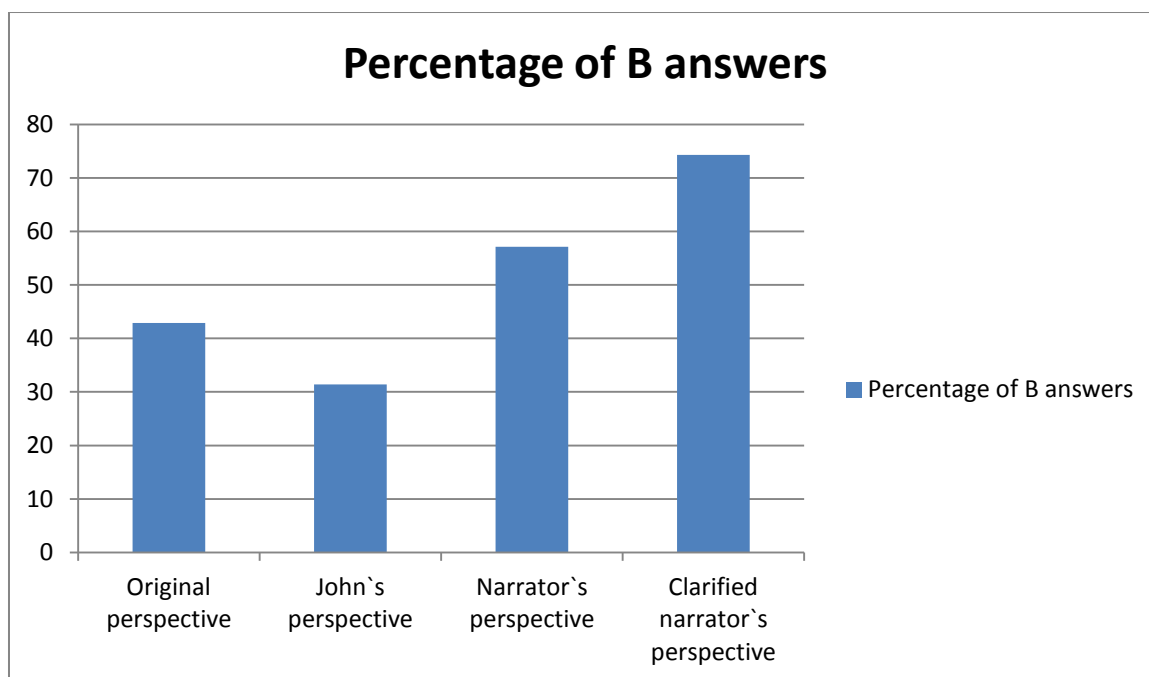


Figure5. The results of study 3. (Sytsma and Livengood 2011: 326)

In the fourth study Sytsma and Livengood gave the original Gödel case to participants. They also added a second paper for which they asked participants to indicate which of the following statements describe their understanding about the test question.

They asked: “1) When John uses the name ‘Gödel’, does John think he is talking about (A) or (B). Or 2) When John uses the name ‘Gödel’, is he actually talking about (A) or (B).”

(Sytsma and Livengood 2011: 326)

The results of fourth study are summarised and presented in a following graph.

