# Experimental Philosophy and the Philosophy of Language

# A Defence of the Ambiguity Thesis

By

**Katharine Szondy Hamilton** 

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

VICTORIA UNIVERSITY OF WELLINGTON

#### Abstract

In this thesis I employ the experimental method to inform three important debates within the philosophy of language. These three debates can loosely be characterised as the following: Strawsonianism vs. Russellianism about the meaning of definite descriptions (Chapter 2), Millianism vs. Descriptivism about the meaning of proper names (Chapter 3), and Internalism vs. Externalism about natural kind terms (Chapter 4). To investigate these debates I use surveys to test the intuitions of ordinary language users, that is, nonphilosophers, about the meaning of various terms and phrases in natural language. This included New Zealand undergraduate students, students in China, and participants in the US in order to investigate any cross-cultural differences. The results of these three studies indicate substantial variation in the intuitions held among ordinary language users. I use this variation to defend an ambiguity thesis. According to this thesis, some terms and phrases as they occur in natural language (specifically, proper names, natural kind terms, and definite descriptions) have multiple meanings associated them. No one disambiguation is correct outside of a context of utterance. If the ambiguity thesis is accepted, various philosophical puzzles disappear. I will also address a number of objections that face the general program of this thesis.

### Acknowledgments

I would like to thank a number of people who helped and supported me throughout the process of producing this work. First and foremost I would like to thank my supervisor Stuart Brock whose continual support and guidance saw me through this project. In particular, Stuart's generosities with his time and his knowledge were invaluable to me and have contributed hugely to the betterment of this work. I would also like to thank my secondary supervisor Justin Sytsma whose thorough comments on drafts and various discussions is greatly appreciated.

Many thanks also go to various friends and colleagues from the department for their support over the last three years. Being surrounded by such an incredible group of people has made this project much more enjoyable and fruitful. I would finally like to thank my family and friends for their continual care and words of encouragement.

# Contents

Chapter 1: Experimental Philosophy and the Philosophy of Language	11
1.1 Introduction	11
1.2 The Ambiguity Aversion	14
1.3 The Objection Against Ordinary Language Philosophy	20
1.4 The Objection from the General Methodological Problems of Surveys	27
1.5 The Expertise Objection	31
1.6 Present Thesis	53

Chapter 2: Definite Descriptions and the Ambiguity Thesis	55
2.1 Introduction	55
2.2.1 Russell's theory of descriptions is false as applied to ordinary language	56
2.2.2 Experimental philosophy	62
2.2.3 Study 1: The Basic Survey	67
2.3.1 Definite descriptions are ambiguous between at least three readings	74
2.3.2 Study 2: The Conceptual Survey	81
2.3.3 Summary so far	91
2.4.1 This ambiguity is a semantic rather than pragmatic ambiguity	92
2.4.2 Study 3: The Implicature Survey	99
2.4.3 Study 4: The Default Meaning Survey	106
2.5.1 General discussion	113
2.5.2 Possible alternative explanations	120
2.5.3 Conclusion	123

Chapter 3: Revisiting Cross-cultural Semantics About Proper Names	125
3.1 Introduction	
3.2 Experimental philosophy	
3.3 Present studies: Studies 5 and 6	
3.4 Cognitive reflexivity	140
3.5 Study 5: Experimental method	

3.6 Experimental results	144
3.7 Study 6: US sample follow up	148
3.8 Discussion	150
3.9 Defence of intuitions	151
3.10 The Ambiguity Thesis	153
3.11 Problems with traditional theories	156
3.12 More theories of meaning	163
3.13 The semantic pragmatic distinction	165
3.14 Conclusion	166

# Chapter 4: An Experimental Perspective on Semantic Externalism About Natural Kind

Terms	
4.1 Introduction	
4.2 Meta debate	
4.3 Intuitions about natural kind terms	
4.4 Philosophical bullying and philosophical filtering	
4.5 Experimental philosophy	
4.6 Present studies: Studies 7 and 8	
4.7 Study 7: Experimental method	
4.8 Experimental results	
4.9 Study 8: US sample follow up	
4.10 Discussion	
4.11 Philosophical bullying/filtering	210
4.12 Alternative explanations	211
4.13 Conclusion	223

Chapter 5: Concluding Remarks	
5.1 Concluding remarks	

References	231
Appendices	249

## Chapter 1: Experimental Philosophy and the Philosophy of Language

#### 1.1 Introduction

Experimental philosophy is an approach to philosophical questions that has become popular over the last decade. Rather than relying solely on the *a priori* methods utilised by philosophers in the twentieth century, experimental philosophers advocate the use of methods borrowed from psychology and cognitive science to empirically investigate philosophical questions. By carefully examining the intuitions of non-philosophers, or 'the folk,' this methodology is able to inform on important philosophical debates as it offers a new source of evidence. Among other things, this kind of data can challenge the status of commonly held philosophical opinions held by philosophers working in the field.<sup>1</sup>

The aim of this thesis is to defend and employ the experimental method in order to inform three ongoing debates within the philosophy of language.<sup>2</sup> These three debates can loosely be characterised as the following:

- 1. Strawsonianism vs. Russellianism about the meaning of definite descriptions
- 2. Descriptivism vs. Millianism about the meaning of proper names
- 3. Internalism vs. Externalism about the meaning of natural kind terms

The focus on these particular debates is for the following reasons: they are some of the most important debates in the history of analytic philosophy, as yet they remain unresolved and, they have only recently had the experimental method applied to them. Although there has been a large amount of experimental work conducted within the philosophy of language literature, most of this work has not focused directly on meaning but has focused instead on reference. While I presuppose an intensional semantics—and therefore that meaning

<sup>&</sup>lt;sup>1</sup> This is one common project of experimental philosophy. For a discussion on some of the other projects see Nadelhoffer & Nahmias (2007).

<sup>&</sup>lt;sup>2</sup> One note on terminology: throughout this thesis I use 'experiment' to mean any kind of empirical study; mostly I mean survey studies. In a technical sense 'experiment' might be restricted to studies that include a control group and the studies I will present and discuss would be regarded as empirical rather than experimental. However, I will follow previous experimental philosophers who have adopted the term 'experiment' in a loose sense to mean the kinds of studies I will be discussing (e.g., Jyllkä et al., 2009; Machery et al., 2004). Thus, throughout this thesis I will use 'experiment' and 'study' interchangeably.

determines reference in all possible worlds— the two notions can come apart. In particular, it is possible for terms to have different meanings and yet have the very same extension. These cases are in many ways the most interesting but have been largely ignored in the experimental literature. What I aim to show is that the experimental studies have left a lot of room for further investigation. Consequently, the experiments I will present in this thesis are able to provide new evidence and advance the debates listed above.

Specifically, I will use the experimental method to test the intuitions of ordinary language users, namely New Zealand undergraduate students, students in China, and participants in the US. What I will argue is that the experimental results indicate that definite descriptions, proper names, and natural kind terms are all semantically ambiguous. To this end, I will propose and defend what I call 'The Ambiguity Thesis' — that is to say, the thesis that some referring terms and denoting phrases in language have more than one possible meaning associated with them, and which meaning they have will depend on their context of utterance. According to the Ambiguity Thesis, a native speaker can use these terms and phrases with any of the semantic contents available; no one disambiguation is correct outside of a context of utterance.

My PhD thesis will be structured as follows: in Chapter 2 I will present Studies 1-4 which examine folk intuitions about definite descriptions. Specifically I will look at whether participants have intuitions consistent with a Russellian theory, according to which a definite description is a special kind of general proposition, or intuitions consistent with the Frege-Strawson Theory, according to which definite descriptions are directly referential expressions or singular terms. The data of four experiments, I argue, are best interpreted as indicating that definite descriptions are ambiguous between at least three different readings. Moreover, two of these experiments (Studies 3 and 4) suggest that this ambiguity is not easily explained in pragmatic terms but are indicative of a genuine semantic ambiguity. Here I will introduce and defend the Ambiguity Thesis more fully. This chapter is a novel departure from previous experimental work in the area. In fact, discussion of definite descriptions is so

sparse that in a recent collection of papers titled *Advances in Experimental Philosophy of Language* (2015) there is no discussion of definite descriptions.<sup>3</sup>

I follow up this study in Chapter 3 with Studies 5 and 6 designed to test whether participants have Millian or descriptivist intuitions about proper names. That is, whether participants consider the semantic value of a proper name to be its actual referent, or whether they consider the meaning to be synonymous with something like a definite description. This experiment is the most similar to the bulk of previous experimental work within the philosophy of language, which has indicated large amounts of variation in the intuitions of ordinary language users about the reference of proper names (for example see, Machery et al. 2004). What these previous experiments have suggested is that Kripkean intuitions are not as universal or widely shared as it has been commonly assumed by philosophers. The present study is based on this research but instead of focusing on cases where intuitions about the reference of a given proper name diverge, this study will focus instead on intuitions about the meaning of proper names, even in cases where the reference does not change. Again, the results indicate support for the Ambiguity Thesis which would also explain the variability previously found in intuitions about reference.

Lastly, in Chapter 4 I will present Studies 7 and 8 which test whether participants have Internalist or Externalist intuitions about natural kind terms. These experiments also provide support for the Ambiguity Thesis. Here I will also consider whether global divergence but local convergence in philosophers intuitions about this issue are the result of a kind of 'philosophical bullying' by local authorities, which shapes intuitions through the course of their study, discussions, etc.

Chapter 5 is a brief conclusion that considers further experimental research that might be done to extend the findings in this thesis.

Before presenting these three studies I want to address a number of objections that face the general program of the present thesis. Firstly, the Ambiguity Aversion is a general objection based on a bias against the endorsement of ambiguity theses in philosophy. Secondly, the Objection Against Ordinary Language Philosophy is an objection to the informativeness of

<sup>&</sup>lt;sup>3</sup> See Jussi Haukioja (Ed.) (2015): Advances in Experimental Philosophy of Language.

the ordinary language program in philosophy, a program partially relied on in experimental philosophy. Thirdly, the Objection Against the General Methodology of Surveys is a collection of concerns about the use of surveys to gather data about intuitions. Lastly, the Expertise Objection is an objection to the general program of Experimental Philosophy made by philosophers who challenge the informativeness of this approach to philosophy. Each of these objections calls into question the significance of the present program. For this reason the aim of this chapter is to respond to each objection in turn.

#### **1.2 Ambiguity Aversion**

The Ambiguity Aversion objection is a general objection to the postulation of ambiguities in the philosophy of language and is therefore a challenge to the thesis I will be advocating, that is, the Ambiguity Thesis. At bottom, it is a (perhaps reasonable) bias against the postulation of ambiguities, but we can put the objection in the form of an argument as follows:

- (1) A philosophical theory about the meaning and semantic value of types of terms and phrases in a language may or may not postulate an ambiguity
- (2) Any philosophical theory about the meaning and semantic value of types of terms and phrases in a language that does not postulate an ambiguity (a univocal theory) is to be preferred, all else being equal, to any theory that does postulate an ambiguity (an equivocal theory)
- (3) Therefore, all else being equal an ambiguity thesis in philosophy should be rejected in favour of a univocality thesis

Many philosophers are sympathetic to this kind of argument and are therefore suspicious of ambiguity theses. In particular the postulation of ambiguities has been very controversial within the philosophy of language literature. For example, we can see this come out in the literature about definite descriptions. For while there are some philosophers who take even *simple* sentences containing definite descriptions to be ambiguous (e.g., Amaral, 2008; Devitt, 2004; Reimer, 1998; Wettstein, 1980), the overwhelming majority of philosophers

reject such a hypothesis in favour of a 'no-ambiguity' account (e.g., Blackburn, 1986; Schoubye, 2010; von Fintel, 2004; Zacharska, 2010). I will discuss this specific debate in Chapter 2.

Here I want to note that I too am sympathetic to this argument. However, I think attention needs to be drawn to (2), and in particular to the phrase "all else being equal". When I present an ambiguity thesis it is because all else is not equal, but that we have good reason to think an ambiguity is present. The aim of chapter 2 will be to make good on this claim.

The Ambiguity Aversion objection was made famous by Saul Kripke (1977) in his famous response to Donnellan (1966).

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

When Kripke makes this comment it is in relation to Donnellan's proposed distinction between attributive and referential uses of a definite description. Kripke argues that the distinction is not the result of a semantic ambiguity in sentences containing definite descriptions since we can account for the very same phenomenon by appealing to pragmatic considerations; that is, we can account for the two uses as a difference between speaker's reference and not a difference in semantic reference.<sup>4</sup>

Here I think Kripke is right. However, his comments are sometimes taken as a dismissal of ambiguity theses in general. Kripke, though, is more nuanced than this, indicating a preference for univocal theories, allowing that there may be trumping evidence in support of ambiguity theories, and putting the onus on the ambiguity theorist to provide that evidence.

This last comment raises the question as to what evidence would warrant the postulation of an ambiguity. Indeed, the reason for such controversy about the postulation of ambiguities

<sup>&</sup>lt;sup>4</sup> Kripke's argument against postulating a semantic ambiguity in sentences containing definite descriptions will be discussed in depth in Chapter 2.

in philosophy arises from the very fact that it is notoriously hard to determine when genuine ambiguities exist in language. For this reason a number of linguistic tests have been designed to test for ambiguities in language (see Sennet, 2015; Zwicky & Saddock, 1975). To give some examples:

*The Contradiction Test*: Use the possibly ambiguous term or phrase in a syntactically contradictory sentence and see whether there is a non-contradictory interpretation. For example, 'a bank is not a bank' can be interpreted in a non-contradictory way indicating that 'bank' is semantically ambiguous (Machery, 2014).

*The Conjunction Reduction Test*: Take two sentences containing the possibly ambiguous term, for which each sentence uses the term with a different meaning, and conjoin them so that the possible ambiguity is only included once. If the conjoined sentence feels zeugmatic, that is, sounds strange due to the term being exploited for multiple meanings, then there is evidence of ambiguity. For example, the conjoint of the following sentences:

- (1) The colours are light.
- (2) The feathers are light.

produces the zeugmatic sentence:

(3) The colours and feathers are light.

which indicates that 'light' is semantically ambiguous.

*Ellipsis* Test: Similarly to the conjunction reduction test, conjoin two sentences containing the possibly ambiguous term so that the second instance of the term is replaced by 'too'. For example:

(3\*) The colours are light and the feathers are too.

This should force a single interpretation of the possible ambiguity. Any sense of zeugma indicates an ambiguity.

*Disjunction*: Put the possibly ambiguous term into different contexts and see whether it forces a single interpretation. If the term only expresses one meaning in each context rather than either meaning in all contexts, then it is ambiguous. For example:

- (4) The man deposited money into his bank account.
- (5) The woman had a picnic down on the bank.

Here we see that 'bank' is ambiguous because in each context, 'bank' *either* means 'money lending institution' *or* it means 'side of a river'. In neither (4) nor (5) does 'bank' mean 'either money lending institution or side of a river' (see Roberts, 1984; Zwicky & Saddock, 1987).

Inconstancy Under Substitution: Substitute the possibly ambiguous term with different synonyms and determine whether the meaning is preserved. For example, if 'cardinal' is not an ambiguous word and is synonymous with the term 'Papal dignitary', then 'Papal dignitary' should be substitutable for it in all extensional contexts without changing its meaning. But it is not. '1 is a cardinal (number)' clearly does not mean the same as '1 is a Papal dignitary'. So either 'Papal dignitary' isn't a true synonym or (more likely) 'cardinal' is ambiguous.

For our present purposes, however, there are problems with these tests which mean they are not viable tests to determine whether an ambiguity is really present. For one, these tests mostly only attempt to distinguish semantic ambiguities from under-specification or vagueness. However, this is not what we need the tests to do. For our purposes we need tests that can determine whether any ambiguity is a genuine semantic ambiguity or whether it can be given a pragmatic explanation. This is exactly what these tests cannot do. For instance, there might be cases in which the Contradiction Test indicates an ambiguity, but the ambiguity is not a semantically significant one. It might be that we try to interpret the sentence as not contradictory even if, strictly speaking, it is. Machery found that participants were willing to endorse seemingly contradictory sentences like 'in a sense tomatoes are vegetables' and 'in a sense tomatoes are not vegetables' about half the time. Moreover, even without the hedge 'in a sense' participants' still endorsed each of these kinds of sentences more often than when one of the sentences were obviously false (e.g.,

'lions are not animals'). However, this does not show that 'tomatoes' is an ambiguous term; just that we can if needed interpret it in a non-contradictory (perhaps figurative) way. That is, we can give a non-contradictory interpretation pragmatically. If these tests were applied in the following experiments to determine whether the terms and phrases of interest are semantically ambiguous there would be no way to rule out the possibility of a pragmatic explanation of the data. For example, if participants were presented with seemingly contradictory sentences about definite descriptions, such as 'the King of France is bald' and 'the King of France is not bald', endorsement of both sentences would not be evidence that definite descriptions are semantically ambiguous.

Furthermore, these tests are only positive tests for ambiguity. It could be that some words or phrases are genuinely ambiguous but are not picked up with these tests. This in fact is very likely in cases of polysemy. Polysemy is a type of ambiguity where the different meanings have a very close conceptual relationship. While paradigm cases of ambiguity involve homonymous words-- single word types that have multiple but conceptually distinct meanings—in polysemes the different meanings are conceptually very similar. For example, 'the USA' can mean a land mass, or a population of humans, or an economy.<sup>5</sup> Each associated meaning is only subtly different from the others, which is what makes it a case of polysemy, the ambiguity might be too subtle to register with these tests. Zeugma in particular is very subtle and may not be consciously registered. So, for example, the following two sentences are clearly true:

(6) The USA voted in its first African-American President in 2008.

(7) The USA is a part of the North American continent.

But the following conjunctive contraction:

(8) The USA, which is part of the North American continent, voted in its first African-American President in 2008.

<sup>&</sup>lt;sup>5</sup> Lycan in 'Metaphysics and the Paronymy of Names' (unpublished manuscript) gives a similar example about 'Australia'. He argues that proper names are 'paronymous' (essentially polysemous) and that this is especially apparent in the case of fictional characters. For example, 'Romeo' refers to either the character or the non-actual corresponding person.

might not seem zeugmatic unless the reader was alert to the possible ambiguity.

Sennet (2002) uses the Conjunction Reduction Test to argue that definite descriptions are not semantically ambiguous because conjunctions of sentences using 'the' in purportedly different ways do not yield any sense of zeugma.<sup>6</sup> However, as I have just illustrated, this is not good evidence that definite descriptions are univocal because any ambiguity that exists would quite likely be very subtle. Indeed, applying *any* of these tests in the present experiments would not rule out the possibility that the terms and phrases are ambiguous because they are examples of polysemy.

When I argue for the Ambiguity Thesis it is because I believe there is good reason to think that a genuine ambiguity is present. I will contend that the data I present is not better accounted for by other explanations. Because I think the ambiguity in each instance is a case of polysemy, I will not rely on the standard linguistic tests to provide evidence of ambiguities in the relevant terms. Instead, I will use some different measures. In the case of definite descriptions, two experiments directly attempt to rule out any possible pragmatic interpretation of the results. One experiment tests participants' competency to correctly identify cases of entailment and cases of mere implication (or implicature). Only participants who demonstrated a competency with the distinction were tested on cases of definite descriptions. In this way the experiment distinguished what participants considered a definite description to *mean* and what a definite description *merely implies*. The other experiment asks what participants take the meaning of definite description to be when it is taken completely out of context. By removing contextual cues—including information about the speaker—the experiment focuses on semantic meaning and not speaker meaning.

For natural kind terms, pragmatic responses were controlled for by providing participants with sufficient information to distinguish between speaker and semantic meaning and to direct their responses appropriately. This was not done in previous studies such as Jyllkä et al. (2009) for which the simple 'yes' and 'no' answers did not distinguish between semantic meaning and speakers' meaning. By providing more detailed responses, this experiment was better able to distinguish between the two. For proper names, speakers' meaning was controlled for by asking participants' whether two agents mean the same thing despite the

<sup>&</sup>lt;sup>6</sup> Sennet's (2002) use of the conjunction reduction test will be discussed in greater depth in Chapter 2.

agents having different beliefs (i.e., the agents would not think they meant the same thing). Previous experiments about proper names (e.g., Machery et al., 2004) asked about the reference rather than the meaning of a proper name, which is also open to pragmatic concerns. By asking specifically about semantic meaning and requiring them to distinguish it from speaker meaning, most of these concerns are dispelled. Moreover, a control case was included, for which there was a 'right' answer, which checked that participants understood the questions.

The considerations made in each of these three studies strongly support the presence of a genuine ambiguity which fulfils Kripke's criteria of having compelling grounds to postulate an ambiguity thesis.

#### 1.3 The Objection Against Ordinary Language Philosophy

The Objection Against Ordinary Language Philosophy challenges the informativeness of experimental philosophy because it works within the general program of ordinary language philosophy, a program which is widely considered refuted. What this objection claims is the following:

- Experimental philosophy relies on the general program of ordinary language philosophy
- The program of ordinary language philosophy is not able to inform traditional philosophical questions
- Therefore, experimental philosophy is not able to inform traditional philosophical questions

The argument, I maintain, is unsound. It is unsound because both of its premises are false. Let me start with (1). While experimental philosophy has some points in common with ordinary language philosophy, it distances itself from the ordinary language programme at exactly those points where it is most problematic. Ordinary language philosophers, such as Ryle, Austin, Strawson, and the later Wittgenstein, believe along with experimental philosophers that by paying careful attention to everyday uses of language that we can gain philosophical insights. To this end, ordinary language philosophers and experimental philosophers alike take it as important to consider what the folk would say. However, to find out what the folk would say, ordinary language philosophers tended to rely on their own individual intuitions from the armchair about what that would be.<sup>7</sup> Experimental philosophers recognize that this anecdotal methodology is deeply problematic, and instead look to a wider more representative population to uncover the actual intuitions of ordinary language users. That is, rather than consider what the folk *would* say, experimental philosophers empirically test what the folk *do* say.

Peter Strawson (1950) gives a particularly pertinent illustration of the ordinary language methodology that experimental philosophers repudiate. Regarding definite descriptions, he says the following:

Now suppose some one were in fact to say to you with a perfectly serious air: "The king of France is wise ". Would you say, "That's untrue"? I think it's quite certain that you wouldn't. [p. 330]

While Strawson infers this is the case, the experimental philosopher would actually test it. Indeed, it is this claim that I will test in the next chapter.

Despite their methodological differences, though, I do take it that experimental philosophy and ordinary language philosophy are inextricably tied. Both traditions look to the linguistic practices and intuitions of natural language speakers for evidence that informs philosophical questions. I therefore take experimental philosophy to be something of an extension to the method of the ordinary language project, and an improvement on its methods for uncovering the kind of evidence both traditions view as informative. But if that's right, there must also be something wrong with (2). And indeed there is.

<sup>&</sup>lt;sup>7</sup> With the possible exceptions of Tennessen (1960) and Naess (1953; 1966). In Chapter 2 I will discuss a possible experiment conducted by Tennessen about definite descriptions. There are also indications that Naess used surveys to investigate ordinary language use; for example, see the following quote from Naess's obituary: "Working in teams, the Oslo School's adherents used questionnaires to investigate the meanings that ordinary people assigned to terms like "truth," "free enterprise" and "democracy"

<sup>(</sup>http://www.nytimes.com/2009/01/15/world/europe/15naess.html?\_r=2&emc=eta1&).

Proponents of the Objection to Ordinary Language Philosophy argument take it that (2) is true because ordinary language philosophy has been refuted, and not practiced anymore. According to this accepted mythology, there was a kind of war between ideal, or logical, language philosophers and ordinary language philosophers on the other. Sometime in the 1960s (so the story goes) ideal language philosophy won a decisive victory. Experimental philosophy can, it is supposed, be seen as a return to the bad old days when people thought ordinary language philosophy could inform on philosophical questions. Ideal language philosophers criticise the ordinary language project for the reason that ordinary language is often vague and imprecise, or ambiguous. Philosophical problems though require precise and unambiguous solutions. Russell in particular is noted for criticising philosophers who consider ordinary language to be philosophically insightful as he comments:<sup>8</sup>

Common speech is full of vagueness and inaccuracy. .... Everybody admits that physics and chemistry and medicine each require a language which is not that of everyday life. I fail to see why philosophy, alone, should be forbidden to make a similar approach towards precision and accuracy [Russell (1959) p. 178]

For this reason ideal language philosophers have worked with a constructed language which gets rid of the ambiguities in natural language, regiments the language, and makes it more precise. Quine (1960) in *Word & Object* makes this clear as he discusses the regimentation of language when we move from ordinary language to logical notation:

Hence to paraphrase a sentence of ordinary language into logical symbols is virtually to paraphrase it into a special part still of ordinary or semi-ordinary language; for the shapes of the individual characters are unimportant. So we see that paraphrasing into logical symbols is after all not unlike what we all do every day in paraphrasing sentences to avoid ambiguities. ...If we paraphrase a sentence to resolve an ambiguity, what we seek is not a synonymous sentence, but one that is more informative by dint of resisting some alternative interpretations. Typically, indeed, the paraphrasing of a sentence *S* of ordinary language into logical symbols will issue in substantial divergences. Often the result *S'* will be less ambiguous than *S*, often it will have truth values under circumstances under which *S* has none, and often it will even provide explicit references where *S* uses indicator words [Quine (1960) p. 159-160]

Here we see Quine note the respective differences between ideal and ordinary language, one being the elimination of ambiguities.

<sup>&</sup>lt;sup>8</sup> Thanks to Hanfling (2000) for this reference.

The real fall into disrepute of ordinary language philosophy is widely assumed to have happened from the 1960's because of (1) the establishment of Gricean conversational implicatures, and (2) the development of systematic semantic theory (see Hansen, 2014). Soames' *Philosophical Analysis in the Twentieth Century* (2009) in particular is famous for claiming that ordinary language philosophy died with the Gricean distinction between what is meant in conversation and what is literally said. This is because, he argues, what we would or would not say might be an indication of a pragmatic rule of language use rather than an indication of a term's meaning, and as such we should not take ordinary language to be indicative of the literal meaning of what is said. For this reason, Soames claims that we need to drop the method of examining ordinary language use.

Weatherson (2007), however, has challenged Soames' representation of the history of ordinary language philosophy. For one, Weatherson argues that it is not clear why we should think that Grice overthrew ordinary language philosophy when, (1) ordinary language philosophers like Austin recognised the very distinction that Grice points out between speaker's meaning and semantic meaning, and (2) there's nothing inconsistent between the systematic semantics or ideal language project and the ordinary language philosophy project. Weatherson notes:

[T]here are plenty of very systematic formal semanticists who take Strawson's work on descriptions seriously, and try and integrate it into formal models. So we might wonder why Grice's work shouldn't have led to a kind of ordinary language philosophy where we paid more careful attention to system-building. [p. 432]

What we see is that it is not clear exactly what the challenge to ordinary language philosophy is, or, incidentally, whether ordinary language philosophy is actually refuted. Therefore there is not a clear reason to accept (2).

Moreover, despite no longer openly claiming to be ordinary language philosophers, many philosophers implicitly reject (2) and continue to work within the framework of ordinary language philosophy, seeing the merit it in. For example, we see Strawson (1950) respond to Russell's challenge that his theory of descriptions was not intended (as Strawson supposes) to be an analysis of definite descriptions as used in everyday language. In response, Strawson says:

I think it is true to say that Russell's Theory of Descriptions ... is still widely accepted among logicians as giving a correct account of the use of such expressions in ordinary language. [1950, p. 321]

Strawson, then, can be thought of as responding to philosophers who think of the theory of descriptions as applying to ordinary language even if he is not objecting to Russell himself. Indeed, there are many such philosophers. For example, immediately after Stephen Neale (1990) spells out what Russell's theory of descriptions is, he explains his project as follows:

Our task now is to spell out exactly how and when the formal statement of the Theory of Descriptions is to be applied to sentences of *natural language*. [p. 33, my emphasis]

Here is another example. In the mid-1990's a debate raged about who first discovered that proper names are rigid designators. Traditionally it was thought to be Kripke (1980), but at the time some thought it was in fact Ruth Barcan Marcus in earlier writings and that she had not received adequate credit for the discovery. Weighing in on this dispute, John Burgess defends Kripke's claim (in part) by observing that:

Marcus' well-known early work on axiomatic systems of modal logic, in Marcus (1947) and companion papers, is only indirectly relevant to the present study, since it was artificial rather than natural languages that were in question there. [1996, p.2]

Commenting on more directly relevant work of Marcus's, Burgess says the following:

As the foregoing quotations show, part of what Marcus assumes about 'names' is clearly not intended to hold literally for names in the ordinary sense. This circumstance raises doubts as to how much of the rest of what Marcus says is intended to apply to names in the ordinary sense, in natural language. [1996, p.2]

So if Burgess is right, while Marcus and Kripke both conceived of names as rigid designators, Marcus was engaged in the ideal language project and Kripke was engaged in the ordinary language project. Kripke's insight, that supposedly set him apart from Marcus, was that proper names *in natural language* are rigid. My purpose here is not to adjudicate between the two sides of this debate about the history of philosophy. Rather, it is to emphasise what was at stake, namely a thesis about ordinary language and not about an ideal or artificial language. It is here that the important work in philosophy is supposed to have been done.

Kripke's ordinary language approach is also seen in his treatment of natural kind terms, which we also see in Putnam (1973, 1975). Putnam is clearly concerned with what we would and would not say, which is reflected in his Twin Earth example; see for example:

[I]n this example, one would *not* say "on Twin Earth the meaning of the word 'water' is XYZ" [Putnam (1975) p.140].

Moreover, Kripke maintains his emphasis on ordinary language use as we him extend his previous work in the recently published Reference and Existence (2013). Here he frequently makes reference to what ordinary language speakers would ordinarily say:

It seems that *in ordinary language we clearly do talk as if* we applied existence to individuals. For example, someone can raise the question whether Napoleon really existed. [p.7, my emphasis]

It is quite true that *one shouldn't dismiss ordinary language here*: someone who marked, according to some Russellian paradigm, 'Hamlet soliloquized' 'false' on an English test in school would get a bad grade justly, not a good one, and he shouldn't say that Russell's analysis shows that the statement is false. Russell's analysis is being misapplied here if it is taken to show that the statement is false. [p.60, my emphasis]

*Everything seems to me to favour attributing to ordinary language* an ontology of fictional entities, such as fictional characters, *with respect to which ordinary language* has the full apparatus of quantification and identity. [p. 69-70, my emphasis]

In the type of thought experiments Kripke and Putnam present we see them rely on their intuitions about what the folk would say under certain conditions. It is this attention to ordinary language use which they consider to be essential to determining the meanings of various words.

We can also see Kripke appeal to ordinary language use when he challenges Donnellan's distinction concerning definite descriptions. Here Kripke famously presents two arguments to show that Donnellan's distinction is not indicative of the meaning of definite descriptions. In one he asks us to imagine an ideal language (a language stipulated to be Russellian) and argues that such a language would not be discernible from ordinary language. In the second

he looks to various natural languages use of the definite article<sup>9</sup>. Both of these arguments rely on the careful investigation of ordinary language use.

In addition to seeing contemporary philosophers *implicitly* using, defending, and appealing to the ordinary language programme, we also see more and more contemporary philosophers do so explicitly and self-consciously. As they do so, they mount a sustained defence of the framework against (2) (e.g., Baz, 2012; Hanfling, 2000; Hansen, 2014). Indeed, this approach seems to be enjoying a renaissance. Moreover, there are some more contemporary philosophical theses which have their roots in the ordinary language philosophy project (see Recanti, 2004). For example, Contextualism, the view that the meanings of some expressions are context relative, is derived from this project (Recanti, 2004; Searle, 1979). Seegar (2010) has defended a position he calls Constitutivism, a positions that holds that the truth of some intuitions, and in particular Twin Earth type intuitions, is determined simply by whether most people share the intuition or not. That is, that ordinary language use at least partly determines the meanings of some expressions. In this sense investigating the intuitions of ordinary language users is entirely relevant because it is their language and they in part determine the semantic value of their terms.

I think there is good reason to reject (2) and disregard The Objection Against Ordinary Language Philosophy. However, the last point I want to make is even if (2) were accepted, that is, if we conceded that ordinary language philosophy is to be given up in favour of ideal language philosophy, it would not be the case that experimental philosophy is deemed uninformative. The Objection Against Ordinary Language Philosophy needs to be able to point out what the problem with the ordinary language project is. As noted, generally philosophers have argued that ordinary language is ambiguous, a feature which is eliminated in ideal language. To this end experimental philosophy, and more specifically this thesis, could be seen as support for the ideal language project by empirically demonstrating the ambiguity in ordinary language use. Lutz (2009) makes a similar point. He argues that experimental philosophy is more closely associated to the ideal language project than usually considered because experimental philosophy shows the premises of ideal language to be true. That is, it is actually through experimental philosophy that we get evidence for

<sup>&</sup>lt;sup>9</sup> This will be more fully discussed in Chapter 2.

(2) because it uncovers any biases, inaccuracies, and ambiguities that exist in ordinary language. Endorsing the ideal language project actually provides reason for experimental philosophy.

It is worth noting here that in defending the project of ordinary language philosophy, I am assuming that intuitions themselves a form of evidence. While this is a widely held view in this literature, many philosophers would not agree (e.g., Williamson, 2007). Indeed, many experimental philosophers take their work to show that our intuitions are not a source of evidence (termed the 'negative program' of experimental philosophy by Alexander, Mallon & Weinberg, 2010). While I attempt to argue that intuitions *are* informative, at least in these instances, and moreover, that *ordinary language users* intuitions are informative, it may be left open for strong proponents of the view against the informative of intuitions to maintain their position.

#### 1.4 The Objection from the General Methodological Problems of Surveys

Opponents and proponents of experimental philosophy alike tend to agree that the use of the survey methodology, utilised frequently by experimentalists, requires careful attention. There are various practical problems associated with the formulation, deployment, and interpretation of surveys and survey data that can call into question any conclusions made. Some philosophers take a stronger stance against the use of surveys in philosophy and hold that because of the problems associated with the survey methodology we cannot derive any substantive conclusions from them (e.g., Kauppinen, 2007). These two attitudes track the two versions of The Objection from the General Methodological Problems of Surveys that I want to present and reject.

The first is the stronger version; it goes as follows:

- 1) Survey responses are not indicative of our philosophical intuitions
- 2) Experimental philosophy is concerned with uncovering our philosophical intuitions

 Therefore, survey responses are not appropriate evidence in experimental philosophy

The claim here is that the survey methodology cannot inform on experimental philosophy because the type of responses elicited by surveys are not the type of responses that we are concerned with. We can take this a step further and claim that, because at least most experimental philosophy uses the survey methodology, experimental philosophy is not an appropriate methodology to inform on philosophical questions. We can see this version of the argument come out in Ludwig (2007):

[R]esponses to surveys about scenarios used in thought experiments are not ipso facto intuitions, that is, they are not ipso facto judgments which express solely the subject's competence in the deployment of the concepts involved in them in response to the scenario. The responses that people make in such circumstances are apt to be the product of a number of different factors, among which are how they understand the task, their background beliefs, empirical and nonempirical, how they think what they say will be taken, loose analogies they may draw with other sorts of situations, how they understand the scenario, whether they pay adequate attention to relevant details, whether they think clearly and hard enough to see what to say in response to the kind of question asked, assuming they understand it correctly, how they think that their interlocutor will (or interlocutors generally would) understand what they say or more generally what they would be trying to convey by what they say or how they respond, as well as perhaps various shortcuts or rules of thumb in reasoning, or plain mistakes. [p.144]

Here Ludwig suggests that the responses elicited by surveys are not intuitions, or at least not the intuitions we are concerned about. This is because we cannot determine what factors come into play to produce a given response so we cannot know if participants are responding to (and only to) the relevant factors. For this reason Ludwig thinks we should abandon experimental philosophy and any reliance on the survey methodology.

I want to argue that (1) is too strong. Even if survey responses are affected by extraneous factors, that is, factors other than those which are relevant to the case at hand, this is not enough reason to think that surveys can never give us relevant defeasible evidence. There are some domains in which surveys will be particularly apt to uncover the relevant intuition. This, I think, is especially true when we are considering intuitions about language. Unlike some other areas of philosophy for which it matters very little what responses are collected

in a survey, in the philosophy of language survey responses seem directly relevant. Language is socially constructed to allow for communication among people. At least in some cases, it is very relevant what different people take the meaning and reference of our terms and phrases to be, and means that surveys are an appropriate source of evidence to uncover the intuitions held. This makes the philosophy of language very different to other areas which tend to be concerned with questions outside the control of people. Haukioja (2015) makes a similar point to defend more recent experimental work in the philosophy of language. In this subfield, he argues, we are typically concerned with our language and how it actually functions rather than all languages and how they must function. For this reason he takes empirical considerations to be particularly relevant to questions of language compared to other areas of philosophy. We can see this view come out in Seegar's (2010) Constitutivism (discussed in Section 1.3 above) and some comments by Cohnitz (2015) who argues that the semantic facts about theories of reference are not independent of the intuitions and usage of competent speakers, but are in fact constituted by them. If we do take the semantic value of our terms to be directly constituted by ordinary language users' intuitions, then surveys must be directly relevant here.

However, we still have the weaker problem that survey responses are affected by many irrelevant factors unintentionally. Here we might present the argument in a different form; call it The Objection from the General Methodological Problems of Surveys 2:

- 1) The survey methodology is difficult to perfect and prone to errors
- 2) If a methodology is difficult to perfect and prone to errors it is not informative
- Experimental philosophy utilises the survey methodology to inform on philosophical questions
- 4) Therefore, experimental philosophy cannot inform on philosophical questions

I take it that (1) is uncontroversially true; surveys are notoriously hard to design and are liable to various practical problems. Moreover, philosophers are not typically expert in using the experimental design. This means that very often experimentalists in philosophy are not careful enough in their use of surveys or their interpretation of the data collected from

those surveys. We can look at some of the kinds of problems encountered with the utilisation of surveys.

Firstly, it is hard to determine whether a survey is formulated so that it will elicit the *right* concept, or elicit the right concept reliably across participants. As Ludwig notes above, the responses participants give are affected by a number of contextual and psychological factors. For this reason it is hard to discern whether participants are responding in the right sort of way, that is, to the relevant concept under the right circumstances, and whether they are doing so consistently. Cullen (2010) too has argued that it is a mistake to think that we can simply 'read off' what intuition is held among people from a survey because so many contextual features are unwittingly placed into surveys. Contextual cues can affect how participants respond which indicates that surveys do not reliably track the relevant intuition. For this reason Cullen urges experimentalists to reconsider their methodological assumptions.