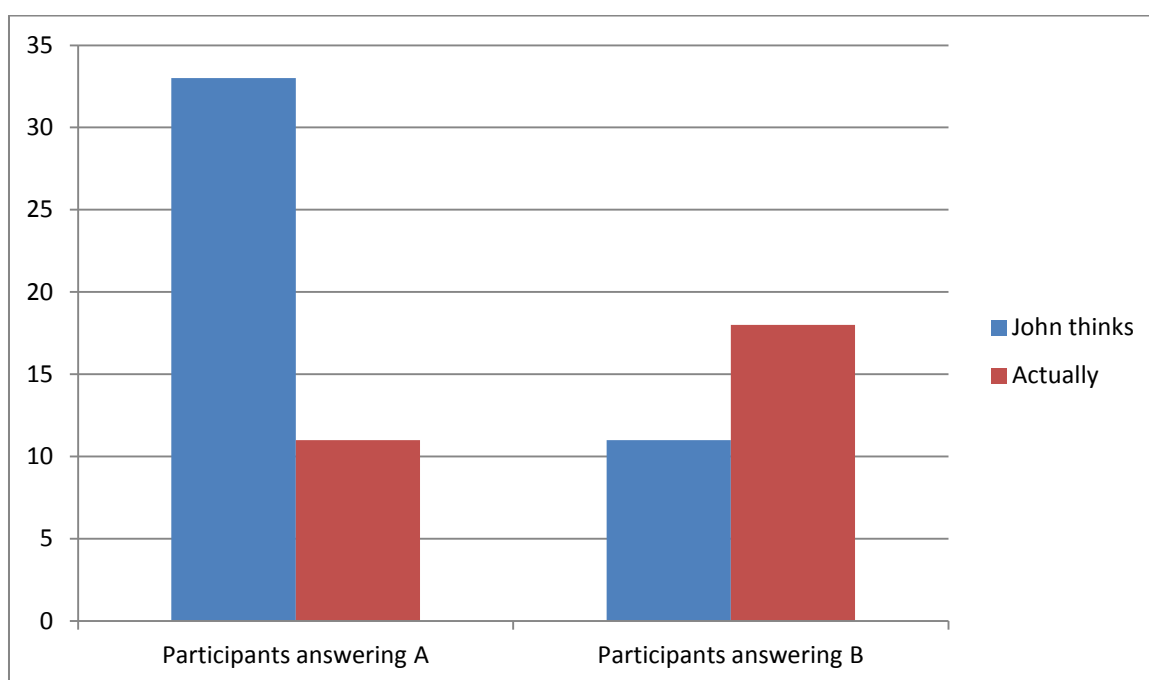


Figure6. The results of fourth study (Sytsma and Livengood 2011: 327).

The summary of all four studies raises many questions about the methodological design of testing referential intuitions. Several perspectives that have not been taken into account may influence the empirical results of the original probe. New perspectives that experimental philosophers did not include in the test explain the significant variations in intuitions. As new

empirical study has shown, if all possible interpretational perspectives are being studied carefully, the variation in intuitions will not be so remarkable anymore. Thus, our philosophical intuitions can still be promising and trustworthy. In addition, it would not be a problem to trust lay speakers' intuitions either, while they have shown similarities in answering to the hypothetical thought experiments.

Some philosophers may still argue that Sytsma and Livengood tested only Westerners and therefore these conclusions do not explain the variations in intuitions. I can agree with the opponents that for a scientific interest it would have been reasonable to test East Asians as well. Sytsma and Livengood published their four empirical studies in 2011. One year later James Genone mentions in his paper called "Theories of Reference and Experimental Philosophy" that Sytsma and Livengood have run a series of unpublished experiments (Genone 2012: 158). Collected data involved the intuitions of Japanese participants using the similar revised materials that Sytsma and Livengood used in their four studies. However, this time Japanese responses did not correspond to the initial results that Westerners gave. Rather the interpretation of empirical study confirmed their stronger inclinations to a descriptivist account. But at the same time we cannot avoid the fact that experimental philosophers' methodological enterprise was not sufficient to give adequate test results. It is correct to say that the methodology of the original Machery et al. experiments was faulty and the empirical findings are not so convincing in the philosophy of language anymore. Although cross-cultural variations remain in the revised tests, then I assume that this leads to another problematic aspect of empirical research, which needs to be answered. There are still totally different philosophical intuitions about reference across different cultures. This makes me ask whether we have understood the nature of reference properly.

4. The Dual Nature of Reference

In the last chapter Sytsma and Livengood demonstrated that the method of testing intuitions that the experimental philosophers used in reporting their significant findings appears to be broken. The design of the experiment allows too many variations or different perspectives that can affect peoples' responses to the thought experiment. Surprisingly, even the original perspective to the Gödel's case showed remarkable differences in the test results. However, after Sytsma and Livengood managed to do some extra tests, they showed that Japanese people are inclined to have descriptivist intuitions, which means that the old cross-cultural debate remains. This was also something that the experimental philosophers (Machery et al.) found in 2004 when they published their paper called "Semantics, Cross-Cultural Style." Due to the ambiguity and methodological errors they were not able to reach similar results like Sytsma and Livengood did, but the interpretational side of this cultural debate is still worth considering. We have not found the reason why Westerners give responses that are associated with Kripkean intuitions and why East-Asians are giving responses that are associated with descriptivist intuitions. All this makes philosophers reconsider that maybe we have not understood what is the true nature of reference.

4.1 A New Ground for Reference

It seems that the majority of philosophers of language assume there is one unique theory for our referential judgments. They have believed that our referential intuitions correspond strictly either to the descriptivist theory of reference or the causal-historical view of reference. During the experimental study no other accounts are involved in the test. I believe that this is one reason why philosophers have not reached a common viewpoint.