Secondly, we need to consider who is participating in the experiments. Kauppinen (2007) has noted that competence and performance error are both problems with surveys. To ensure the right concept is being applied by participants, experimentalists need to consider who is competent with the relevant concepts and control for such a competence among participants. And even if we can somehow distinguish those that are competent from those that are not, we need to concede that even competent users will make mistakes. Performance error is a general problem with experimentation and experimentalists need to be careful to create conditions that restrict the possibility of errors as much as possible to reduce this effect. Although he expresses doubt about the possibility of producing such conditions, Kauppinen refers to such a condition as 'ideal circumstances' (Kauppinen, 2007).

Lastly, there are also issues with the interpretation of the data collected from surveys. Keller (2011), for one, expresses concerns with the interpretation of survey data as an experimenter is able to use data in a number of different ways depending on their goals by presenting their findings in generic claims. Generic claims about philosophical issues ("humans have free will") can easily be interpreted in different ways. Keller takes this to mean that in best practice an experimenter needs to be clear about what their data shows

and how to best present it to avoid over reaching the data, that is, claiming that the data shows more than it does, or eliciting misinterpretations among readers.

Although I accept that the survey methodology is prone to these kinds of problems, I do not take it that (2) is true. Brock (2014), for one, argues that all this shows is that we need to be careful in our use of experiments in philosophy, not avoid them altogether. It is not the case that surveys cannot tell us anything; it just means that we need to be cautious. These are problems in practice not in principle and I think it is entirely possible to constrain these kinds of problems associated with survey methodology to such a degree that they do not compromise the conclusions drawn from the data. Philosophers traditionally have not been trained to conduct experiments, and indeed there are mistakes in the way experimental philosophy has been carried out in the past. However, philosophers are becoming trained and expert in this area and the methodology is being more carefully applied. Indeed, critics of the survey methodology have actually helped to improve the methodology as philosophers can use the criticisms to improve their survey design.

One early example of an experimental paper that was severely criticised based on its methodological design is an experiment by Machery et al. (2004) about the reference of proper names. I will discuss this case in more detail in Chapter 3. For now, however, it is worth noting that in this experiment Machery et al. (2004) found a lot of variation in the intuitions Westerners' and East Asians held about the reference of proper names. However, their findings were criticised based on the survey design; for examples, about the formulation of the question (Martí, 2009), about the linguistic competence of participants (Lam, 2011), and about ambiguities in the presentation of the survey (Deustch, 2009; Ludwig, 2007; Sytsma & Livengood, 2011). These criticisms allowed for the experiment to be re-done with an improved survey design (e.g. Machery et al., 2010) indicating that many of the practical problems associated with the survey methodology can be mitigated with better practice.

#### **1.5 The Expertise Objection**

The Expertise Objection is best regarded as a defence of 'armchair philosophy' against the experimentalist challenge, this being a challenge against the informativeness of philosophical intuitions. The experimentalist challenge (sometimes called the negative thesis or program of experimental philosophy) regards the large amounts of variation in folk intuitions as indicative that philosophical intuitions are not reliable (e.g., Machery et al., 2004, for intuitions about the reference of proper names). The Expertise Objection presents a defence of philosophical intuitions. Although there are different characterisations, in essence what it states if the following kind of argument:

- 1) The intuitions of professional philosophers are better or more reliable than the intuitions of non-philosophers ('the folk') about philosophical questions
- Experimental Philosophy tests the intuitions of non-philosophers ('the folk') about philosophical questions
- 3) Experimental Philosophy is not informative about philosophical questions

Proponents of this argument take it that (1) is true, which we see come out in the following passages (note that there will be further examples below):

[I]t seems prima facie eminently plausible that people who have thought long and hard about a certain issue will also make better intuitive judgments in their area of expertise. [Horvath (2010) p. 465]

[Philosophers have] skills in responding to questions about described scenarios on the basis of one's competence in concepts involved. [Ludwig (2007) p.138]

[A] totally uneducated person may reflect very little.... semantics is notoriously hard and the folk are a long way from being experts....I think we have a principled basis for preferring the intuitions of philosophers to those of the folk on these semantic matters because the philosophers are more expert. [Devitt (2011) p. 426]

The general themes that come up in these kinds of passages are that philosophers have thought harder about the relevant issues, are more familiar with the relevant concepts, and have special training in abstract and objective reasoning and so are less influenced by

irrelevant factors and less likely to commit mistakes in their reasoning. Accordingly, it is thought that philosophers are less likely to fall victim to fallacious reasoning with regard to philosophical issues in the same way that someone trained in mathematics or statistics is less likely to commit the gamblers fallacy, the base rate fallacy, the Prosecutor's fallacy, etc. (Hofmann, 2010). Moreover, as we see come out in the last quote, in science and other disciplines the judgments of laypersons are not given equal weight to those of experts in the field, and so it seems we should do the same in philosophy. This kind of 'analogical argument' comes up frequently in defence of (1).

What I will argue is that all the kinds of reasons that have been given in support of (1) are all fallacious in some way. Philosophers do not, contrary to the argument, have superior intuitions about the sort of cases experimental philosophers test. The way I will do this is by extracting themes from the given passages for reasons philosophers give in favour of (1) to use as testable hypotheses. I will then draw on empirical research that disconfirms each hypothesis. As such, it cannot be used as support for (1). Thus, I will argue, the Expertise Objection begs the question, because its proponents have given the experimental philosopher no good reason to think that (1) is true.<sup>10</sup>

Before doing this, however, it is worth briefly considering (2), as well as the argument as a whole. The argument as it is presented here is not valid; (1) and (2) together do not entail (3). If it is the case that philosophers have better intuitions than non-philosophers, as (1) claims, it does not mean that the intuitions of non-philosophers are not informative at all, just that they are not as good a source of evidence about philosophical issues. Experimentalists never suggest that folk intuitions are the only kind of evidence to take into account when investigating philosophical questions; as the metaphor runs, experimental philosophy is only put forward as one tool in the philosopher's toolbox.

Premise (2) is also problematic. Proponents of The Expertise Objection take it that (2) is true; however, as we will see, there is a lot of experimental work that does not only rely on or investigate non-philosophers' intuitions. Some of that work empirically investigates the intuitions of *philosophers*. I will discuss this research in relation to (1) below. And some

<sup>&</sup>lt;sup>10</sup> Note that many advocates of experimentalist methods have challenged the Expertise Objection on the grounds that philosophical expertise should not put a limit on whose intuitions are relevant (e.g., Feltz, 2009; Genone, 2012; Knobe & Nichols, 2007; Machery, 2012).

other experimental work does not investigate intuitions at all. Some philosophers use brain imaging (e.g., Greene & Paxton, 2009; Greene, Nystrom, Engell, Darley & Cohen, 2004), reaction times (e.g., Chemla & Bott, 2010; Schwarz, 2013), and other independent sources of evidence. For example, Schwitzgebel (2009) used library records to find that both classic and contemporary ethics books, that is, those most likely to be borrowed by moral philosophers, were more likely to be missing than non-ethics books. Using the experimental method he was able to infer that there are at least some cases in which moral philosophers behave no more morally than others. A slightly weaker version of (2) can still do the work required by the proponent of The Expertise Objection. It is still the case that *most* Experimental Philosophy tests the intuitions of non-philosophers. If (2) is modified to:

2\*) Experimental Philosophy generally tests the intuitions of non-philosophers ('the folk') about philosophical questions

Then the proponent of the argument can avoid this last concern.

So what might be said in support of (1)? Why think that professional philosophers, compared to the folk, have superior intuitions about philosophical cases? Five reasons have been given in the literature, although, as we will see, these reasons are not completely independent and are often highly related:

- 1. philosophers are less influenced by irrelevant factors
- 2. philosophers reflect more on their intuitions
- 3. philosophers are more familiar with the relevant concepts
- 4. philosophers show convergence of opinion
- philosophers are like experts in other fields who are considered to have superior intuitions

#### 1.5.2 Philosophers are less influenced by biases and irrelevant factors

A number of philosophers have argued that the folk intuitions are influenced by irrelevant factors and are therefore unreliable. We can see this view come out in the following passages:

[Philosophers] have intuitions about hypothetical cases that typically do not vary with irrelevant factors.... the charge of epistemic untrustworthiness are explicitly restricted to non-philosophers. [Horvath (2010) p.464]

[W]e need robust intuitions that are elicited only when conditions in (A) obtain—that is, when failures of competence, failures of performance, and influence of irrelevant factors are ruled out. [Kauppinen (2007) p. 138]

Proponents of this view tend to call on experimental work which has uncovered evidence of the biases that the folk display. For example, Lombrozo (2009) found that when given 'push' and 'switch' versions of the trolley case, undergraduate students made different responses to each case depending on whether they received only one case or both cases together. That is, participants thought that it was more morally permissible to push if they were shown both the 'push' and the 'switch' case instead of just the 'push' case. Moreover, undergraduate students' show an actor-observer bias to the trolley case such that they are less likely to think it is morally permissible to press the switch when they are asked to imagine that they themselves are the agent (Nadelhoffer & Feltz, 2008).

More recent experiment work, however, indicates that philosophers suffer from the same biases that non-philosophers do. For one, philosophers show the same order effects in a number of moral judgements as the folk (Schwitzgebel & Cushman, 2012). When presented with the trolley case, both philosophers and non-philosophers showed the exact same bias to rate the 'push' and 'switch' cases as equivalent if the 'push' case was shown first. That is, they rated 'push' as more morally good if presented after 'switch'. Moreover, in a recent experiment Schwitzgebel and Cushman (up) attempted to reduce this bias in philosophers intuitions in four different ways: by limiting the philosophers to those with specific expertise on the subject matter of the question used, by limiting the philosophers to those who

reported stable intuitions, by encouraging reflective responses,<sup>11</sup> and by pairing scenarios that differed in phrasing but not content. They found that none of these measures reduced the bias in philosophers intuitions compared to non-philosophers.

Moreover, philosophers also fall victim to an actor-observer bias to moral judgements, such as the trolley case, although they showed the opposite effect to non-philosophers (Tobia, Buckwalter & Stich, 2013). In this case philosophers thought that there was a greater moral obligation to switch when they were the actor compared to when they were the observer. The folk, on the other hand, thought that there was a greater moral obligation to switch when they were the observer compared to when they were the actor.

In another experiment, Tobia, Chapman and Stich (up) found that the moral judgements of both philosophers and non-philosophers are affected by smell. They presented some of their participants with surveys that had been sprayed with disinfectant and found that this manipulation affected how harshly the participants rated various actions as morally wrong.

Non-philosophers in the 'clean' group (received a survey sprayed with disinfectant) rated the actions more harshly than non-philosophers in the control group. This bias was also observed by philosophers, but only when they were the actor in the scenarios. When presented with cases for which they were the observer there was no difference between the ratings of philosophers in the 'clean' group and philosophers in the control group.

As a final example, while previous research has indicated that the personality trait 'Extroversion' is a predictor of compatibilist intuitions about freewill (Feltz & Cokely, 2009; Nadelhoffer et al., 2009), Schulz, Cokely and Feltz (2011) found that this bias continues even after controlling for free will expertise.

These studies indicate that, while sometimes the intuitions of philosophers and nonphilosophers differ, the intuitions of both groups are affected by seemingly irrelevant manipulations. Therefore, the claim that philosophers will be resistant to such biases is false. If this is the best basis for holding that (1) is true, proponents of The Expertise Objection need to postulate and test some specific bias and show that it is a bias that the

<sup>&</sup>lt;sup>11</sup> Reflective thinking was encouraged by a request to think carefully and a delayed time before responses were made.
folk fall victim to but philosophers do not. As far as I am aware no such bias has been identified.

# 1.5.3 Philosophers reflect more on their intuitions

Some proponents of the Expertise Objection have suggested that philosophers have superior intuitions to non-philosophers because they have more reflective intuitions (e.g., Devitt, 2011, pg. 426; Horvath, 2010, p. 464; Kaupinnen, 2007). Devitt (2011) specifically argues that we should not trust the intuitions of the folk about the reference of proper names because the cases presented are complex and the folk may not reflect much on their intuitions. Indeed, Devitt has argued that intuitions "are empirical theory-laden central-processor responses to phenomena, differing from many other such responses only in being fairly immediate and unreflective, based on little if any conscious reasoning" (2006, p. 103). He suggests, however, that experts are more likely to have trustworthy intuitions because they are based on previous reflective practices (Devitt, 2011, p. 425).

Although not in relation to The Expertise Objection, Livengood et al. (2010) have argued that philosophers are particularly reflective and that this is what distinguishes them from nonphilosophers. Furthermore, they actually tested this claim by measuring the reflective thinking of over 4000 participants using a well-known test from psychology studies, Frederick's (2005) Cognitive Reflection Test (CRT). The CRT measures the likelihood one has to challenge their intuitions rather than accepting what is first bought to mind or what seems to be the obvious response. It is made up of three items which each have an intuitive answer which comes to mind immediately but which turns out to be incorrect; see the following:

(1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? \_\_\_\_\_\_ cents

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? \_\_\_\_\_ minutes

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? \_\_\_\_\_ days

In these cases, the right answers are 5, 5, and 47, although the unreflective or intuitive answers are 10, 100, and 24.<sup>12</sup> Participants are assigned a score 0-3 depending on how many questions they correctly answered. What Livengood et al. found was that CRT score increased with philosophical training, even after controlling for level of education. They suggested that this finding indicates that people inclined toward philosophy are more reflective in temperament, although, as they note, it might be that philosophers are taught to be more reflective through their course of study. Either way, this study suggests that proponents of the Expertise Objection are right to think that philosophers have more reflective intuitions than non-philosophers.

I agree that the CRT gives good evidence that philosophers are particularly reflective. However, I think it is a mistake to think that this means that philosophers have better philosophical intuitions. In fact, there seems to be good reason to think that this is not the case. In some recent experiments about the meaning of different terms and phrases, which I will be presenting in Chapters 1-3, I incorporated the CRT into my surveys and found that there was not any relationship between reflectiveness and philosophically relevant intuitions. For each of the three studies conducted, participants received a survey that first contained a version of the CRT followed by a number of philosophical questions to elicit their intuitions about definite descriptions, proper names, or natural kind terms. There were three versions of the CRT used: the control version which did not include a copy of the CRT questions, the 'unreflective' version which contained only the three CRT questions followed immediately by the philosophical questions, and the 'reflective' version which had the CRT guestions prefixed with the sentence "The three problems below are all difficult (more difficult than they might first seem). Think carefully before you answer" and the philosophical questions prefixed with the sentence "Now turn over and answer the English language questions. These questions may also be difficult, so again reflect carefully before you answer." The reason for this manipulation was to see whether reflective thinking could be primed for. In each of the three studies CRT score was not related to philosophical intuitions, nor was priming for reflective thinking related to philosophical intuitions. What these results indicate is that participants who scored highly on the CRT were not more likely to hold one philosophical intuition over another. For example, in the natural kinds study, CRT

<sup>&</sup>lt;sup>12</sup> See Frederick (2005) for why it is the case that these are the answers.

score was not related to whether participants had Internalist or Externalist intuitions. Similarly, in the proper names study, only a minimal effect of CRT score was found for one question. There was similarly no difference in philosophical intuitions between the three CRT versions indicating that participants in the 'reflective' group did not answer differently on the philosophical questions than participants in the control or 'unreflective' groups.<sup>13</sup>

Another reason to suspect that cognitive reflectiveness is not indicative of better philosophical intuitions is that there are gender differences on the CRT which are not also present in philosophical questions. Although many studies indicate that females robustly tend to do worse on the CRT than males (e.g., Frederick, 2005; Livengood et al., 2012), I found that in the proper names study although females did worse on the CRT than males, they did better on the philosophical control question. There were also no gender differences on the philosophical questions.

The data from these experiments indicates that being reflective or having reflective intuitions does not influence, or thereby improve, philosophical intuitions. This suggests that even if philosophers are more reflective, we should not expect them to have different or better philosophical intuitions than non-philosophers.

## 1.5.4 Philosophers show convergence of opinion

The huge amount of disagreement uncovered about the intuitions held by non-philosophers is sometimes used as a reason to reject the usefulness of the folk intuitions (Kauppinen, 2007; Ludwig, 2010). The argument states:

- 1) If intuitions are reliable you would expect convergence
- 2) There is no convergence among the folk intuitions
- 3) The folk intuitions are unreliable

<sup>&</sup>lt;sup>13</sup> But see Paxton et al. (2011), Pinillos et al. (2011), and Knobe & Samuels (2013) for experiments that link CRT score to moral judgements.

We tend to think that convergence is a sign of reliability; that if many people come to the same conclusions, based on similar information, then that conclusion is more likely to be correct. Take physics for example. Experts in physics converge on what principles guide the physical world and so we trust their intuitions on such matters more than the folk intuitions because the experts are considered more reliable. Two points can be made about the argument. One is that there is not convergence among philosophers either, so according to this argument philosopher intuitions are also unreliable. Secondly, even if it is the case that philosophers converge at least sometimes, or in some places, or on some issues, there is good reason to doubt that convergence of this kind is an indication of reliability.

On the first point, philosophers notoriously disagree on many, if not all, philosophical problems. Indeed it seems that philosophy is an area where access to more information about a question is no guarantee of holding a particular view about it. Even cases where there is thought to be a consensus among philosophers (where there is not among the folk) turn out to be unsettled. Consider theories of reference. While experimental philosophy indicates that among the folk there are large amounts of disagreement about which theory is intuitive (e.g., Machery et al., 2004), we tend to think that at least among philosophers there is consensus in favour of Kripke's (1980) causal-historic theory. However, there are, in fact, many dissenters from this popular positon, that is, philosophers who have read Kripke but still follow something in the manner of the traditional descriptive theory of reference (e.g., causal-descriptivist's such as Jackson, 1998b; Kroon, 1987).

Bourget and Chalmers (2013) have experimentally revealed the extent of disagreement among philosophers in a large scale study of almost 2000 current philosophers online. They presented philosophers with a survey of thirty questions across many philosophical domains; for examples:

- a) Abstract objects: Platonism or nominalism?
- b) Epistemic justification: internalism or externalism?
- c) External world: idealism, skepticism, or non-skeptical realism?
- d) Free will: compatibilism, libertarianism, or no free will?
- e) Mental content: internalism or externalism?

- f) Normative ethics: deontology, consequentialism, or virtue ethics?
- g) Personal identity: biological view, psychological view, or further-fact view?
- h) Proper names: Fregean or Millian?
- i) Time: A- or B-theory?
- j) Trolley problem (five straight ahead, one on side track, turn requires switching, what ought one do?): switch or don't switch?

What they found was large amounts of disagreement among philosophers. In fact the greatest agreement was 81.6% for the option 'non-skeptical realism' on the question of the status of the external world. They also found many effects based on age, gender, specialisation, and location. To give a few examples, there were positive relationships between continental European philosophers and subscribing to a Fregean theory about proper names, US philosophers and subscribing to a Millian theory about proper names, Australasian philosophers and subscribing to consequentialism and the B-theory, and female philosophers and subscribing to nominalism. There were also negative relationships between female philosophers and considering zombies metaphysically possible, and age and subscribing to externalism about mental content or subscribing to the B-theory.

Not only does Bourget and Chalmers' (2013) survey empirically quantify the lack of consensus among philosophers, it also allows for some comparison to the folk. The breakdown for of support in the philosophy of language questions on proper names and mental content is similar to the breakdown seen in the folk in some recent data I have collected. Although not directly comparable, in a survey about proper names I found that approximately 30% of US and NZ folk supported a descriptivist theory compared to a Millian theory. Similarly Bourget and Chalmers found that 28.7% of philosophers subscribed to a Fregean theory about proper names compared to a Millian (or 'other') theory. Moreover, in a survey about natural kind terms I found that NZ folk tended to support externalism (50-66%); Bourget and Chalmers also found a majority of philosophers supported externalism, although about mental content. In both these cases we see similarities in support between philosophers are asked directly which theory they support while the non-philosophers give indirect support for a theory through their responses to thought experiments.