One way to provide a solution to our linguistic debate is to explain whether theories of reference or at least the fragments of these theories can function together. This idea has led some philosophers to find a better solution. For instance, Genone and Lombrozzo (2012) have tried to establish an account which creates suitable grounds for the existence of both theories

of reference. In their work they are inspired by Evans' (1973) example that described the story of Marco Polo. The story tells that when Marco Polo learned the name 'Madagascar' from local sailors, he used the same reference as the locals did. However, Polo mistakenly believed that he is referring to the mainland, not to the island which would have been the correct referent. In this case Kripkean theory would say that the reference was unsuccessful. However, this does not explain the change in the reference of the name. This story which could describe very easily our everyday linguistic practice has led Evans to suggest that descriptions must play a significant part in our linguistic usage concerning the causal-historical view. He has also proposed that each use of a name is associated with a file of information that an individual stores about the referent of the name. At the same time, some information in that file can be false as well. So it should be clear that a theory which involves both causal and descriptivist factors in determining reference would best explain the reference of proper names (Genone and Lombrozzo 2012: 720).

4.2 An Experiment for Dualistic Account

Genone and Lombrozzo's next step was a careful examination of these factors among participants. They involved 192 undergraduate students who completed a questionnaire about the Tyleritis' case:

“There is a small island in the Indian Ocean called Alpha. Natives of Alpha, called ‘‘Alphians,’’ sometimes catch diseases not found anywhere else in the human population. When this happens, they consult Alphian doctors. One of the diseases on Alpha is tyleritis. Tyleritis is a disease that affects muscles and causes muscle pain. It is caused by exposure to a rare mineral, can be diagnosed with a blood test, and can be cured by an injection (Genone and Lombrozzo 2012: 724).”

Genone and Lombrozzo gave several perspectives about the Alphian disease called 'tyleritis'. First they explained general facts about tyleritis, then they presented Alex's and his doctor's beliefs about the tyleritis. All factual contents of these perspectives were similar. Now the story goes on in the examination room where Alex and his doctor are talking:

Alex thinks: 'I might have tyleritis.'

His doctor thinks: 'I wonder if this patient has tyleritis?'

When Alex and his doctor each have a thought in the examination room, are they having a thought about the same disease?

(A) Yes, they are both thinking about the same disease.

(B) No, they are not thinking about the same disease. (Genone and Lombrozzo 2012: 724-725)

The story goes on to when Bob from the island Brom had a different disease, which was also called tyleritis. What makes this hypothetical experiment tricky is that if Bob and Alex should meet and talk about tyleritis, then we can ask whether or not they talk about the same disease. Meanwhile, the circumstances of this story changed in many ways. There were four distinct parts each involving a certain relation between description and causal origin.

Experimenters tested participants' who were gathered to evaluate this hypothetical case. Genone and Lombrozzo managed to identify at least two factors that had influence participants' decision-making process when judging the correct referent in the Tyleritis case. This is descriptive information and shared causal origin which had a significant influence on folk judgments of shared reference for concepts. Both factors played a significant role in the experiment. From this research it should be clear that participants' responses are sensitive to manipulations of both causal and descriptive information. Thus, both factors can play a significant role in individual reference judgments.

These empirical results have certainly a great influence on the findings of Machery et al., because according to Genone and Lombrozzo's interpretation there is no pure descriptive and causal-theory of reference. While pure distinction between theories of reference was an important part of the formers' conclusions, then it is another setback for the experimental philosophers.

4.3 Discussion about the Dual Nature of Reference

The project about referential behaviour, what experimental philosophers have run since 2004, has influenced many philosophers. Most of the philosophers who have been mentioned in this paper show readiness to accept folk in testing referential intuitions, but disagree on a second aspect which says that our intuitions vary within and across cultures. Sytsma and Livengood have successfully showed how ambiguities in a probe can influence our judgments about the Gödel's case. In addition to that, Genone and Lombrozzo have questioned the variations by

showing that variations in intuitions are due to the dual nature of reference. Both of them have good arguments against the variations in intuitions, but at the same time, they do not criticise Machery et al. on similar grounds. This aspect makes them a little bit controversial to each other. In this section I am going to argue how both accounts can find a common ground for a criticism against the empirical findings of experimental philosophy.