Kauppinen (2007) has defended armchair philosophy against experimentalism and suggested that philosophers' intuitions do converge, at least some of the time. He argues that experimental surveys only yield surface intuitions, not the robust reliable intuitions that we should be interested in. It is using the 'Dialogue and Reflection Models', methods that are used by philosophers, which give us reliable intuitions. Indeed, Kauppinen thinks that by engaging in the practices of dialogue and reflection philosophers *do* tend towards convergence, at least some of the time.<sup>14</sup> That is, even though there may be significant disagreements among philosophers, the accumulation of examples and counterexamples can lead to convergence (p. 33).

On the second point, even if Kauppinen is right about there being some convergence among philosophers, we should not accept, without further argument, that this means that philosophers are more reliable than the folk. There are many examples of convergence without reliability. Take homeopathy for example. Experts in homeopathy converge on their beliefs; however, we do not take their beliefs to be more reliable than other peoples beliefs on the matter (in fact, we might take the homeopaths to be less reliable). We also see large amounts of convergence concerning matters of religion, although it cannot be that every religion is right, so convergence here does not indicate reliability.

#### 1.5.5 Philosophers are more familiar with the relevant concepts

Some proponents of The Expertise Objection have suggested that philosophers have more reliable philosophical intuitions because they are more familiar and competent with using the concepts appealed to in these intuitions. Ludwig (2007) in particular has argued that surveying the folk about philosophical concepts is uninformative because philosophers have a better underlying competence with the concepts elicited by thought experiments in their area of expertise. One way to put this is considering scientists' use of scientific terms. A scientist using the terms 'weight' and 'mass' might mean something different to someone without a scientific background who uses the terms. We should then expect in some

<sup>&</sup>lt;sup>14</sup> Kauppinen (2007) gives an example of Internalism about moral motivation (p. 33-35). Here he argues that there is a convergence in intuitions about moral internalism over time through philosophical dialogue and reflection.

circumstances that the scientist will have better intuitions about the nature of mass because they will be employing the concept in the right kind of way.

It might be argued that the concepts employed in philosophy are not like the concepts used in science. We might think that, at least for the most part, non-philosophers are very competent with the concepts of 'knowledge', 'freedom', and 'right action', and maybe to a lesser extent, 'necessity', 'sufficiency', and 'logical consequence'. Ludwig appears to disagree with this claim. He argues that being able to use a term or *understanding* it is different from *possessing a concept*. Concept possession involves being able to articulate the application conditions of that concept, rather than just being able to use the term correctly. The role of thought experiments is to elicit intuitions of the subject based purely on the competence one has deploying the concept of interest, and it is important, according to Ludwig, that the subject is aware of this. It is philosophers, he suggests, that are in the better position to do this. Indeed, he continues that there is a difference between simply responding to a thought experiment, on the one hand, and having semantic intuitions about it, on the other. Ludwig claims that while the semantic intuitions of philosophers and of the folk do not differ, it is the philosophers' responses to thought experiments that are more likely to be indicative of those intuitions. That is, non-philosophers are more likely to respond to thought experiments without employing their semantic intuitions (p. 152).

Nonetheless, if it is the case that philosophers are more competent with employing the concepts of their area of expertise, there is the further question as to whether we should expect philosophers to then have reliable intuitions about areas of philosophy that are not their specialty. Because philosophy is such a diverse subject, there is a huge disparity in the areas of familiarity and competence that philosophers have. If we do not consider a non-philosopher to have access to the concepts appealed to in thought experiments, should we still consider an epistemologist to be competent with the concepts employed in moral thought experiments, or political philosophers to be competent with the concepts employed in thought experiments about personal identity. Philosophers are not experts in philosophy *per se*, only in a specific domain. Therefore, we should not expect philosophers in general to have better intuitions than the folk about philosophical questions, or at least not when they are considering questions outside of their field of expertise.

Perhaps, though, we might expect experts *in a specific philosophical domain* to have superior intuitions within that domain as compared to non-philosophers. Machery (2012) has challenged this proposal by specifically comparing the intuitions of philosophy of language experts to the folk about reference. In specific response to Devitt's (2012) Expertise Objection to experiments about reference, Machery argues that linguistic and philosophical expertise does not improve intuitions about reference. That is, having expertise in these areas does not mean that one is more likely to have reliable intuitions about thought experiments in the area. He says:

Philosophers of language and linguists are aware that one must explain how words refer; they understand what is at stake in explaining reference; they are well acquainted with the (by hypothesis, poor) theories of reference put forward by philosophers and linguists; and so on. It is unclear why this kind of advantages would make their intuitions more reliable than laypeople's. [Machery (2012) p. 43-44]

To demonstrate this, Machery compared the expert intuitions of linguists and philosophers of language, separated into nine groups based on specialisation, to similarly educated folk. He found that all the groups tended to have Kripkean intuitions, and were very confident in their answers, but that the proportions varied from 66.7% to 88.6% across the groups suggesting some variation even among experts. Machery then separated the experts into two groups; those likely to have read Kripke's *Naming and Necessity* (philosophers of language and semanticists), and those who would likely be influenced by the descriptions associated to a name, and again compared their intuitions to the folk. What he found was that experts who had likely read Naming and Necessity were more likely to have Kripkean intuitions (86.4%) than the other experts (68.7) or the folk (76.9%). If expertise leads to convergence to the truth we should not expect this variation— especially given that the groups of experts do not converge more than the folk.

What Machery does suggest is that, rather than expertise producing better intuitions, expertise in philosophy might actually *bias* intuitions. If this is the case, then if anything philosophers have less reliable intuitions than the folk, who are free of theoretical commitments. The idea that philosophers are biased in their intuitions may be demonstrated in Bourget and Chalmers' (2013) study from the effects having to do with specialisation and location. That Australasians are more likely to have consequentialist

intuitions, or that philosophers of language are more likely to have externalist intuitions about mental content suggest that these are not 'natural' or obvious intuitions to have, but are intuitions that have been ingrained through their academic setting. This might also suggest the possibility of a philosophical bullying phenomenon occurring within academia. It is this possibility which I will take very seriously in Chapter 4.

To sum up: philosophers are not, generally, experts in philosophy, only of some part of it. Therefore, we should not consider their intuitions, generally, to be better than nonphilosophers. Moreover, Machery's study suggests that even when restricted to areas of specialisation, philosophers do not necessarily have better intuitions than non-philosophers.

# 1.5.6 Philosophers are like experts in other fields who are considered to have superior intuitions

The last argument I want to consider in favour of The Expertise Objection is the analogical argument. This argument is made frequently against the use of folk intuitions by considering other disciplines in which we do not give any credence to the folk; instead we defer to experts in that field. Philosophers, it is argued, are akin to these experts. Consider the following examples of the analogical argument from the relevant literature:

[W]e should acknowledge that not all intuitions are created equal . . . . For example, the physical intuitions of professional scientists are much more trustworthy than those of undergraduates or random persons in a bus station. [Hales (2006) p. 171]

Consider the hypothesis that professional physicists tend to display substantially higher levels of skill in cognitive tasks distinctive of physics than laypeople do [Williamson (2011) p. 220]

Scientists have and rely on physical intuitions, intuitions that are trained, educated, and informed and yet are good indicators of truth for those very reasons. In the same way, the modal intuitions of professional philosophers are much more reliable than either those of inexperienced students or the 'folk'. [Hales (2006) p. 171]

[I]t matters exactly zero what folk intuitions about free will, twin earths, Chinese rooms and the like are, just as it matters zero what folk intuitions about logic, mathematics, physics or the game of chess are to the practices in those fields. [Pigliucci (2011)]

Here we see proponents of The Expertise Objections appeal to research that shows that experts in other fields clearly have better intuitions about that field than non-experts. The kinds of examples drawn on are mathematicians, logicians, physicists, and chess experts. For example, studies show that physicists have better intuitions about physical objects; e.g., about the trajectory of moving objects (Kaiser et al., 1986; see McClonskey, 1983). The idea is that experts in these domains all show some kind of resistance to biases which makes their judgements more reliable. Therefore, the argument goes, we should expect the same about philosophical judgements. Commonly mathematicians and logicians are brought up, although that they have superior intuitions about their fields of expertise is generally only assumed to be true. We see this come up in Ludwig (2007) who asks us to imagine the following survey is conducted:

Consider the following two number series:

(a) 0, 1, 2, 3, . . .

(b) 1, 3, 5, 7, . . .

Which of the following claims is correct?

- (A) (a) has more members than (b).
- (B) (b) has more members than (a).
- (C) (a) and (b) have the same number of members.

Ludwig theorises that undergraduate students without mathematical training would overwhelmingly incorrectly respond (A) while those with a certain level of training would correctly respond (C). Nado (2014a; 2014b) similarly takes this survey *thought* experiment to be a good demonstration of improved judgements among experts:

We naturally assume that scientists, mathematicians, and the like have special expertise in their fields; therefore, it's reasonable to assume that philosophers have expertise in their field as well, even in the absence of experimental evidence. [Nado (2014a) p. 632]

For the first, take Ludwig's (2007) undoubtedly correct prediction that mathematicians would display fewer errors than untrained subjects when comparing the cardinalities of various infinite sets. [Nado (2014a) p.636]

Ludwig (2007) appeals to a plausible case of what might be called improved intuition among mathematicians. He notes that no mathematician would feel troubled by the discovery that naïve subjects judge that the series of natural numbers contains more members than the series of odd positive integers. We would of course expect trained mathematicians to show improved performance when judging such a case. [Nado (2014b) p. 6]

Williamson makes a similar assumption about the improved skill of physicists:

Consider the hypothesis that professional physicists tend to display substantially higher levels of skill in cognitive tasks distinctive of physics than laypeople do. The hypothesis could be tested by systematic experiment. But even before that has happened, one can reasonably accept it [Williamson (2011) p. 220].

The problem is that this argument relies on an empirical claim that experts have better or 'improved' intuitions. However, as the research I have discussed above suggests, there is no detected improvement of intuitions among philosophers (Nado, 2014a, p.636). The analogical argument, consequently, is a non-starter. We can, however, consider why the analogy fails.

There seem to be some relevant differences between philosophy and the kinds of cases brought up in the analogical argument. It is assumed from the limited examples given that all expertise results in better intuitions or judgements in that domain. However, there are other areas where this does not seem to be the case and it is these cases which make a better analogy to philosophy. Machery (2012) discusses linguists and statisticians as such examples. For one, research has indicated that linguists do not have more reliable syntactic intuitions than the folk (see Machery, 2012). When asked to judge the acceptability of sentences, linguistic expertise does not produce more reliable intuitions, that is, there is not more agreement among experts (see Culbertson & Gross, 2009). Furthermore there are disagreements about whether or not a sentence is ambiguous (in Machery, 2012). Similarly, statisticians do not appear to have better intuitions about drawing statistical inferences than the folk for complex data sets (see Machery, 2012). Similarly, Philip Tetlock (2005) in Expert Political Judgement found that political and economic experts were not more reliable at estimating probabilities or making predictions about how events would unfold with respect to their areas of specialisation than non-experts. Experts were, however, more confident in their responses and better at explaining away or defending their bad predictions. The

question that arises, then, is what might we take the relevant differences to be between the cases where intuitions are improved and those where they are not?

For one, the cases in which there are improved intuitions are cases where there is a lot of feedback, confirmation, and disconfirmation available. That is, there can often be almost instantaneous feedback to a given response. One reason that this feedback is available is because these are cases in which there is consensus among experts about what the correct response is. For example, when learning physics, mathematics, or chess, a student can be told whether they made a correct or incorrect, good or bad response immediately so they can recalibrate their concept accordingly (see Weinberg et al., 2010). In his study of expertise, Ericsson has argued that at least some forms of expertise require deliberative practice by which he means an effortful acquisition and maintenance of expertise which relies on fast, informative feedback (Ericsson, 2003; Ericsson, Krampe & Tesch-Romer, 1993). Weinberg et al. (2010) take up this research to argue that philosophy is a field that is deficient in the kind of feedback required. There is often little consensus among experts and so no clear feedback is available. Similarly, political predictions about the future cannot be verified immediately, nor is there a clear correct or incorrect response that students can be told in linguistics because of the lack of a 'right' answer (Machery, 2012).

It is also not clear that the cases discussed in the analogical argument should count as *intuitions*, or at least as the same kind of intuition appealed to in philosophy and philosophical thought experiments. The intuitions in the case of mathematicians, physicists and chess experts are based on previous experiences and learned beliefs and principles from their study. Experts in these fields tend to have explicit awareness of why they have the relevant intuition, and what principles are guiding them (Nado, 2014b). Philosophers do not share this kind of knowledge; they do not have transparent access to the mechanism behind their intuitions (Nado, 2014b). This difference in the opacity of intuitions points to a methodological difference between the two kinds of cases. As Nado (2014a) points out, in mathematics and physics the intuitions of experts are not used to inform on what theory is correct, like in philosophy, but are the direct result of learning that a specific theory is correct. Philosophers' intuitions, on the other hand, are used explicitly as evidence for a theory (p. 637). This difference in the quality and the role of intuitions in different

disciplines suggests that the analogical argument does not provide a convincing analog to philosophy to drive the Expertise Objection.

Another difference between the two different *types* of intuitions that we are currently interested in is that the intuitions of mathematicians and physicists can, presumably, become better with the acquisition of expertise. That is, the intuitions can be changed over time with further learning. This is not such a standard feature of philosophical intuitions. Mizrahi (2015) compares the kinds of intuitive judgements we make in philosophy to perceptual judgements. Perceptual judgments are sensory seemings that cannot be improved with training (being an expert on the atmosphere does not change the colour of the sky perceived). Likewise, she argues, intuitive judgments are intellectual seemings that cannot be improved with training.

### 1.5.7 Philosophical expertise and philosophical intuitions

The Expertise Objection fails to undermine experimental work in philosophy because there does not seem to be a principled reason to prefer the intuitions of some group over another, at least for the kinds of intuitions that experimentalists probe for. Indeed, if anything experimental philosophy looks to be particularly important to tell us when we should be wary of our intuitions and of putting too much emphasis on their significance. I think the experimentalist has a very strong defence against the Expertise Objection no matter what reason is given to think that philosophers have superior intuitions [that is, for (1) above]. What I want to do here is briefly discuss where this leaves philosophical expertise Objection is aimed at philosophers who claim that experimental work suggests that intuitions are not a reliable source of evidence at all. That is, that because of the variation in intuitions uncovered by experimental work, we cannot consider intuitions to be informative with respect to philosophical questions. If the Expertise Objection fails, which I believe it does, it seems that we are again left with this conclusion, that we cannot use intuitions to do philosophy. I will be presenting a different interpretation of experimental

data which rejects this claim, but does so by appealing to the Ambiguity Thesis rather than to expertise.

What then can we conclude about philosophical expertise? Given the experimental data, some philosophers have claimed that there is "no such thing" (Mizrahi, 2014).<sup>15</sup> Taken in context though, this claim can only be reasonably thought to mean that there is no expertise when it comes to having correct philosophical intuitions. Experimentalists would not (generally) want to concede that philosophers are *not* expert in their field. However, it does not look like their expertise can lie in having superior intuitions.

Despite the huge emphasis on experimentalism and the Expertise Objection, only one paper that I know of within the expertise debate has sought to locate philosophical expertise outside of philosophical intuitions.<sup>16</sup> Nado (2014b) argues that the comparison of philosophy to the sciences is still a useful analogy to make, but that it does not provide grounds to challenge experimental philosophy. She argues that it is useful to motivate the claim that we have philosophical expertise, but not that the expertise lies in intuitions. Instead she offers a weaker version of the analogical argument, what she calls the 'basic' version. See the following:

Scientists (e.g.) possess some form of scientific expertise, and therefore philosophers can be expected to possess some form of philosophical expertise. This version [of the analogical argument] is, however, obviously too weak; it does nothing to motivate the idea that the expected expertise should involve, specifically, a resistance to the particular biases that appear to exist in folk reactions to thought experiments [p. 6].

Here we see Nado suggest that the analogical argument need not appeal to intuitions at all as a measure of expertise. This basic version of the argument she claims is better than the more commonly held version of the argument which either explicitly or implicitly states that philosophical expertise results specifically in *improved intuitions*. She says:

<sup>&</sup>lt;sup>15</sup> See Mizrahi (2014). 'Philosophical Expertise' in Nadelhoffer's blog *Experimental Philosophy* <a href="http://philosophycommons.typepad.com/xphi/2014/10/philosophical-expertise.html">http://philosophycommons.typepad.com/xphi/2014/10/philosophical-expertise.html</a>.

<sup>&</sup>lt;sup>16</sup> Some philosophers have argued that we do not do philosophy by relying on intuitions. In particular, see Cappelen (2012) in *Philosophy without Intuitions* where he argues that the philosophical method does not include the use of intuitions.

On the most popular approach [of the analogical argument], by contrast, the analogy appeals not merely to the existence of some form of special expertise in other fields, but specifically to the existence of improved intuitions in other fields. Thus, our license for assuming improved intuition in philosophers would be due simply to the fact that we assume that physicists have superior physical intuitions, that mathematicians have improved mathematical intuitions, and so forth. This version is explicitly endorsed by Devitt, by Hales, and by Horvath, each of whom refers to the existence of improved intuitions in other fields. Though this version of the analogical argument might initially seem to rectify the weakness of the "basic" version, it in fact results in a much weaker analogy—and for reasons independent of the experimentalist findings demonstrating biases among philosophers. [p. 6]

Here Nado suggests that while this version of the analogical argument is more specific than the basic version, it results in a weaker analogy; weaker because unlike experts in the sciences, philosophers do not demonstrate an improved intuition (as we have seen). Thus, the mistake of the Expertise Objection is to put so much emphasis on the intuitions appealed to by experimentalists and thought experiments. Indeed, although intuitions are important in philosophy, they are never taken as definitive proof by philosophers.

If this is right, and I think it is, we can then ask where philosophical expertise might be located. Nado (2014b) suggests that the "task of evaluating a thought experiment" (pg. 9-10) and the "reflection, argumentation, and dialog" (p. 11) is how we should conceptualise philosophical expertise. We might reasonably expect that philosophers are better at evaluating arguments, constructing and interpreting thought experiments, reconciling inconsistent intuitions and beliefs, among other things. Indeed, consider cases in which thought experiments demonstrate that we hold inconsistent intuitions, for example, about what we should do in trolley case scenarios. Trolley case scenarios indicate that we experience a change in intuitions based on seemingly irrelevant factors, such as whether we are the actor or observer in a given case. Thought experiments in these cases are not designed because we think that philosophers will have faultless intuitions; they are meant to show that we *do* have error prone intuitions, which is why we need to rely on stable theories to decide what the right course of action is. In this conceptualisation of philosophical expertise, the expertise does not lie in *what* intuitions are held, but in what can be argued given the intuitions elicited.