First of all, I will come back to the design of the Gödel's thought experiment which leaves some open questions in an interpretation of the empirical results by Machery et al. What need critical examination are two rather forced questions at the end of the probe than the probe itself. So far I have argued that hypothetical thought experiments are quite reasonable ways to get feedback from philosophers and lay people. However, while the empirical results are not unanimously understood among the philosophers who have run different experiments, it would be reasonable to seek other alternative viewpoints for coming out of this unclear situation.

Genone and Lombrozzo argue that our variations are due to the dual nature of reference. If it is so, then the design of the Gödel probe is not able to measure the nature of reference as Genone and Lombrozzo have claimed, nor are they correct the empirical findings that experimental philosophers hold. To put this into other words, the 'experimental tool' must be meant for measuring how something really is. The experimental design of experimental philosophy is not able to do so. The measurements are complicated, because the two forced questions at the end of probe would divide the referential judgment into two. A person who is going to answer to the empirical test can choose between A the person who really discovered the incompleteness of arithmetic? Or, B the person who got held of the manuscript and claimed credit for the work? If one answer corresponds to the descriptivist account and the second to the causal-historical account, then by choosing one of them a person is going to exclude another option for himself. But if both theories are legitimate in a certain way, then we cannot make such a distinction at all.

Also, Sytma and Livengood have already shown what it means for test results when different perspectives are considered. People can have similar responses to the probe, but only when all hypothetical options for interpretation have opened and clarified. This gives rise to the chance that the nature of reference is more complicated. There are lots of perspectives of how to examine a hypothetical thought experiments, and respectively, there may be several ways how to judge it. In one situation a person can strictly have descriptivist intuitions, however, in

another situation he can have totally Kripkean intuitions. Or he judges that more than one perspective seems to make sense in a limited way. Thus, I believe that the experimental philosophers need to explain how all situational perspectives and the nature of reference match together. So far the empirical results of Machery et al. will lose their significant value for the philosophy of language.

Moreover, for this kind of research as I explained above, it would be better to test our philosophical intuitions about reference not only on the Gödel case, or similar, but also on cases that are structured differently. The Gödel case bases too much on Kripke's account, which he used as contra-argument for descriptivist viewpoint. For a better overview it would also be reasonable to test a story where the causal-chain of speakers fails to give a correct referent. Manipulation with the causal chain takes place only towards the competent speakers in this story, but at the same time, it should not actually eliminate the real person of whom the speakers are talking about. This leaves an open possibility to support the causal-historical view of reference even if it is less desirable choice in a new probe. In that way we can learn something new about our philosophical intuitions. For example, below I have drawn a possible candidate for this measurement.

The Hound of Baskerville

A man called Johnson is driving across England. His route goes through the village of Baskerville. When he arrives there, he decides to stay for a longer period. On the first day in Baskerville he meets with some village people in a local pub. They warn Johnson not to go to the forest, because there is a monstrous beast there who is guilty of killing two villagers. All villagers call this unknown beast the Hound of Baskerville. After they have introduced this name for the beast, Johnson starts using it. Fortunately, there is a brilliant detective, Sherlock Holmes, who comes to the village. He would like to investigate the murder case. During his investigation, Holmes finds something extraordinary. The monstrous beast appears to be a man called Mr. Frankfurt who was wearing a costume that made him look like a monster. What is more, Mr. Frankfurt was actually guilty of killing those two villagers, not the monstrous beast as the other villagers had believed. In order to avoid undermining his subsequent investigation, Holmes does not tell the villagers or Johnson about the results of his investigation.

The Baskerville case is slightly different to Gödel's case. This probe is inspired by the famous detective story which has been modified for testing intuitions. The challenge for reference is also a little bit different. In the Gödel's case John was a more independent character than Johnson, because in this new probe, Johnson is an essential link of competent speakers. What is more, we can see how this causal-historical chain fails to give correct referent where Johnson is the last speaker in the causal chain. Now it would be interesting to see whether this semantic manipulation influences peoples' choice. The question is do they change their responses compared to Gödel's case. For instance, if they change their responses from causal-historical view to more descriptivist view of reference, then it will be a sign that the nature of reference is more complicated than experimental philosophers have expected.

4.4 Complicated Nature of Reference

The last theorisation in the previous section has motivated me to do more research in order to make sure that my assumptions about the complicated nature of reference have an empirical proof. In order to confirm my hypothesis, I decided to create a suitable strategy for measuring our nature of reference. The aim of this strategy was to measure a possible shift in given responses to the thought experiments. If people have a dualistic understanding of reference, then they must show their inclinations to both accounts. The shift in responding to the thought experiments demonstrates my conclusions well.