Two further questions that arise are: what are philosophical intuitions, and what is their role in philosophy. Here I think it will be useful to elaborate on what I take 'intuition' to mean in this thesis, and what it is that experimental philosophy, in the kinds of cases discussed, is doing by eliciting the intuitions of the folk (and, as we have seen, more recently those of philosophers). The typical role of intuitions in philosophy is that they are used as evidence in some way. As evidence for a philosophical thesis, against it, as evidence against the informativeness of intuitions in general, etc. Traditionally in armchair philosophy philosophers conducted thought experiments to elicit an intuition and presented their own intuition to the case as the obvious response. Indeed, these philosophers often seem to assume that what seemed to be the case to them would be shared equally with others; that is how obvious some of their claims seemed to be to them. For example, Strawson often makes claims about the obviousness of the truth conditions for sentences containing definite descriptions:

Now it is *obviously quite* false that the phrase 'the table' in the sentence 'the table is covered with books', used normally, will 'only have an application in the event of there being one table and no more'. [(1950) pg. 332, my emphasis]

Because philosophers have traditionally considered their intuition to be shared, or universally held, that they hold the intuition seemed to be good evidence for that truth of that intuition.

Experimental philosophy challenges the assumption that traditional philosophical theses are as intuitive as it has been thought by testing whether the intuitions used as evidence are as obvious or widely held as it has been assumed. One role of experimental philosophy is thus simply to test the obviousness of such claims. That is, test whether there is uniformity in our intuitions. Another role of experimental philosophy is to analyse and systematise these intuitions. That is, to investigate intuitions systematically and in different scenarios. For this we need replications and progression among the research of experimentalists. This is one goal of the present thesis.

It is a mistake, however, to think that experimentalists consider the responses to surveys to be decisive. Consider Plato's writings in which Socrates intentionally elicits an intuition and then describes how it is wrong. Often our intuitions elicited by thought experiments are

taken to show that our intuitions are not trustworthy, or to challenge other intuitions. So intuitions at best play a role in philosophy, but one that only makes sense in the context of some philosophical dialog.

But then what is it that we are actually eliciting in these cases. There is a debate about what intuitions are. Philosophers have offered different accounts of what counts as an intuition in philosophy. Some accounts are fairly restrictive concerning what an intuition is (e.g., Cappelen gives a very restrictive account). But I think the best way to think about intuitions is not in a restrictive sense. Stich and Tobia (2015) and Williamson (2007) have each argued that restrictive accounts of intuitions do not account for all the ways we actually employ the term 'intuition'. Williamson for example states: <sup>17</sup>

Although we could decide to restrict the term 'intuition' to states with some list of psychological or epistemological features, such a stipulation would not explain the more promiscuous role the term plays in the practice of philosophy. [Williamson (2007) p. 218]

Stich and Tobia similarly endorse an inclusive account of 'intuitions'. See the following:

What do the intuitions invoked in these examples have in common? As noted earlier, when things go well, people who are asked about these cases find themselves almost immediately disposed to offer an answer, though they are not consciously aware of engaging in any reasoning that leads them to this answer. We are inclined to think that this is *all* that these cases have in common. [Stich & Tobia (2015) p. 4, emphasis theirs].

Nado (2014) has argued that what we consider to be intuitions are generated by many different mental processes. This, she argues, means that the reliability of one kind of intuition does not inform us of how reliable another kind of intuition is. Stich and Tobia take this to suggest that when we use intuitions in philosophy as evidence, we need to do so on a case by case basis. The role of intuitions will need to be more fine-grained than it has previously been considered.

Here I will be using 'intuition' as a very inclusive term to mean the responses that are given in the kind of scenarios presented by experimentalists. As Stich and Tobia comment, intuitions in this sense are regarded as the immediate response that comes to mind and

<sup>&</sup>lt;sup>17</sup> Williamson (2007) does think that experimentalists are committed to a restricted type of account though. Stich and Tobia (2015) disagree with this claim.

seems correct without conscious reasoning (or at least much conscious reasoning) on it. This kind of response might also be described as an intellectual seeming (Mizrahi, 2015).

# 1.6 Present Thesis

I hope that this discussion has validated the general program of the present thesis by addressing the Ambiguity Aversion Objection, The Objection Against Ordinary Language Philosophy, The Objection to the General Methodology of Surveys, and The Expertise Objection. From here I will employ the experimental method by testing the intuitions of non-philosophers, or *ordinary language users*, to inform on questions central to the philosophy of language tradition.

# **Chapter 2: Definite Descriptions and the Ambiguity Thesis**

## 2.1. Introduction

For more than a century now, philosophers of language have considered the analysis of simple sentences containing definite descriptions an important project within analytic philosophy.<sup>18</sup> In 1905 Russell proposed a theory of descriptions that has since been treated by some as orthodoxy. Others, though, have called it into question periodically over the years. Philosophers, such as Strawson, have questioned the truth conditions of Russell's theory. In this chapter I take up this debate again. My original contribution will involve the employment of empirical methods to test whether ordinary language users hold intuitions consistent with a Russellian theory of definite descriptions. Specifically, I will defend the following three theses:

- A. Russell's theory of descriptions is false as applied to ordinary language in English.
- B. Definite descriptions in ordinary language are ambiguous between at least three readings.
- C. This ambiguity is a semantic rather than a pragmatic ambiguity.

In Section 2.2.1 I will discuss and unpack the first of these claims. In doing so I will introduce two theories of definite descriptions; these are Russell's theory of descriptions and the Frege-Strawson Theory. I will discuss a challenge from ordinary language philosophy against Russell's theory, a challenge the Frege-Strawson Theory aims to overcome. In Section 2.2.2 I will discuss how experimental philosophy can play a role in this debate, following which I will present Study 1 as evidence for A in Section 2.2.3.

Section 2.3.1 will be concerned with claim B. it is here that I will introduce 'The Ambiguity Thesis' and distinguish it from other theories in the literature. This thesis I take to be the best explanation of the data from Study 1. I will then extend the findings of Study 1 and

<sup>&</sup>lt;sup>18</sup> A special issue of *Mind* in 2005 was entirely devoted to this debate which exemplifies its significance among philosophers.

present Study 2 as evidence for B in Section 2.3.2. I will then give a brief summary of the two studies in Section 2.3.3.

Section 2.4.1 will be concerned with claim C. Here I will present some arguments that challenge the Ambiguity Thesis, arguments for the claim that any apparent semantic ambiguity in simple definite descriptions can be explained by appealing to pragmatic considerations. I will then present two studies (Studies 3 and 4), in Section 2.4.2 and Section 2.4.3, designed to overcome this challenge to the Ambiguity Thesis.

Finally, in Section 2.5.1 I provide a general discussion of the experimental results presented here and argue that the Ambiguity Thesis has many advantages over traditional theories of definite descriptions. In Section 2.5.2 will discuss and reject some alternative explanations of the data. I will end with a brief conclusion in Section 2.5.3.

## 2.2.1 Russell's theory of descriptions is false as applied to ordinary language

The first hypothesis I will investigate is that Russell's theory of descriptions is false as applied to ordinary language. That is, Russell's treatment of definite descriptions does not capture the way we actually use definite descriptions in the English language. However, it is important to note that Russell himself was not concerned with giving a theory that was true of natural language. Russell (1957), in response to Peter Strawson's (1950) critique, makes it clear that his theory is not intended as an analysis of natural language, for the reason that ordinary language is imprecise and inaccurate. We can see this in the following passages:

This brings me to a fundamental divergence between myself and many philosophers with whom Mr. Strawson appears to be in general agreement. They are persuaded that common speech is good enough not only for daily life, but also for philosophy. I, on the contrary, am persuaded that common speech is full of vagueness and inaccuracy [Russell (1957) p. 387].

My theory of descriptions was never intended as an analysis of the state of mind of those who utter sentences containing descriptions. [p. 388]

Instead Russell is interested in the ideal language project (cf. Section 1.3), because he intends to give an account of descriptions that overcomes the defects of ordinary language and can be used as a precise philosophical tool.

Although it was not Russell's intention, I *will* be testing his theory as applied to ordinary language. I think this is warranted because, as argued in Section 1.3, the ordinary language project is insightful as it tells us something about how language actually functions (i.e., is a descriptive project). I also noted that any ambiguity uncovered here could in fact be used to support the ideal language project and Russell's rejection of ordinary language philosophy. Lastly, Russell's theory has recently been defended as a theory of descriptions in ordinary language (e.g., Kripke, 1977; Neale, 1990, 2004).

The theory I will therefore be evaluating, a theory I will for convenience refer to simply as *Russellianism*, consists of the following three theses:

(a) THEORY OF DESCRIPTIONS: A definite description with the *grammatical* form 'the *F* is *G*' does not contain a singular referring expression, but instead can be paraphrased as follows: (i) there is an *F*, (ii) there is at most one *F*, and (iii) it is *G*.

(b) UNIVOCALITY THESIS: Definite descriptions are not semantically ambiguous.

(c) ORDINARY LANGUAGE THESIS: The Theory of Descriptions is a gives the correct meaning and truth conditions of definite descriptions as they occur in natural language (as opposed to a normative claim about what they should mean, or a claim about what they mean in an ideal language).

According to (a), definite descriptions are restrictive quantifiers such that a sentence of the form 'the F is G' expresses a general proposition which has the following *logical* form:

$$\exists x(F(x) \& \forall y(F(y) \rightarrow x=y) \& G(x))$$

Thus, on Russell's analysis, a definite description asserts the existence and uniqueness of an F that also satisfies the descriptive content of *G*. Conjointly these three claims, (i)-(iii), give the truth-conditions for a sentence containing a definite description. In a case where one of

these conditions is not satisfied the entire proposition is rendered false. To illustrate, consider an utterance of the following sentence:

(1) The present King of France is bald

According to Russell, (1) is false because condition (i) is not satisfied. That is, nothing satisfies the descriptive content of 'the F' (i.e., empty definite descriptions) because there is not a present King of France.

Some philosophers, most notably Strawson (1950), have challenged (c), the ordinary language thesis, by arguing that Russellianism is not true of ordinary language. We can see this come out in the following passage:

Now suppose some one were in fact to say to you with a perfectly serious air: "The king of France is wise". Would you say, "That's untrue"? I think it's quite certain that you wouldn't. [Strawson (1950) p. 330].

Instead, the truth-conditions of Strawson's theory are intended to give an analysis true of natural language.

Although Gotlob Frege only briefly mentions definite descriptions specifically (he focuses on proper names), we can see him in 'Sense and Reference' (and 'Function and Concept'<sup>19</sup>) work with both the ideal language project and the ordinary language project. On the one hand he suggests that an empty definite description *should* be stipulated to denote 0:

In accordance with what was said above, an expression of the kind in question must actually always be assured of meaning, by means of a special stipulation, e.g. by the convention that it shall count as meaning 0 when the concept applies to no object or to more than one. [Frege (1892) fn. 9]

However, as we will see below, when analysing ordinary language he suggests that empty definite descriptions have no truth value (see Elbourne, 2013, p. 43-44). It is Frege's analysis of definite descriptions in natural language that has been of most interest to philosophers and for that reason he is generally classed alongside Strawson (see Elbourne, 2005;

<sup>&</sup>lt;sup>19</sup> See Elbourne, 2013.

Elbourne, 2013; Heim & Kratzer, 1998). For this reason I will be contrasting Russellianism with, what I call, the Frege-Strawson Theory.

The Frege-Strawson Theory looks like it gives a more natural interpretation of simple sentences containing empty definite descriptions. Consider the following case again:

(1) The present King of France is bald

While the Theory of Descriptions delivers the verdict that (1) is false, philosophers like Strawson and Frege have argued that, when used in ordinary language contexts, (1) is neither true nor false. Strawson is quite clear on this point:<sup>20</sup>

Similarly, if I ask: "Is the sentence [containing an improper definite description] true or false?" I am asking an absurd question, which becomes no less absurd if I add, "It must be one or the other since it's significant". The question is absurd, because the sentence is neither true nor false any more than it's about some object. [Strawson (1950) p. 329]

Frege is less clear on this point. However, we can get some insight about his view in the following passages:

It may perhaps be granted that every grammatically well-formed expression representing a proper name always has a sense. But this is not to say that to the sense there also corresponds a reference. The words 'the celestial body most distant from the Earth' have a sense, but it is very doubtful if they also have a reference. The expression 'the least rapidly convergent series' has a sense; but it is known to have no reference, since for every given convergent series, another convergent, but less rapidly convergent, series can be found. In grasping a sense, one is not certainly assured of a reference. [Frege (1892) p. 28]

Here we see that Frege takes empty descriptions to have no referent. Keeping in mind that Frege considers definite descriptions to be 'compound proper names' we get an idea of what Frege thinks the outcome is for sentences containing an empty definite description:

Is it possible that a sentence as a whole has only a sense, but no reference? At any rate, one might expect that such sentences occur, just as there are parts of sentences having a sense but no reference. And sentences which contain proper names without reference will be of this kind.

<sup>&</sup>lt;sup>20</sup> Here I use 'improper' definite descriptions to denote definite descriptions that are not true of exactly one entity (see Neale, 2004). This, of course, includes empty definite descriptions; however, Strawson is not discussing empty definite descriptions exclusively in this passage.

The sentence 'Odysseus was set ashore at Ithaca while sound asleep' obviously has a sense. But since it is doubtful whether the name 'Odysseus', occurring therein, has a reference, it is also doubtful whether the whole sentence has one. [Frege (1892) p. 32]

Here we see that, according to Frege, a sentence containing a proper name does not refer. Taken together with the claim that Frege considers the referent of sentences to be a truth value, Elbourne (2013) concludes that Frege must take sentences containing empty proper names (and therefore empty definite descriptions) to have no truth value (see Elbourne, 2013, p. 44).

The reason sentences like (1) are considered neither true nor false on the Frege-Strawson Theory is because containing definite descriptions are regarded as singular terms. That is, the function of a sentence containing a definite description is to refer to a specific entity and is therefore object-dependent.<sup>21</sup> Thus, a sentence containing a definite description of the form 'the *F* is G' is equivalent to '*X* is G' where the referent of *X* contributes directly to the proposition expressed. And so, 'the *F*' functions like 'that *F*'. Importantly, according to this view, the use of a sentence containing a definite description does not assert the existence of some (unique) entity, as Russellianism does, but rather presupposes it. Frege says:

If anything is asserted there is always an obvious presupposition that the simple or compound proper names used have reference. If one therefore asserts 'Kepler died in misery', there is a presupposition that the name 'Kepler' designates something; but it does not follow that the sense of the sentence 'Kepler died in misery' contains the thought that the name 'Kepler' designates something. [Frege (1892) p. 40]

In cases where this presupposition fails, such as in (1), the proposition is considered neither true nor false because the question of a truth value does not arise when there is no entity for the proposition to be true of. On this account, presuppositions are located at the level of semantics and so constitute part of the meaning of the sentence. Moreover, they do so even when the sentence is negated. That is, sentences like:

(2) The King of France is not bald.

still presuppose the existence of the King of France (see Schwarz, 2015).

<sup>&</sup>lt;sup>21</sup> Devitt calls these intensions "singular thoughts" (2004); one might also call them *de re* thoughts (see Devitt, 1996).

As we can see, cases like (1), that is, simple sentences containing empty definite descriptions, provide a pivotal point of disagreement between Russellianism and the Frege-Strawson Theory.<sup>22</sup> The same disagreement may also arise in cases of non-uniqueness, such as:

(3) The Prince of England is married.

However, the disagreement, at least about natural language speakers' intuitions, may be illusory. While many philosophers take sentences like (3) (i.e., incomplete descriptions) to favour the Frege-Strawson Theory, this is controversial. Russellian might seem to be committed to saying that (3) is false. After all, *Russellianism* holds that simple sentences containing a definite description are false when the description is satisfied by more than one entity. But (3) might be interpreted as containing implicitly restricted quantifiers, in which case native speakers would view it as true.<sup>23</sup> Moreover, while Frege took the uniqueness condition as important. Strawson differed from Frege in this respect as he took it as obvious that a definite description could be used successfully to refer to a unique individual even when more than one entity satisfied the descriptive content of 'the *F*' . Consider the following passage:

Russell says that a phrase of the form 'the so-and-so', used strictly, 'will only have an application in the event of there being one so-and-so and no more'. Now it is obviously quite false that the phrase 'the table' in the sentence 'the table is covered with books', used normally, will 'only have an application in the event of there being one table and no more'. [(1950) pg. 332]

One way to flesh out this suggestion is to maintain that the context determines which entity is referred to, but whichever it is, it satisfies the descriptive content of the description. Thus, in cases like (3), the definite description refers to a contextually salient individual, or an individual who uniquely satisfies the predicate in a

 <sup>&</sup>lt;sup>22</sup> What I will discuss in Section 2.5.1 below is that disagreements about sentences containing empty definite descriptions might only hold in cases in which the definite description is in the topic position of the sentence (otherwise Russell gets it right).
<sup>23</sup> This is sometimes called the 'implicit approach' to solving the problem of incompleteness. Some other

<sup>&</sup>lt;sup>23</sup> This is sometimes called the 'implicit approach' to solving the problem of incompleteness. Some other notable solutions are 'the explicit approach' and the 'hybrid approach' (Neale, 2005; also see Ludlow, 2013). Whether or not incomplete descriptions pose a genuine problem for Russellianism is still debated as many philosophers have challenged these solutions (e.g., Devitt, 2007; Devitt & Sterelny, 1999; Ludlow & Segal, 2004). For this reason cases of incomplete descriptions do not provide clear support for the Frege-Strawson Theory over Russellianism.

contextually restricted domain (cf. Elbourne, 2013; Heim & Kratzer, 1998; and Schwarz, 2015).

Because Russellians and Strawsonians may well give *the very same* predictions about folk intuitions regarding the truth value of statements of (3), experimentally testing folk intuitions about such cases is fraught with difficulty. For this reason, when I test between Russellianism and the Frege-Strawson Theory, I do so using the clearer cases in which nothing satisfies the description rather than cases in which more than one entity fits the descriptive content.

# 2.2.2 Experimental philosophy

The intuitions of ordinary language philosophers (e.g., Strawson) provide us with some reason to think that Russellianism is false. However, it is only by appealing to ordinary language users that we can decisively reject (or confirm) (c), the Ordinary Language Thesis; the intuitions of ordinary language users will provide better evidence for rejecting Russellianism than the intuitions of a small sample of analytic philosophers.<sup>24</sup> Thus, my aim here is to experimentally test this claim by investigating the intuitions of native English speakers.

So far there has been very little experimental work conducted about definite descriptions. Notably, a very early study mentioned by Herman Tennessen in 1960 (well before the rise of experimental philosophy) seems to have tested the intuitiveness of the Frege-Strawson Theory. Here Tennessen claims:

Strawson is wrong: Of about 1,500 informants tested in some recent experiments, no one seemed to act in accordance with Strawson's predictions. [p. 187-188]

<sup>&</sup>lt;sup>24</sup> Here it might be argued that the judgements of philosophers working in this area will be more expert and should be preferred to the judgements of ordinary language users. However, as I have argued in the previous chapter, the judgements of ordinary language philosophers are entirely relevant in this domain (see Chapter 1, Section 1.3). Moreover, because C explicitly makes a claim about ordinary language use we should be interested in ordinary language users in this instance. Also note my response to the Expertise Objection (Chapter 1, Section 1.5).

At first appearance, this study looks like it could actually provide support for Russellianism. However, Tennessen gives no reference for the study, nor does he give any other details about the specific methodology used or how the results were analyzed. This makes it understandably difficult to take much away from his claim.

More recently there have been some relevant experiments that do not rely explicitly on participant's intuitions but use reaction times or memory tasks as dependent measures. That is, they are not survey experiments. Firstly, of only tangential relevance is some experiments conducted by Chemla and Bott (2013) in which they investigated how we process sentences containing false presuppositions, particularly under cases of negation. For example, the sentence "Zoologists do not realize that elephants are birds" introduces the false presupposition that elephants are birds. They noted that participants could interpret sentences of this kind as true or false. For one, the sentence could be interpreted so that the presupposition falls outside the scope of the negation, i.e., 'Elephants are birds & NOT 'zoologists believe so'; on this interpretation the sentence would be considered false. Alternatively, the sentence could be interpreted so that the presupposition falls within the negation, i.e., 'NOT (Elephants are birds & zoologists believe so)'; on this interpretation the sentence would be considered true. To test which interpretation was more natural for ordinary language users, Chemla and Bott asked participants to make truth value judgements for a number of sentences of this form on a computer based questionnaire which also allowed them to measure reaction times. What they found was that 'false' responses were made more frequently and faster than 'true' responses. This suggests that while both interpretations are available to participants, the more natural interpretation is one in which the presupposition is interpreted independently of the operator.

In this investigation, the ordinary interpretation of simple sentences containing definite descriptions, and in particular non-referring (or non-denoting) definite descriptions is of interest. To this end it is hard to interpret Chemla and Bott's results as evidence as evidence either for or against (c), because they did not look specifically at definite descriptions. Their study does, however, suggest an alternative way of looking for evidence to adjudicate

between Russellians and Strawsonians, something that might be investigated in a future study.<sup>25</sup>

Within psychology there have been a number of memory experiments that support a presuppositional account of definite descriptions specifically. For instance, Loftus (1975) presented participants with a clip of a traffic accident. Participants were then asked a number of questions about the clip about some entity that was not in the clip (e.g., a school bus). Participants either received questions that included a definite description and so presupposed the existence of the entity (e.g. "Did you see the children getting on the school bus?"), or they received questions that included an indefinite description of the existence of the entity ("Did you see a school bus in the film?"). A week later, all participants were given the indefinite question. Loftus found that participants who initially received the questions containing definite descriptions were more likely to incorrectly affirm the existence of the non-existing entity compared to participants who initially received the questions containing indefinite descriptions (and control participants who did not receive either question initially). While these kinds of experiments do not count as evidence against Russellianism, they do indicate that ordinary language users presuppose the existence of a described entity when presented with a sentence containing a definite description, and do so to the point of forming false memories on the basis of that sentence.

A more relevant experiment that relied on reaction times was conducted by Schwarz (2015). In this experiment, schematic displays were used to test how quickly simple sentences containing definite descriptions were judged to be 'false' (i) when nothing had the property expressed by the predicate in the *subject* expression ('presupposition false'), and (ii) when nothing had the property expressed by the predicate in the *subject* expression ('presupposition false'), and (ii) when nothing had the property expressed by the predicate in the *object* expression ('predicate false'). For example, in one experiment participants were presented with sentences like 'The boy with an outing on Tuesday is going to play golf' with an image displaying a boy and a calendar with an activity printed on it where either the activity (predicate false) or the day (presupposition false) did not match the sentence. Schwarz found that participants took longer to reject sentences based on a false presupposition than sentences with a false predicate. Schwarz argues that this is evidence against Russellianism, which would not

<sup>&</sup>lt;sup>25</sup> How a time trial experiment might be able to inform on the present debate will be discussed in Chapter 5.

distinguish between the falseness of each case.<sup>26</sup> Although a presupposition account (the Frege-Strawson Theory) would say that cases of false presupposition are neither true nor false, rather than false, participants were not able to make this response. It makes sense, according to this theory, that participants' would reject these sentences if they do not have a better option available and would find it harder to reject non-referring definite descriptions.<sup>27</sup> Nonetheless, the evidence against Russellianism in this case is hardly decisive, as participants all gave an answer (namely 'false') to the 'presupposition false' cases, an answer which is consistent with Russellianism but *in*consistent with the Frege-Strawson view.

There is, however, evidence that even when given the opportunity to give responses consistent with truth valuelessness that participants still tend to favour 'false' responses. In the most relevant set of experiments for present purposes, Abrusán and Szendrői (2012) in 'Experimenting with the King of France' tested participants' truth value judgements for a number of sentences containing non-referring definite descriptions. This is the only published study that I am aware of that specifically uses a survey methodology to test the intuitions of ordinary language users about definite descriptions. They used a number of definite descriptions that lacked actual referents; such as the king of France, the emperor of Canada, and the Belgian rainforest. Each definite description was then presented in a number of different positive and negative existential sentence types; such as *is/isn't bald*, is/is not on a state visit to Australia this week, is/is not married to Carla Bruni. The aim of the study was to use experimental evidence to distinguish between different presuppositional accounts of definite descriptions<sup>28</sup> and so in doing so they assumed something like the Frege-Strawson Theory is correct, and did not explore Russellianism in any detail. Interestingly, what they found was a strong tendency for participants to give 'false' responses to the positive existential statements, for example, almost 80% of responses to sentences like 'The king of France is bald' were 'false', compared to less than 20% of 'can't say' responses (which were considered indicative of truth valuelessness). These findings seem to lend support for Russellianism; however, Abrusán and Szendrői

<sup>&</sup>lt;sup>26</sup> I.e., between cases of no *F* compared to cases of is not *G*.

<sup>&</sup>lt;sup>27</sup> It might be that some version of Russellianism can account for the difference in reaction times between cases in which the *F* is not *G* and cases in which there is no *F*. For example, if the proposition (iii), that whatever is *F* is *G*, is assessed first.

<sup>&</sup>lt;sup>28</sup> A discussion of these variations of the Frege-Strawson Theory will be given in Section 2.5.1.

discard this suggestion due to the responses to the negative cases. Here the results were less clear-cut as responses varied extensively depending on the sentence type. Even so, a substantial proportion of participants gave 'false' responses to all sentence types (between 15-50%) compared to the proportion that gave 'can't say' responses (between 9-34%). For example, for the sentences of the form 'X isn't bald', 44.9% of participants responded 'false', 33.9% responded 'can't say', and 21.3% responded 'true'. What I will argue is that this kind of variation in intuitions is good evidence for an ambiguity account of definite descriptions. What Abrusán and Szendrői argue, however, is that something like the Frege-Strawson Theory (Strawson or some other truth-value gap theory for them)<sup>29</sup> is the correct semantic theory of definite descriptions, and the variation from the predicted 'can't say' responses should be given a pragmatic explanation, such as those considered in various presupposional theories (e.g., Laserhohn, 1993; von Fintel, 2004). Von Fintel (2004), for one, defends a variant of the Frege-Strawson Theory in which all sentences containing nonreferring definite descriptions are truth valueless at the semantic level, however, they may be interpreted as true or false based on independent information which means the presupposition failure can be bypassed or ignored. So, 'The king of France is married to Carla Bruni' can be rejected based on independent knowledge about Carla Bruni (namely that she is married to Nicholas Sarkozy) and so knowledge of the lack of a French monarchy is not needed. It is this possibility of a pragmatic interpretation that I am going to try to rule out experimentally in Section 2.4.2 and Section 2.4.3.

Abrusán and Szendrői go on to suggest that the high proportion of 'false' responses could be due to participants rejecting the sentence, or interpreting 'false' to mean 'not true', rather than them interpreting the sentence itself as false. It could also be suggested that a problem lies with their use of 'can't say' as an indicator of truth valuelessness. Participants possibly did not consider this as an appropriate response to capture their truth valueless intuitions; indeed, they might have interpreted 'can't say' to mean something like 'I do not understand'. If this was the case, participants might have opted for 'false' as a best alternative option. To this end, it would be better to use a label such as 'neither true nor

<sup>&</sup>lt;sup>29</sup>Abrusán and Szendrői (2012) consider three different truth value gap theories: Strawson (1971), Laserhohn (1993), and von Fintel (2004). These theories will be discussed in Section 2.5.1.

false' in order to track intuitions of a truth-value gap.<sup>30</sup> It is these kinds of methodological problems with surveys, discussed in Section 1.4, which I will try to rule out in Studies 3 and 4.

The aim of the present studies is to expand on this experimental work about definite descriptions by testing the hypothesis that Russellianism is false as applied to ordinary language. The results will hopefully be clearer and less equivocal than previous studies. Given the previous literature, it is expected that the data will provide substantial support in favour of Russellianism, with some variation. Indeed, Tennessen's allusion to a study in which participants did not give responses in line with the Frege-Strawson Theory, as well as the high proportion of 'false' responses in Abrusán and Szendrői's experiment, give us some basis to suspect that a majority of participants will have Russellian intuitions. However, whether there will be a consensus for Russellianism is unclear. I will consider the results to support the hypothesis that Russellianism is false if there is up to 20% disagreement with this theory. At this threshold of disagreement we can hold that there is not a consensus, even if there is a majority of support for Russellianism.

# 2.2.3 Study 1: The Basic Survey

*Participants:* Fifty-four undergraduate students from Victoria University of Wellington, New Zealand participated voluntarily during their normal lecture time. Females made up 58.5% of the sample.<sup>31</sup> Forty-nine participants indicated that they were native English speakers; thirty-five participants identified as being of NZ-European descent, four as Maori, one as Pacific Islander, nine as Asian, and ten as 'other'.<sup>32</sup> Only five participants indicated that they had previously taken a philosophy course.

*Survey and Procedure:* The final survey comprised five cases, three test cases and two control cases.<sup>33</sup> For each case, participants were asked to make a judgement about the truth

<sup>&</sup>lt;sup>30</sup> Abrusán and Szendrői (2012) do discuss this point (p. 12-13).

<sup>&</sup>lt;sup>31</sup> One participant identified as 'other'.

<sup>&</sup>lt;sup>32</sup> Some participants identified with more than one ethnic group.

<sup>&</sup>lt;sup>33</sup> In an initial pilot study a sixth case was also included that resembled (S4) above. However, for the reasons discussed in Section 2.2.1, that cases of non-uniqueness may be hard to interpret, this case was cut. The other questions were also adapted from the initial pilot study, so the results of that study are not included here.

value of a sentence of the form 'the F is G'. For the three test cases, the sentence contained an empty definite description. These sentences were the following:

(S1) The present king of France is bald.

- (S2) The President of England is wise.
- (S3) The Chinese Pope is courageous.

Participants in the US sample only received (1) and (2) while participants in the New Zealand sample received all three sentences. The two control cases contained referring definite descriptions; in one of the cases the predicate was false of the actual referent:

(S4) The present queen of England is a teenager.

In the other control case the predicate was true of the actual referent:

(S5) The present Prime Minister of New Zealand is male.

The control cases were included to detect for participants answering randomly as well as to separate the test cases as distractors. The idea with separating the test cases was that participants would be forced to consider each case independently instead of simply that they should give the same response to all the test cases. The order of the questions was always the same: (S4), (S1), (S5), (S2), (S3). The available responses were: 'True', 'False', 'Neither', and 'Both'. For the test cases, 'False' was considered indicative of Russellianism, and 'Neither' was considered indicative of the Frege-Strawson Theory. 'Both' was included to account for the intuition that there is a sense in which a sentence was true and a sense in which it was false.<sup>34</sup>

For each case a corresponding presupposition question was also included to check that participants knew whether or not a definite description was referring (i.e., whether or not there was a presupposition failure). These questions simply asked whether there was the relevant *F*; for example, "Is there a present king of France?" For participants in the New

<sup>&</sup>lt;sup>34</sup> 'Both' was included because of some arguments made by Hanuong et al. (2013) in their paper 'Mind the (truth-value) gap'. Here they criticised experiments which did not include truth value gaps and gluts (both true and false) as possible responses (in particular Chemla and Bott's, 2013, study). 'Both' does make an interesting response in Study 2.

Zealand sample, each presupposition question directly preceded its corresponding case. For the US sample this order was adapted so that all of presupposition questions came at the end. The reason being was that it was thought that the presupposition questions could bias intuitions on the corresponding cases. Because participants from the US sample had no way to change their responses after they were made their answers were not biased by the presupposition questions.

*Results:* For each case, only the data from participants who had correctly answered the corresponding presupposition question was included for final analysis. That is, only participants who had responded "No" to the corresponding presupposition question for the test cases, and "Yes" for the control cases were included in the analysis of each case. Overall, the responses to the two control cases were very accurate. The proportion of participants who correctly answered the presupposition questions that corresponded to each control question was 90.7% and 98% for (S4) and (S5) respectively. Of those participants, the proportion who correctly answered each case was 94.9% for (S4) (answered "False") and 100% for (S5) (answered "True").

The proportions of participants included in the final analyses of the test cases were 48.1%, 83.3%, and 70.4% for cases (S1), (S2), and (S3) respectively. Across all three test cases, the proportions of responses were 1.8% for 'True', 34.9% for 'False', 58.7 for 'Neither', and 4.6% for 'Both'; for the proportions of responses to each test case see Figure 1 (this is also shown in Table 1 below). A Kruskal-Wallis test indicated no significant difference between responses to each test case [p = .604]. Only 23 participants (42.6%) correctly answered all of the presupposition questions and were included for analysis on all three test cases. Of these participants, 17 (73.9%) answered consistently across the three test cases (i.e., gave the same response to each case); 6 responded 'False' to all test cases and 11 responded 'Neither' to all test cases.



Figure 1. Proportions of responses given to each test case.

These results support the hypothesis that Russellianism is false because a substantial proportion of participants gave a response consistent with the Frege-Strawson Theory. A binomial test found that the proportion of 'Neither' responses on the test cases was significantly greater than a 'noise' threshold set conservatively at 20% (p < .001). Indeed, as we can see from Figure 1, about 60% of responses were consistent with the Frege-Strawson Theory. Theory.

*US sample follow-up*: Additional data was collected online. Ninety-eight United States residents participated voluntarily as part of an online free personality test.<sup>35</sup> In this sample the mean age of participants was 32.09 (SD = 12.25) and included 35 males and 63 females. All participants indicated that they were competent with English.

This follow-up survey was conducted and analysed in the same way as for the NZ sample except for two changes that were made to the survey questions. Firstly, this sample did not receive (S3); thus, only two test questions [(S1) and (S2)] were used. Secondly, the original sentence (S5) was adapted to 'The present President of the United States is male' to be more relevant for US participants.

The proportions of participants included in the final analysis were 34.7% and 75.5% for (S1) and (S2) respectively. Across the two test cases, the proportion of each response were 19.9% for 'True', 28.6% for 'False', 45.4 % for 'Neither', and 6.1% for 'Both'. Table 1 shows the proportions of responses to each test case, as well as for the NZ sample for comparison.<sup>36</sup> A Mann-Whitney U test indicated a significant difference between responses to each test case [U = 985.5, p = .047]. We can see that this is due to a lower proportion of 'Neither' responses to Case (S1) (and a greater proportion of 'False' responses) compared to Case (S2). Only 29 participants (29.59%) correctly answered both of the presupposition questions and were included for analysis on both test cases. Of these participants, 16 (55.17%) answered consistently across the two test cases (i.e., gave the same response to each case); 6 responded 'False' to both test cases and 10 responded 'Neither' to both test cases.

<sup>&</sup>lt;sup>35</sup> Although these participants are all most likely US residents it is possible that persons outside the US accessed the survey. Note all participants included indicated that they were competent English speakers. <sup>36</sup> While a direct comparison between the NZ sample and the US sample might not be warranted due to relevant differences between the surveys, for interest, Mann-Whitney U tests did not find any difference in responses between each sample [p = 170 and p = 634 for (1) and (2) respectively].

		Case (S1)	Case (S2)	Case (S3)
NZ	TRUE	0	4.4	0
	FALSE	42.3	28.9	36.8
	NEITHER	57.7	57.8	60.5
	BOTH	0	8.9	2.6
US	TRUE	11.8	12.2	
	FALSE	44.1	23.0	
	NEITHER	44.1	56.8	
	BOTH	0.0	8.1	

Table 1. shows the proportions of responses NZ and US samples gave to each of the three test cases: the King of France case, the President of England case, and the Chinese Pope case.

Again, the results of this follow-up study support the hypothesis that Russellianism is false as most participants gave responses consistent with the Frege-Strawson Theory. This is borne out again as the proportion of 'Neither' responses was significantly greater than 20% (p < .001). While there was a smaller proportion of 'Neither' responses on Case (S1) than Case (S2) (as well as compared to the proportions of 'Neither' responses in the NZ sample, see Table 1), looking at this case in isolation the proportion of 'Neither' responses is still greater than this noise threshold (p < .001).

*Discussion*: The results of this study support the hypothesis that Russellianism is false. Indeed, the expectation that a majority of responses would support this theory was not borne out. Overall, the results indicate the following three things. (a) If there is a bias in
intuition, it is in favour of the Frege-Strawson Theory as approximately 60% of responses were consistent with this theory. We can see this by the proportions of 'Neither' responses in Table 1. This finding is not consistent with Tennessen's claim, that no one holds Strawsonian intuitions in these kinds of cases. If we take this claim as: there will be less than 20% of responses in favour of the Frege-Strawson Theory (to account for noise, or performance error), we see that this claim is not confirmed by the present data. (b) Even though the majority of responses support the Frege-Strawson Theory, there is still substantial support for Russellianism. We can see this by the proportions of 'False' responses which fluctuate between 23% and 44% for the different cases. Moreover, on Case (S1) the US sample actually indicated equal support between Russellianism and the Frege-Strawson Theory. It is this finding which suggests that definite descriptions are ambiguous between the two theories. (c) This ambiguity is further borne out by the observation that, of the participants whose data was included in all final analyses, approximately 26% of the NZ sample and 45% of the US sample responded differently across the test cases.<sup>37</sup> The finding of variation within individual participants suggests that participants found that different readings of 'the F is G' the most appropriate in different cases and that individual participants were comfortable switching between different readings. That simple sentences containing definite descriptions are ambiguous between at least these two readings will be discussed more fully in the next section.

One potential worry is that some participants responded 'True' to test cases, especially in the US sample. Considering that these participants had responded 'No' to the presupposition questions, it is hard to account for these responses. One possibility is that participants who made these responses were considering the counterfactual case in which there was an individual who satisfied the definite description. If this was the case then it is perhaps not so surprising that they would think that *if* there was a King of France then it could be true that he would be bald, or, that *if* there was a President of the United Kingdom then they would be wise.

<sup>&</sup>lt;sup>37</sup>It is notable that a greater proportion of the NZ sample answered consistently compared to the US sample considering that participants in the NZ sample were given three test cases and participants in the US sample were only given two.

#### 2.3.1 Definite descriptions are ambiguous between at least three readings

In this Section I will present the Ambiguity Thesis as the best explanation of the results of Study 1. In doing so, I will distinguish the Ambiguity Thesis from other types of ambiguity theories and other types of 'hybrid theories' that are discussed in the philosophy of language literature. Moreover, I will present Study 2 that extends the results of Study 1 to give evidence for a third reading of definite descriptions, namely a conceptual reading.