For this test purpose I decided to gather 40 participants from different fields of science, including philosophy. All of them had very good skills in English. I asked them to read the Gödel and the Hound of Baskerville's probes. All of them were designed so that they took into account the criticism of Sytsma and Livengood about the different perspectives to the probe. The best viewpoint for this challenge appears to be the clarified narrator's perspective, which according to Sytsma and Livengood was able to diminish variations in intuitions within culture. Having read the probes, the participants had to answer two questions. The first question controlled their general ability to understand the text. Fortunately, all of them passed

this question well, which means that they were all competent speakers to answer the second question. The second question was meant for testing their philosophical intuitions about reference respect to the clarified narrator's perspective.

According to the Gödel story, I asked the participants the following:

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?
- (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?

And when the participants read the story about the Hound of Baskerville, I asked them as follows:

Having read the above story and accepting that it is true, when Johnson uses the name 'Hound of Baskerville', would you take him to actually be talking about:

- A) A monstrous beast in the forest that allegedly killed two villagers, or
- B) Mr. Frankfurt, the man who murdered two villagers wearing a costume

Also, at the end of each thought experiment there was a scale, where participants could show their confidence to the answer they were given to the probe. On a scale from 0 to 100 they had the chance to choose the best number that indicated their confidence. The lower numbers were indicating that they were not sure and the higher numbers were indicating that they were sure.

When I started to analyse the results of my experiment, I got quite interesting outcomes. According to the answers that participants gave, it is possible to make a distinction between four distinct groups. All groups are distinguishable due to the answers they gave to the second question in the Gödel and Baskervilles cases. This means that there is a group of people who

responded A to the Gödel case and also to the Hound of Baskerville's case. I call this group as the A-A Group. Respectively there was also a B-B Group, A-B group and B-A group, where the first capital letter assigned always the answer given to the Gödel case, and the second to the Baskerville's thought experiment.

The most surprising part of this experiment is the fact how groups were proportionally divided. The results are shown in the following figure:

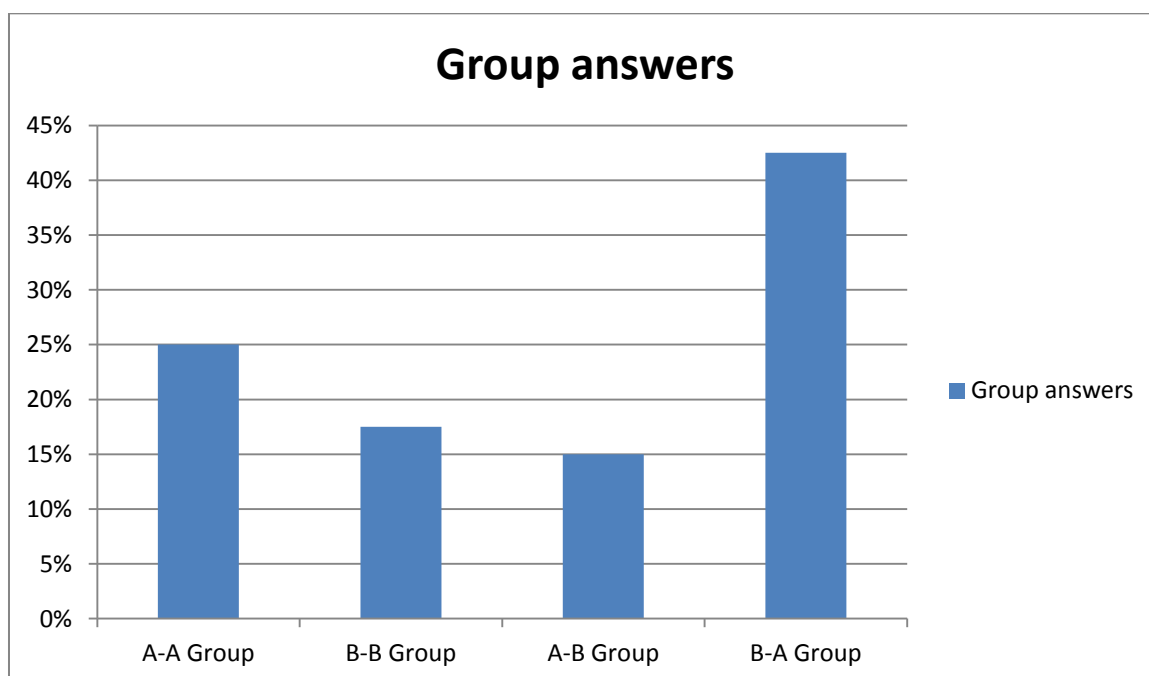


Figure7.

The results indicate that our nature of reference is more complicated than the experimental philosophers have anticipated. There are not just people who can be categorised with respect to the descriptivist or causal-historical view of reference. The complexity of intuitive thinking seems wider and maybe even random, depending largely on the different circumstances that have been taken into consideration. If my experimental research had based only the results of Groups A-A and B-B, then they would have been comparable to the results that the

experimental philosophers have found in testing East Asians. However, here we cannot ignore the groups of people, especially the B-A group, who were changing their responses during an experimental study. Their inclination to change their final judgments about the cases shows that the both theories of reference may play a significant role in our everyday cognition. There is no reason to underestimate one theory to another.

What is more, it seems that our confidence is also a small contributor in analysing the final results. People in groups B-B and A-B are more confident when they had to make their final judgment. For example, the B-B persons' average confidence was between 90-92%, they also scored more often the maximum percentages than any other group. The A-B group's average was more than 80%. Meanwhile, people who can be associated with predominantly A answers, such as A-A and B-A participants. Their average confidence was between 70-80%. Also, they seemed to give the lowest percentages. Some of them were even less certain than 50%.

In order to demonstrate what these empirical findings tell us about the big picture, then it would also be reasonable to calculate the overall score of all answers and to show how a significant shift actually has happened during the experimental research. The results are given below:

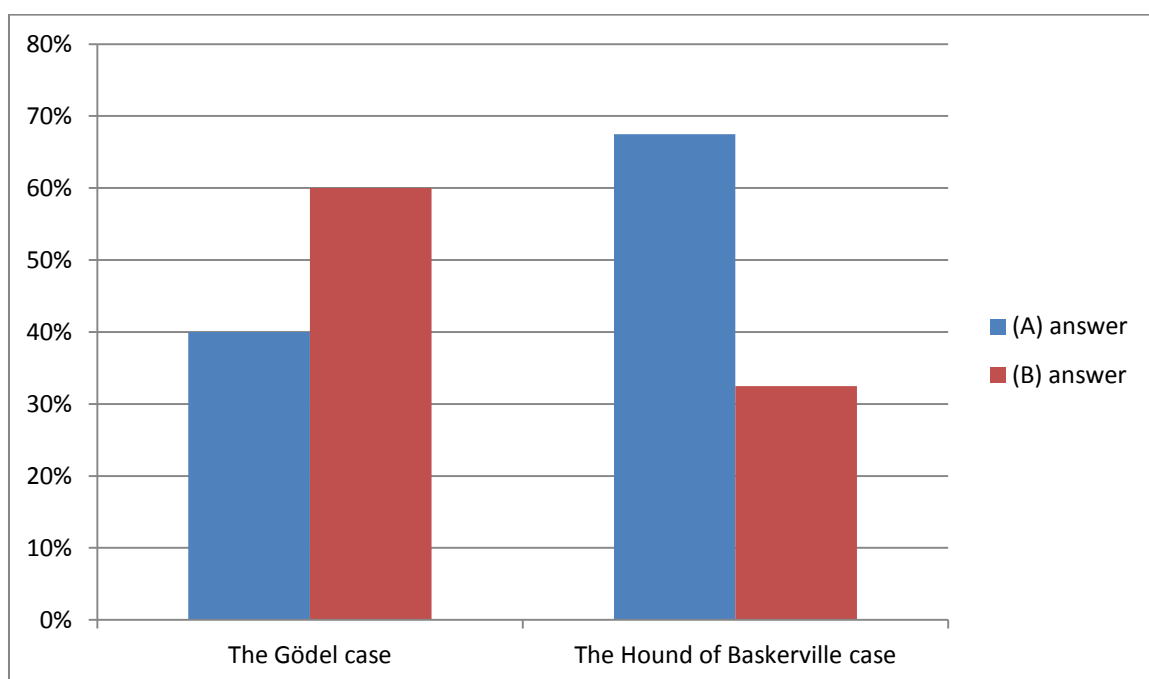


Figure8.

Figure 8 shows that according to the Gödel case the participants were mostly inclined to have B answers, which is interpreted as a support for the causal-historical view of reference. 60% of all participants found this option more favourable for them. The results changed upside down when participants were let to consider the Hound of Baskerville story. The percentage of B decreased rapidly down to 32.5% by making the A answer more favourable to the audience. Thus, my first predictions about the shift in intuitive thinking are real. People can easily change their intuitive responses to the thought experiments when circumstances are going to be changed. The manipulation with the failure of causal-historical chain in the Hound of Baskerville story is a good example of it.