In this debate philosophers have tended to come down on one side; that is, they hold that either Russellianism (e.g., Bach, 1981,1987; Kripke, 1977; Ludlow & Neale, 1991<sup>38</sup>; Neale, 1990; Salmon, 1991) or the Frege-Strawson Theory (e.g., Atlas, 2004; Elbourne, 2013; Laserhohn, 1993; Schoubye, 2009; von Fintel, 2004) is the correct *univocal* theory of definite descriptions.<sup>39</sup> Certainly, we can see substantial support in favour of each theory. Notably, however, some philosophers have insisted that definite descriptions are ambiguous between the two analyses (e.g., Amaral, 2008; Devitt, 2004; Reimer, 1998; Wettstein, 1981). We can see this come out in the following passage(s):

According to RD [referential descriptions], it is similarly a semantic not pragmatic fact that a certain description token refers to the person the speaker has in mind and that the containing utterance depends for its (literal) truth on that reference. And it is a semantic fact about the description type that it has such a use as well as its use as a quantifier. [Devitt (2004) p. 2 fn. 5]

Definite descriptions are ambiguous in the sense that they can be used (literally) in sentences of the form *The F is G* to express singular or general propositions, depending upon the communicative intentions with which they are used. When used attributively, they function semantically as generalized quantifiers (roughly in accordance with Russell's Theory). The proposition literally expressed will be (roughly) of the form: *There is exactly one F and whatever is F is G.* In such cases, the (univocal) linguistic meaning of the definite description (involving uniqueness with respect to *F*-hood) contributes directly to the proposition literally expressed. When used [directly] referentially, definite descriptions function semantically as *indexicals.* The proposition expressed will be of the form: *x is G*, where the identity of *x* will vary with the

<sup>&</sup>lt;sup>38</sup> Here Ludlow & Neale generalise a Kripkean defence of Russellianism to indefinite descriptions. Thus they claim that referential uses of definite descriptions and indefinite descriptions should not be considered significant at the semantic level.

<sup>&</sup>lt;sup>39</sup> There are dissimilarities in the views of the philosophers that I have classed as supporters of the Frege-Strawson Theory. All these theories are in line with the general theme of the Frege-Strawson Theory and hold that sentences that contain empty definite descriptions are truth valueless (or neither true nor false) at the semantic level. Some of these accounts will be discussed in more depth in Section 2.5.1.

context of utterance. In such cases the object or individual referred to (*x*) contributes directly to the proposition literally expressed, provided it satisfies the (univocal) linguistic meaning of the definite description: provided it is the (contextually) unique *F*. Thus, the contribution of the linguistic meaning in such cases in merely *indirect*. [Reimer (1998) p. 93]

In the above passage, Marga Reimer suggests that definite descriptions can be used in the manner of Russellianism, as generalised quantifiers, or in the manner of the Frege-Strawson Theory, as directly referential expressions. Here we can see that Reimer also seems to take it that 'the *F*' only needs to be contextually unique, a point I discussed earlier in Section 2.2.1. Not only does Reimer note that definite descriptions *can* be used with these two functions, she is clear that both of these uses provide a different semantic basis for a sentence containing a definite description. That is, they express different propositions.

I believe the results of Study 1 are best supported by this kind of account. Study 1 provides evidence for the hypothesis that Russellianism is false because it indicates that Russellian intuitions are not the consensus (i.e., there is not less than 20% disagreement with the predictions of Russellianism). In fact there was a *greater* proportion of intuitions consistent with the predictions of the Frege-Strawson Theory, a finding which was not expected given the previous experimental work. However, the results do not support the Frege-Strawson either; most notably because there was still substantial support for both Russellianism (up to 44%) as well as the Frege-Strawson Theory (up to 60%). In light of these results I propose that the best explanation is captured by the Ambiguity Thesis, which can be stated as follows:

**The Ambiguity Thesis:** simple sentences containing definite descriptions, of the form 'the *F* is *G*', are semantically ambiguous and have multiple meanings associated with them. A native speaker can use a definite description either with the content suggested by Russellianism or as suggested by the Frege-Strawson Theory depending on their communicative intention; no one disambiguation is correct outside of a context of utterance.

The Ambiguity Thesis accounts for the variation uncovered in Study 1 in terms of participants disambiguating the sentences containing definite descriptions differently. Some participants, given the context, interpreted the cases according to Russellianism and some

participants interpreted the cases according to the Frege-Strawson Theory. Notably, some participants disambiguated the different cases in different ways (these are the participants who did not respond consistently across the cases). The Ambiguity Thesis accounts for this variation as well as these participants finding one interpretation more natural in some cases and the other interpretation more natural in other cases due to a difference in context. Thus, the Ambiguity Thesis provides an explanation for all three of the main findings of Study 1 [i.e., (a)-(c) above].

The Ambiguity Thesis needs to be distinguished from some other ambiguities that have been noted by philosophers of language. One kind of ambiguity that is often mentioned in this kind of context is the notion of a scope ambiguity. A scope ambiguity occurs in natural language when it is unclear whether an operator takes wide or narrow scope over a quantifier (or similar expression) in the language. Russell famously exploits this idea when explaining how a negative existential like:

(4) The Present King of France does not exist.

can be true despite the fact that it looks as though it could not be (because the subject expression does not pick out anything). Russell explains away the appearance of paradox by pointing out that (4) is ambiguous in English, depending on whether we give the negation operator wide or narrow scope. If it is given narrow scope, as in:

(4a) The Present King of France is such that he does not exist.

then it is false, but if the operator is given wide scope, as in:

(4b) It is not the case that the Present King of France does exist.

then it is true. The kind of ambiguity I am proposing here is *not* related to scope ambiguities. The ambiguity I am proposing arises even in *simple* sentences containing a definite description.

Another kind of ambiguity is noted by Kripke (1980) and Soames (2002), who suggest that if proper names are directly referential expressions they can still have different semantic contents, specifically when different people have the same name. So, on this view,

'Aristotle' is ambiguous between the ancient philosopher Aristotle and Aristotle Onassis the shipping magnate (cf. Kripke 1980: 8). If definite descriptions are referring expressions, as Strawson supposes they are, a similar kind of ambiguity might arise for descriptions. This is not the kind of ambiguity I am proposing here. Indeed, according to the Ambiguity Thesis I advocate, a definite description can be ambiguous even if it turns out that all disambiguations pick out or have the very same referent.

Finally, the Ambiguity Thesis needs to be distinguished from a related distinction made by Donnellan (1966). Donnellan's suggestion was that definite descriptions function differently in different contexts. Sometimes they function *attributively*, that is, in the Russellian manner, by picking out whomever or whatever fits the description 'the *F*'. For instance:

Consider the sentence, "Smith's murderer is insane." Suppose first that we come upon poor Smith foully murdered. From the brutal manner of the killing and the fact that Smith was the most lovable person in the world, we might exclaim, "Smith's murderer is insane." I will assume, to make it a simpler case, that in a quite ordinary sense we do not know who murdered Smith (though this is not in the end essential to the case). This, I shall say, is an attributive use of the definite description. [Donnellan (1966) p.285]

Other times definite descriptions function *referentially* to pick out a particular entity that a speaker has in mind. For instance:

Suppose that Jones has been charged with Smith's murder and has been placed on trial. Imagine that there is a discussion of Jones's odd behaviour at his trial. We might sum up our impression of his behaviour by saying, "Smith's murderer is insane." If someone asks to whom we are referring, by using this description, the answer here is "Jones." This, I shall say, is a referential use of the definite description. [Donnellan (1966) p. 286]

Although similar to the Frege-Strawson Theory, this referential sense differs in an important respect. While Frege, Strawson, and Donnellan all regard definite descriptions as referring expressions, a referential use of a definite description, according to Donnellan's characterisation, can still function successfully even if the intended referent does not satisfy the descriptive content 'the *F*'. Consider the following:

But in the second case, where the definite description is simply a means of identifying the person we want to talk about, it is quite possible for the correct identification to be made even though no one fits the description we used. [Donnellan (1966) p. 286] Suppose one is at a party and, seeing an interesting-looking person holding a martini glass, one asks, "Who is the man drinking a martini?" If it should turn out that there is only water in the glass, one has nevertheless asked a question about a particular person, a question that it is possible for someone to answer. [Donnellan (1966) p. 287]

The Frege-Strawson Theory, however, is explicit in holding that the intended referent *does* need to satisfy the descriptive content of 'the *F*' to be used felicitously.

The best way to characterise the difference between Donnellan's distinction and the Ambiguity Thesis might be to take Donnellan to be merely giving a pragmatic distinction. Indeed, while some philosophers' people take Donnellan to be endorsing a semantic ambiguity, it is not clear that he is.<sup>40</sup> For example, although Donnellan seems to be right that a definite description can be used successfully to communicate something about an intended referent, even in cases of misdescription, it does not seem that we would say that such a proposition was true. So a speaker who utters 'the man drinking the martini is such and such' *can* communicate something about the intended referent even if the intended referent is, in fact, drinking water, but we would not say that the speaker *actually* says something true. The Ambiguity Thesis, on the other hand, is explicitly endorsing a semantic ambiguity in which the different readings of a sentence containing a definite description offer different truth conditions. In cases of misdescription, the Frege-Strawson Theory would regard any proposition as truth valueless while Russellianism would regard them as false. Whether the Ambiguity Thesis can be given a similar pragmatic interpretation will be investigated in the second half of this chapter.

Apart from other types of ambiguities, the Ambiguity Thesis also needs to be distinguished from some other types of 'hybrid theories'—theories in the close vicinity that have the potential to be confused with an ambiguity view. Many philosophers exploring issues in the philosophy of language explicitly consider a hybrid theory and contrast it with a 'pure' or univocal theory of meaning, according to which there is just one element that counts as *the* content, semantic value or meaning of a term. A hybrid theory of meaning, by contrast, makes the following claim:

<sup>&</sup>lt;sup>40</sup> Reimer (1998), for example, states "according to Donnellan, the distinction is *semantically* significant: for what is *meant* in each case corresponds to what is (literally) *expressed* in each case."

**Hybrid Theory of Meaning:** There are two or more semantic elements associated with an expression A, which work together in some way to determine A's referent or extension in any possible world.

There are many different ways a hybrid view might be understood, and it is important to be clear and precise about which view one has in mind when evaluating it (in later chapters, especially Chapter 4, I will note that experimental philosophers often fail to do this, and as a consequence draw unwarranted conclusions from their data). Here are three possible ways a theory might count as a hybrid theory about a term A:

- Ambiguity View. If a term A is ambiguous, then there are straightforwardly at least two different meanings associated with A in a language. On any occasion of utterance, A will have just one of those meanings associated with it, and which one will be determined by the context. This is the view about definite descriptions, proper names and natural kind terms advocated in this dissertation.
- 2. Pluralist View. On the pluralist view, a term A has multiple meanings associated with it on any occasion of utterance. The term is not (or at least need not be) ambiguous on this conception because the single term can consistently have the same contents associated with it on every occasion utterance. This kind of view is advocated by—or at least consistent with—David Chalmers' two-dimensional semantics.
- 3. **Component View.** Unlike on the ambiguity and pluralist views, the mixed view tells us that a term A might have a *single* meaning or semantic value associated with it, but that different components work together to determine what that meaning is. If you think the (single) meaning of an expression is captured by a conceptual analysis, you will accept that there are simpler concepts that can be thought of as *components* of the meaning that when conjoined make up the meaning of the expression.

In order to give a full defence of the Ambiguity Thesis, we need to be sure that the experimental data not only rules out a pure univocal theory, but also (i) rules out a pluralist theory and (ii) cannot be fully accommodated by a component view.

A pluralist theory about definite descriptions holds that sentences of the form 'the *F* is *G*' have at least two meanings, or semantic values, all of which are expressed in every token use of the sentence. Murday (2014) has recently offered this kind of theory specifically with regards to simple sentences containing definite descriptions. According to Murday, all sentences of the form 'the *F* is *G*' simultaneously express two propositions, a singular proposition (similar to the Frege-Strawson Theory) and a Russellian general proposition. In support of this theory Murday offers two arguments. Firstly, he applies the 'modal profile test'.<sup>41</sup> He imagines a case in which A utters:

(7) The CEO of Microsoft is a ninny.

because A believes CEOs to generally be ninnies, not because A knows who the CEO of Microsoft is or anything about them. Keeping in mind that the actual CEO of Microsoft is Steve Ballmer, we are asked to imagine a world  $\beta$  in which Steve Ballmer is neither CEO of Microsoft nor a ninny. In  $\beta$  Bill Gates is the CEO of Microsoft and he is a ninny. Murday asks whether A has said something true of  $\beta$  by (7) and claims that our intuition is mixed as to whether something true or false is said— the CEO is a ninny, but Steve Ballmer is not. The reason for our ambivalence, he argues, is because two propositions are expressed, one that is true and one that is false. Thus, according to Murday it is our ambivalence that is evidence for a pluralist theory and against an ambiguity theory about definite descriptions.<sup>42</sup> A second argument Murday gives appeals to 'indirect reports'. He again imagines that A utters (7) even though A does not know who the CEO of Microsoft is. In this case, Murday argues that it would be accurate to say to Steve Ballmer 'A said you are a ninny' even though A would not say 'Ballmer is a ninny'. Murday argues that the reason that both of these statements are legitimate is because both the singular proposition, that Steve Ballmer is a ninny, and the general proposition, that whoever is the CEO is a ninny, is expressed by (7).

This differs to an ambiguity theory which holds that while there are multiple semantic contents that a definite description could have, only one is expressed by a token use of a

<sup>&</sup>lt;sup>41</sup> This is the test Kaplan (1989) uses to show that sentences containing indexicals express a non-rigid 'character' proposition and a rigid proposition. Murday draws some parallels between Kaplan's account of indexicals and definite descriptions.

<sup>&</sup>lt;sup>42</sup> This argument considering counterfactual worlds is related to another pluralist theory, two-dimensional semantics. I will discuss this theory in Chapter 4 when I consider a pluralist explanation of my data about natural kind terms.

sentence containing a definite description. Thus, an ambiguity theory would say that (7) is either true *or* false (or neither true nor false), it is not both simultaneously. Whether pluralism can provide a better explanation of the present data will be considered in Section 2.5.2.

Alternatively, we have component theories. According to a component theory, there are at least two different and independent factors that together determine the meaning of a term or phrase—for example, its referent and its descriptive content.<sup>43</sup> In the case of definite descriptions the relevant competing component theory claims that the meaning of a definite description is somehow determined by *both* the singular reference and the general descriptive content associated it.

## 2.3.2 Study 2: The Conceptual Survey

I have argued that the results of Study 1 are explained by appealing to the Ambiguity Thesis. The aim of Study 2 is to expand on Study 1 by testing for an alternative reading of definite descriptions, namely a conceptual reading. It became apparent that all of the cases used in Study 1, as well as the examples discussed in the relevant literature, considered definite descriptions in which the predicate expression was not conceptually tied to the subject, for example, applying the predicate 'baldness' to a non-existent King of France. If, however, the subject and the predicate expressions were conceptually related, it might be that ordinary language users have different intuitions. For instance, cases that involve potential analytic truths like 'the present king of France is a king' might be considered true on some interpretation. Moreover, cases that involve apparent contradictions like 'the present king of France is a commoner' might be considered false. This possibility is not accounted for by either the Frege-Strawson Theory or Russellianism, which each predict that participants will have the same intuitions (either false or truth valueless) in all of these cases.

Thus, in Study 2 the conceptual reading of definite descriptions will be considered alongside Russellianism and the Frege-Strawson theory. According to the conceptual reading, a

<sup>&</sup>lt;sup>43</sup> Note that this kind of theory is often called a 'hybrid theory'. I have changed some of the terminology to clear up some confusion about these theories.

sentence of the form 'the *F* is *G*' is interpreted as meaning 'The concept of *F*-ness contains the concept of *G*-ness'. The conceptual reading thus holds that the truth value of a sentence can be ascertained purely by knowing the logical relationships that hold between the concepts appealed. Specifically, it holds that a sentence containing a definite description is true if and only if the predicate applied to the definite description is analytically true of it. Thus 'analytic' sentences will be judged true on this reading, and 'contradictory' sentences will be judged false.

The rationale for Study 2 is twofold. Firstly, if definite descriptions are ambiguous, as Study 1 suggests, then it would be interesting to discover the extent to which this ambiguity exists. Secondly, if there is a third alternative reading, then previous evidence might have conflated the evidence for Russellianism and a conceptual reading. This is because previous experiments have taken an intuition that a sentence is false if it contains an empty definite description as evidence for Russellianism (as opposed to the Frege-Strawson Theory). Such an inference commits the fallacy of the false dichotomy if there is a third conceptual reading distinct from the Russellian reading that might account for our intuitions that such statements are false. It might actually be that responses that have been considered indicative of Russellianism are produced by this conceptual interpretation. (After all, it is false that the concept of *King-of-France*-ness contains the concept *Bald*-ness).

*Method:* The method was similar to Study 1. A new sample of participants was collected which comprised 158 undergraduate students (57.6% female, 41.1% male).<sup>44</sup>

Again, there were two types of questions: presupposition questions and test questions. On the test questions participants were asked to make a judgement about the truth value of a sentence that contained one of the following three empty definite descriptions: *The present king of France, The Australian golden mountain, The city at the North Pole.* There were three types of sentences that these definite descriptions appeared in: contingent sentences which applied an unrelated predicate to the definite description (like the cases in Study 1), analytic sentences which applied a predicate that was analytically true of the definite description, and contradiction sentences which applied a predicate that contradicted the definite description. This resulted with 9 different sentences (see Table 2 for full list of the

<sup>&</sup>lt;sup>44</sup> Two participants identified as 'other'.

sentences). There were three versions of the final survey; each version contained three test questions, one of each of the 3 definite descriptions and one of each of the 3 sentence types. Each participant received only one version of the survey. Test questions were always directly preceded by their corresponding presupposition question.

Table 2. shows a list of all the test sentences for all sentence types.

Analytic truths:

The present king of France is a king

The Australian golden mountain is a mountain

The city at the North Pole is a city

Contradictions:

The present king of France is a commoner (i.e., not royal)

The Australian golden mountain is made entirely of bronze

The city at the North Pole is a shoe

## Contingencies:

The present king of France is bald

The Australian golden mountain is wet

The city at the North Pole is beautiful

For Study 2, the predictions made by each possible reading of the test sentences were more complex. Russellianism predicts that the sentences are false while the Frege-Strawson Theory holds that they are neither true nor false, regardless of sentence type. The conceptual reading predicts that cases of contingencies and contradictions are false (like Russellianism), but that for cases of analytic truths the sentences are true (see Table 3. for a summary of the predictions of each theory).<sup>45</sup> Thus the analytic sentences can be used to distinguish between the three readings.<sup>46</sup>

	Conceptual	Frege-Strawson	Russellianism
Analytic truths	TRUE	NEITHER	FALSE
Contingencies	FALSE	NEITHER	FALSE
Contradictions	FALSE	NEITHER	FALSE

Table 3. shows the predictions each theory makes for each sentence type.

<sup>&</sup>lt;sup>45</sup> It has been suggested to me that the conceptual reading should opt out of making a prediction in cases of contingencies, perhaps by being completely silent on the matter or by giving the Frege-Strawson Theory 'neither true nor false' response. However, because I have characterised the conceptual reading as being true only in cases of analytical truths and false otherwise, the contingent cases will be regarded as false according to this reading.

<sup>&</sup>lt;sup>46</sup> Initially there were two surveys that constituted this experiment. The other survey was exactly the same, however, each test sentence contained the clause 'whoever/whatever they/it might be'; for example, 'The present king of France, whoever they might be, is a king'. The reason for this survey was that Donnellan (1966) notes that clauses of this kind are indicative of an attributive use. Thus, an additional test was considered to see if this 'Donnellan' survey resulted in more Russellian responses than the survey without this additional clause. However, this survey was not included in the final data analyses (and is not included in this report of the methodology) because the predictions of this survey were not so clear when considering both a conceptual reading and a Russellian reading.