Also, it should be noted that behind those numbers, there is more reason to talk about the dual nature of reference. It seems that people can sometimes have strictly descriptivist intuitions of reference or Kripkean intuition about the causal-historical view of reference, but above that, they can be influenced by the various unknown factors that will shape our decision-making process. This should be an instance of the dualistic nature of reference.

In addition to that, reading the background information of all participants, I can reject more claims from the experimental philosophers. When Sytsma and Livengood showed that there should not be any significant difference between philosophers and non-philosophers, the same I can say according to my empirical research. Philosophers and educated folk were represented in all four groups. It was really difficult to make any substantial conclusion whether or not philosophical training has some advantage. Their judgments did not differ radically from the rest of educated people who have a degree in another field of science. Nor was there any difference between males and females. At least at some point the intuitional diversity seems to be reducible. This might be due to the dual nature of reference which diminishes this kind of variations in intuitions, or makes it even less notable.

This does mean that the whole project of experimental philosophy is hopelessly wrong. Rather it should make the experimental philosophers rethink their empirical findings and to improve the design of methodology suitable for the dual nature of reference. Otherwise, the initial empirical findings of experimental philosophy lose their significant role in the philosophy of language.

Conclusion

In this paper I have argued that the method which Machery et al. use for testing referential intuitions is not appropriate. The design of the empirical method which the experimental philosophers hold is not meant to measure the effect of possible perspectives for judgment making. Nor does it take into account the fact that the nature of reference is more complicated by involving elements of both referential theories. While it is not clear how possible perspectives and ways of judging them match with each other, then we can say that the empirical findings of experimental philosophy lose their significant role in the philosophy of language.

However, this does not mean that the whole empirical research of experimental philosophy is on the wrong track. Hypothetical thought experiments still seem to be attractive tools for testing our referential intuitions by asserting philosophers and non-philosophers in a hypothetical test situation. In addition to that, some philosophers, such as Michael Devitt, think that lay speakers must not be involved in tests, because their lack of expertise and therefore their intuitions should not matter. I have shown that this claim is wrong. These kinds of thought experiments that were presented to participants were not highly technical and sometimes these events that were ascribed in thought experiments can happen in everyday life as well. Also, Sytsma and Livengood have shown that philosophers and non-philosophers are able to answer in a similar way. Thus, there is no reason to make a difference among the participants.

What seems to be problematic is that the probes can be sometimes ambiguous. And those ambiguous perspectives can have an effect on our judgments. Sytsma and Livengood tested perspectives separately and they were able to correct the empirical findings of experimental philosophy so that Western participants were mostly inclined to have Kripkean intuitions and East-Asians showed their stronger inclinations to the descriptivist account. Taking different perspectives into account, Sytsma and Livengood were able to show that intuitions do not vary within cultures. Cross-cultural variations still remained unexplained.

Those clarified findings have made me ask whether we have understood the real nature of reference. Genone and Lombrozzo made several experiments in order to show that variations in intuitions are due to the dual nature of reference. If the real nature of reference is dual, then it will be another setback for the design of experimental methodology because, choosing one answer out of two, the participant excludes a chance to show their willingness to support the other in some circumstances. In that way we cannot measure how much of his support goes to the descriptive or causal-historical view of reference.

While the design of experimental methodology, as Machery et al. and the experimental philosophers maintain, has not been able to measure is the influence of ambiguous perspectives and the dual nature of reference. This certainly reduces the significance of their initial empirical findings for the philosophy of language.

References

- Alexander, Joshua 2012. Intuitional Sensitivity. *Experimental Philosophy: An Introduction*. Polity Press.
- Devitt, Michael 2011. Experimental Semantics. *Philosophy and Phenomenological Research*. Vol. LXXXII No. 2, p. 418-435.
- Devitt, Michael 2011. Whither Experimental Semantics? *Theoria* 72: 5-36.
- Devitt, Michael 2012. Semantic Epistemology: Response to Machery. *Theoria* 74: 229-233.
- Evans, G. 1973. The Causal Theory of Names. *Proceedings of the Aristotelian Society*. Supplementary Volumes, 47, p. 187-208.
- Frege, G 1892/1943. On Sense and Reference (M. Black, trans.) *The Philosophical Review* 57, 207-230. Reprinted in Ludlow (1997).
- Genone, J 2012. Theories of Reference and Experimental Philosophy. *Philosophy Compass* 7/2, p. 152-163.
- Genone, J. and T. Lombrozo 2012. Concept Possession, Experimental Semantics, And Hybrid Theories of Reference. *Philosophical Psychology*. Vol.25, No. 5, 717-742.
- Jackson, F 1998. Reference and Description Revisited. In Tombulin (Ed) *Language, Mind and Ontology* (12) pp. 201-218.
- Kripke, S 1972/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, David 1970. How to Define Theoretical Terms? *Journal of Philosophy*, 67, 427-446.
- Machery, E., Mallon, R., Nichols, S., and Stich, S. 2004. Semantics Cross-Cultural Style. *Cognition* 92: B1-B12.
- Machery, E 2011. Expertise and Intuitions about Reference. *Theoria* 72: 37-54.
- Mallon, R., Pinillos, A., Nichols, S. 2013. Ambiguous Reference. Manuscript in Preparation.

- Nisbett, R. E. 2003. *The geography of thought: how Asians and Westerners think differently ... and why*. New York: Free Press.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. 2001. Culture and systems of thought: holistic vs. analytic cognition. *Psychological Review*, 108, 291–310.
- Searle, John 1958. Proper Names. *Mind*, 67, 166-173. Reprinted in Ludlow (1997).
- Stich, S 1988. Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity. *Synthese* 74, 391-413.
- Stich, S 1990. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press.
- Stich, S., Machery, E. 2011 forthcoming.
- Sytsma, J. and Livengood J. 2011. A New Perspective Concerning Experiments on Semantic Intuitions. *Australasian Journal of Philosophy* 89, p. 315-332.
- Weinberg, Jonathan 2007. How to Challenge Intuitions Empirically Without Risking Skepticism? *Midwest Studies in Philosophy* XXXI Blackwell Publishing, pp. 318-343.
- Zamzow, J. and Nichols, S. 2009. Variations in Ethical Intuitions. *Philosophical Issues* 19, p. 368-388.

Abstract

This thesis critically examines the empirical findings of experimental philosophy in the philosophy of language. I will cast doubt on the findings of experimental philosophers about our referential intuitions. Mostly I criticise them on two different grounds. Firstly, the design of the probe allows too many variations in the test, which can have a direct effect for the final results. Secondly, the experimental philosophers have underestimated the role of the dual nature of reference when they do experiments. People, however, can respond differently in respect to the circumstances. There is no dominant strategy for judging reference. In this thesis there is also a new thought experiment that I am going to use in order to illustrate my claims.

Intuitsioonide testimine osutuse kohta: missuguseid järeldusi saame teha eksperimentaalsetest andmetest?

Minu magistritöö teemal “Intuitsioonide testimine osutuse kohta: missuguseid järeldusi saame teha eksperimentaalsetest andmetest?” uurib kriitiliselt eksperimentaalfilosoofide empiirilisi tulemusi referentsiaalsete intuitsioonide testimisel keelefilosoofias. Ma leian, et need tulemused, mida on saanud eksperimentaalfilosoofid oma uuringute jooksul ei ole piisavalt adekvaatsed. Ma kritiseerin neid peamiselt kahest vaatenurgast. Esiteks sisaldab testimiseks kasutatav mõtteeksperiment liialt palju varieeruvusi, millel on otsene mõju lõpptulemuste jaoks. Teiseks ei ole eksperimentaalfilosoofid suutnud arvestada mõjudega, mida võib tuua duaalne osutamiski eksperimentide läbiviimisel. Leian, et inimesed võivad siiski osutada erinevalt sõltuvalt situatsioonist, kuhu nad on asetatud. Seega, ei ole olemas peamist strateegiat, kuidas inimesed saavad otsustada osutamise üle. Lisaks on selles magistritöös esitatud ja läbi viidud uus mõtteeksperiment, mis toetab minu seisukohti.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Henri Zeigo

(sünnikuupäev: 01.04.1989)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

magistritöö teoreetilise filosoofia erialal pealkirjaga „Testing Intuitions about Reference: What conclusions should be drawn from the experimental data?“ (eesti keeles „Intuitsioonide testimine osutuse kohta: missuguseid järeldusi saame teha eksperimentaalsetest andmetest?“),

mille juhendaja on Prof. Dr. Daniel Cohnitz,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 30.04.2014