Evidence for the Ambiguity Thesis will be borne out by a mix of responses; that is, by there being significant proportions of responses that are consistent with the predictions of each reading. In Study 1 we saw that there was at least of 20% of responses in favour of each of Russellianism and the Frege-Strawson Theory. It was this finding that indicated that an ambiguity is present. Similarly, in Study 2 we should predict that, if the Ambiguity Thesis is correct, at least in some sentence type there should be at least 20% of responses in favour of each of the readings. For example, on the contingent and contradiction sentences we should expect a similar breakdown of responses to Study 1 between 'neither' and 'false' responses. On the analytic truth sentences we should expect to see significant proportions of each of 'true', 'false,' and 'neither'. On the other hand, we would see evidence of a pluralist view (rather than an ambiguity) if there was a significant amount of 'both' responses. As noted above (Section 2.3.1) a pluralist view claims that there are multiple propositions expressed by sentences containing definite descriptions. This was not indicated in Study 1 in which there was only a small proportion of 'both' responses (<10%). However, if there are at least 20% of 'both' responses uncovered in this experiment, it will be taken as some evidence for a pluralist view.

*Results and Discussion:* The results were analysed similarly to Study 1. The proportions of participants who correctly answered 'No' to the presupposition question and were included in the final analyses were 63.9%, 30.4%, and 67.7% for the King of France case, the Australian golden mountain case, and the City at the North Pole case respectively.<sup>47</sup> The proportion of 'True', 'False', 'Neither', and 'Both' responses were then calculated for each test question. Table 4 shows the proportion of responses made to each of the three cases for each sentence type.

<sup>&</sup>lt;sup>47</sup> This comes to 101, 48, and 107 participants for each of the cases respectively.

		Analytic truths	Contingencies	Contradictions
King of France	TRUE	17.1	0	6.3
	FALSE	34.3	50	28.1
	NEITHER	25.7	50	50
	BOTH	22.9	0	15.6
Australian golden mountain	TRUE	46.2	0	4.8
	FALSE	23.1	35.7	71.4
	NEITHER	23.1	50	14.3
	BOTH	7.7	14.3	9.5
City at the North Pole	TRUE	14.7	2.9	5.3
	FALSE	44.1	57.1	65.8
	NEITHER	32.4	31.4	23.7
	BOTH	8.8	8.6	5.3

Table 4. shows the proportions of responses for each sentence.

For the analytic truth sentences and the contingent sentences, Kruskal-Wallis tests indicated no differences between the 3 cases [p = .160, and p = .262 respectively].

A Kruskal-Wallis test did reveal a significant difference between the 3 cases in contradiction sentences (p = .008). This difference seems to stem from the King of France case in which participants were more likely to respond 'Neither' unlike the other two cases in which there were a greater proportion of 'False' responses. Figure 2 shows the proportion of responses for each sentence type after conflating the 3 cases.



Figure 2. Proportions of responses for each sentence type.

Figure 2 indicates that there was a higher proportion of 'True' responses made to the analytic truth sentences (approximately 20%) compared to the negligible proportion for the contingent and contradiction sentences, as well as the cases in Study 1. This was borne out statistically by comparing the proportions of 'True' responses made to each sentence type with ANOVA which confirmed a significant difference between the three sentence types [F (2, 253) = 11.50, p < .05]. Follow-up Bonferroni post-hoc analyses indicated that the differences in proportions of 'True' responses lay between the analytic truth sentences and the contingent sentences and the analytic truth sentences and the contradiction cases (p < .05). This is evidence of a conceptual reading because the conceptual reading predicted that participants would respond 'True' to only the analytical truth sentences. This evidence for a conceptual reading indicates that when simple sentences that contain an empty definite description apply a predicate that is analytically true of the concept described, then the sentence will be considered true in some instances. Although minimal, it is interesting that there was some endorsement of 'True' responses to the contradiction cases, however, this can most likely be put down to 'noise' in data.

Figure 2 also indicates a large proportion of 'False' responses to the contradiction sentences which is predicted by both Russellianism and a conceptual reading. However, an ANOVA comparing the proportions of 'False' responses did not find a significant difference [*F* (2, 253) = 2.88, *p* = .058].<sup>48</sup>

Interestingly, on the contingent sentences the results of Study 1 were not replicated. In Study 1 there was a consistent finding of approximately a 60-40 split of responses between the Frege-Strawson Theory and Russellianism. On all 3 cases in Study 1 there were substantially more 'Neither' responses than 'False' responses. In Study 2, however, there was a greater endorsement of 'False' responses to the contingent sentences. This inconsistency seems to be driven solely by the City at North Pole case. On the King of France case there is equal support for both Russellianism and the Frege-Strawson Theory, while on the Australian golden mountain case there was greater support for the Frege-Strawson Theory (comparable to Study 1). The reverse proportion on the City at the North Pole case is thus the anomaly, although it is unclear why. One possibility is simply that, unlike the other

<sup>&</sup>lt;sup>48</sup> Interestingly this was not far off significant at  $\alpha$  = 5%. It might be that there was not enough statistical power for any difference to reach significance.

cases, the City at the North Pole could be considered a fictional entity. If this was the case, participants may have looked for an interpretation that would render it false, i.e., in the manner of Russellianism.

Another interesting finding here is the proportion of 'Both' responses. In Study 1 the proportion of 'Both' responses was negligible, a finding that is replicated here for the contingent sentences. However, in the analytic truth and contradiction sentences here we see 'Both' making up more than 10% of responses. This is particularly evident for 'the king of France' a case in which 'Both' makes up 22.9% of responses on the analytic truth sentence and 15.6% on the contradiction sentence. However, again, this difference was not borne out statistically as an ANOVA comparing the proportions of 'Both' responses across the sentence types did not find a significant difference [F(2, 253) = 1.68, p = .188]. Nonetheless, it is worth noting that while none of the three readings considered here predict this result, 'Both' was included to be indicative of a pluralist view. Thus the proportions of 'Both' responses might be taken as weak evidence for a pluralist view. In Section 2.5.2 below I will discuss whether pluralism does give a better explanation of the present results. For now it is worth noting that we might expect a greater proportion of 'Both' responses if a pluralist theory was correct; especially on the analytic truth sentences (it is unclear that we should expect 'Both' responses on the other sentences if a pluralist theory is correct). Thus, I will argue that this is not adequate evidence for a pluralist view. Another way to interpret these responses is that some participants were simply torn between two different readings. For example, in the case of the analytic truth sentences, participants might have noted that the sentences were false (or neither true nor false) because the definite description did not refer, but they might have also seen that in some sense they were true (i.e., conceptually true). If given the context of the sentence they might have been more willing to commit to one reading. That is, they do not necessarily hold that two readings are available at all times, just that they are unsure which reading is expressed in this instance. It is harder to see why there are 'Both' responses on the contradiction sentences, albeit a lesser proportion. Here my best explanation would be that participants recognised that the sentences were both in a sense false and in sense neither true nor false (as opposed to both true and false).

The results of Study 2 indicate that there are three possible readings of simple sentences containing definite descriptions: Russellianism, the Frege-Strawson Theory, and also a conceptual reading. The conceptual reading has previously received no attention in the literature. Moreover, I have argued that simple sentences containing definite descriptions are usually *ambiguous* between these readings.

#### 2.3.3 Summary so far

Study 1 and 2 indicate a large amount of variance about what ordinary language users consider the meaning of a sentence containing an empty definite description. Study 1 indicates that Russellianism is false. This is because there was not a consensus among ordinary language users that sentences like 'the present King of France is bald' are false, and a consensus of this kind is what Russellianism predicts. Indeed, the results indicate support for both the Frege-Strawson Theory and Russellianism. Based on these results I contended that the Ambiguity Thesis provides the best explanation for this variance. Indeed, the Ambiguity Thesis holds that sentences containing definite descriptions can be disambiguated with either the content suggested by the Frege-Strawson Theory or the content suggested by Russellianism. This accounts for all the variation uncovered. Study 2 extended this finding by providing evidence in favour of a third possible reading, namely a conceptual reading. Study 2 suggested that in certain contexts, most notably in cases of (apparent) analytic truths, a sentence containing a definite description might be disambiguated with the content suggested by the conceptual reading. In the next part of this chapter I will present what is probably the most threatening challenge to the Ambiguity Thesis. This challenge claims that the data in Studies 1 and 2 is better explained in terms of some kind of pragmatic phenomenon rather than in terms of a genuine semantic ambiguity. I will argue that the results of two more experiments suggest that a pragmatic account does not better explain the present data.

## 2.4.1 This ambiguity is a semantic rather than pragmatic ambiguity

Studies 1 and 2 provide evidence that simple sentences containing definite descriptions are multiply ambiguous. However, opponents of ambiguity theses in philosophy give us some reason to question whether this apparent ambiguity is genuine. This relates to 'The Ambiguity Aversion Objection' raised in Section 1.2. In this section I want to argue that the pattern of responses uncovered in Studies 1 and 2, is not better accounted for in terms of some sort of pragmatic phenomenon, but is actually indicative of a genuine semantic ambiguity.

The main reason to think that a pragmatic explanation could account for the results comes from philosophers who argue against the postulation of unnecessary ambiguities (see Section 1.2). Most notably in this context is Kripke's (1977) challenge to Donnellan (1966). Kripke argues that Donnellan's distinction between the referential and attributive functions of definite descriptions is consistent with Russellianism, so we do not need to posit a semantic ambiguity. However, Kripke's argument can also be taken as a challenge to the Ambiguity Thesis I defend in this thesis. We can take this argument as the following:

- Ordinary language users' responses in Studies 1 and 2 are consistent with both Russellianism and the Frege-Strawson Theory.
- The Ambiguity Thesis can account for these responses by postulating a semantic ambiguity in sentences containing definite descriptions.
- Russellianism can account for all these responses by appealing to pragmatic considerations.
- All else being equal, an ambiguity thesis in philosophy should be rejected in favour of a univocality thesis.
- Therefore, all else being equal we should reject the Ambiguity Thesis in favour of Russellianism.

While I accept this argument I will argue that all else is not equal. Thus, we should not accept 5). This is because I will find in two new experiments evidence that cannot be

accounted for by Russellianism even when appealing to pragmatic considerations. In light of these experiments we will have good reason to consider the pattern of responses uncovered in Studies 1 and 2 to be best accounted for by the Ambiguity Thesis. Before presenting these experiments I will discuss how Russellianism can account for the variation uncovered so far in Studies 1 and 2 by appealing to pragmatic considerations, and thus explain why it might be thought that the first and third premises are true.

As discussed in Section 1.2, Kripke (1977) argues that the attributive sense (i.e., Russellianism) can fully account for the semantics of definite descriptions and therefore that we should not consider simple sentences containing definite descriptions to be ambiguous. To make his argument, Kripke invokes the Gricean distinction between what a speaker actually says (i.e., semantic reference) and what she intends to communicate (i.e., speaker's reference) by an utterance. For example, a speaker who is mistaken about the identity of the person they see might utter the sentence "Jones is raking leaves" when, in fact, observing Smith. In this case, although the speaker says something (false) about Jones, they may still successfully communicate something *about Smith*. Likewise, Kripke argues that referential uses of definite descriptions can be explained in purely pragmatic terms, and have no bearing on the semantics of definite descriptions. That is, a speaker *can* use a definite description as a referring expression, but such an instance is not semantically significant. Accordingly, the truth conditions for simple sentences containing definite descriptions in any context are always Russellian. This suggestion is further pressed by Neale (1990). Neale appeals to our intuitions regarding other quantifiers to argue that there are obvious cases in which we use a quantifier to communicate something, but would be reluctant to consider what's communicated as semantically relevant. Consider the following passage:

[T]he case for an ambiguity in the definite article weakens considerably once it is realized that sentences containing all sorts of quantifiers may be used to convey object-dependent propositions. Suppose it is common knowledge that Smith is the only person taking Jones' seminar. One evening, Jones throws a party and Smith is the only person who turns up. A despondent Jones, when asked the next morning whether his party was well attended, says,

Well, everyone taking my seminar turned up

fully intending to inform me that only Smith attended. The possibility of such a scenario would not lead us to complicate the semantics of 'every' with an ambiguity; i.e., it would not lead us to posit semantically distinct quantificational and referential interpretations of 'every-one taking my seminar' [pg. 87-88]

Thus, Neale argues that simply because a definite description can be used to communicate referentially, that is, object-dependently, this is no reason to think that such a use reflects anything about the meaning of the sentence containing it.<sup>49</sup>

To further support his argument that definite descriptions are not ambiguous Kripke makes the following arguments about the English language:

 The Hypothetical Language Argument: even in a hypothetical language in which it is stipulated that propositions containing definite descriptions are analysed according to Russellianism (i.e., Russell-English) we would still see instances of referential uses emerge:

If someone alleges that a certain linguistic phenomenon in English is a counterexample to a given analysis, consider a hypothetical language which (as much as possible) is like English except that the analysis is *stipulated* to be correct. Imagine such a hypothetical language introduced into a community and spoken by it. If *the phenomenon in question would still arise in a community that spoke such a hypothetical language (which may not he English), then the fact that it arises in English cannot disprove the hypothesis that the analysis is <i>correct* for English. [Kripke (1977) p. 265]

Thus, because there is no discernible difference between English and Russell-English, there is no reason to think that referential uses are semantically significant.

2. The Natural Language Argument: if definite descriptions are genuinely ambiguous then we should expect to see other natural languages use different words for each of the disambiguations, as we do with other ambiguous words in English. So, for example, while 'bank' is ambiguous in English, in French its synonyms 'banque' and 'bord' are not similarly ambiguous. But no languages contain separate words to mark the supposed

<sup>&</sup>lt;sup>49</sup> Reimer (1998) has responded to Neale's example that definite descriptions, unlike other quantifiers, are standardly used referentially and so they constitute a special case. This argument, that standard use is semantically correct, will be discussed in more depth shortly.

distinction between the referential and attributive senses of the definite article. So there is no such semantic ambiguity (cf. Ludlow & Neale, 2006).

Kripke's challenge seems entirely correct and uncontroversial; Donnellan's distinction gives us no reason to think that definite descriptions are semantically ambiguous (it is somewhat controversial whether Donnellan himself intended his distinction to be a semantic one). However, when applied more generally to the Ambiguity Thesis, in which an ambiguity *is* explicitly postulated, Kripke's challenge poses a substantial threat. In this relation to the present hypothesis, Kripke's challenge gives us reason to regard the Frege-Strawson Theory as only a theory about what is *pragmatically* conveyed.

Advocates of an ambiguity thesis have attempted to address Kripke's challenge, as well as provide independent reasons to regard the Frege-Strawson Theory as semantically significant. Firstly, to 'The Hypothetical Language Argument' it has been argued that because Kripke does not present any positive arguments against an ambiguity it is at least reasonable to claim that definite descriptions are semantically ambiguous (see Reimer, 1998). Indeed, while The Hypothetical Language Argument appeals to the principle of simplicity, it provides no obligatory reason to reject the semantic significance of referential descriptions. To this end, Reimer (1998) considers the conditions it would take to say that referentially used definite descriptions are semantic uses, and goes on to argue that standard use is what it would be required. She gives the analogy of dead metaphors in standard use. For example, 'incensed' could be stipulated have the single meaning 'to make fragrant with incense', although it could also be used metaphorically to mean 'to make very angry'. However, through standard use the metaphorical meaning may have become interpreted literally by users and become a second actual meaning. When concerned with this phenomenon, Reimer argues that in Russell-English referential uses of definite descriptions would not be used standardly as they are in English, and so there is a discernible difference between the two languages. The difference is that in Russell-English definite descriptions would not be used referentially, that is, to refer object-dependently, as frequently as they are in English.

With regard to 'The Natural Language Argument' it has been argued that it is not remarkable, as Kripke claims, if no language disambiguates an ambiguous term in English by

containing two (or more) distinct words to express the distinct meanings associated with the English term. Such an observation is common where the ambiguity in question is a case of polysemy (Amaral, 2008). Note from Section 1.2 that a polyseme is an ambiguous word in which the different meanings are conceptually connected. In this case, then, the suggestion is that the possible ambiguity of definite descriptions is unlike other ambiguous words like 'bank' (i.e., homonyms), and should be compared to other cases of polysemy instead.

Besides specifically addressing these arguments, advocates of an ambiguity in sentences containing definite descriptions have made the following observations which give us independent reason to consider the Frege-Strawson Theory semantically significant. For one, both referentially used definite descriptions and complex demonstratives are used similarly in a wide range of situations; this suggests that they have similar meanings (Devitt, 2004; see Amaral, 2008). The Frege-Strawson Theory also accounts for incomplete definite descriptions (Devitt, 2004; Wettstein, 1981).<sup>50</sup> In particular, definite descriptions are frequently used in the manner of the Frege-Strawson Theory, that is, as referential expressions (Devitt, 2004; Reimer, 1989). This last point is particularly germane. As noted above, Reimer (1989) argues that the standard use of a phrase introduces a literal meaning. Take the example from above in which 'incensed' comes to literally mean 'to make very angry' through common usage by ordinary language speaker's. If this is right then we should consider whatever the most natural interpretation of a sentence containing a definite description is (i.e., the default meaning), according to ordinary language users, to be a semantic interpretation. If the Frege-Strawson Theory offers a standard or common use of definite descriptions, as Reimer and Devitt claim it does, then there does not seem to be a principled reason to deny this convention a meaning in its own right.

Sennet (2002) has presented an alternative type of objection to positing an ambiguity in definite descriptions by appealing to a standard test for semantic ambiguity. As discussed in Section 1.2, there are a number of linguistic tests designed to uncover semantic ambiguities in various terms including the conjunction reduction test. This test identifies ambiguities by

<sup>&</sup>lt;sup>50</sup> As I noted briefly earlier (Section 2.2.1), it is controversial whether incomplete descriptions should count as evidence against Russellianism. However, advocates of an ambiguity theory have maintained that cases of incompleteness at least favour the Frege-Strawson Theory, as it easily accounts for these cases, over Russellianism (e.g., Devitt, 2004).

attempting to force a single token of a possibly ambiguous word to contribute multiple meanings to a sentence:

(8) The colours and feathers are light.<sup>51</sup>

The feeling of zeugma produced by (8) stems from the cross-interpretation of 'light' indicating that it contributes multiple meanings. Some other examples of zeugmatic sentences are:

- (9) He called her a fool and a cab.<sup>52</sup>
- (10) Delia ran the family business and a marathon.<sup>53</sup>

Sennet has applied the same test to the definite article to argue that the lack of zeugma shows that definite descriptions are not ambiguous. He uses the following example:

Imagine we are conspiring to incite revolution by killing certain members of the royal family. Furthermore, imagine that a woman was recently smuggled into the castle and wedded to the King, and no one knows who she is. Pointing to a fellow who looks like a prince, I utter:

8. The Prince will die at dawn.

When uttering this, I point at the man I want to kill once the revolution begins. In fact, his being a prince is not particularly relevant, as he would still be a target to the conspirators since we hate that particular fellow. We don't really care all that much if the description fits the salient individual. This is a context in which Donnellan should happily grant that my description is used referentially. Next, you utter:

9. The Queen will die at dawn.

As noted above, the identity of the Queen is unknown – my use of "The Queen" is paradigmatically attributive. Donnellan suggests a test for attributive uses: If you can you sensibly add 'whoever she/he/it is' after the description, it is attributive. Presumably we can happily do this in this case:

10. The Queen, whoever she is, will die at dawn.

<sup>&</sup>lt;sup>51</sup> From Ludlow (2013) 'Descriptions' in the *Stanford Encyclopaedia of Philosophy*.

<sup>&</sup>lt;sup>52</sup> From Laserhohn (1993).

<sup>&</sup>lt;sup>53</sup> Sennet's (2002) own example.

(9) is thus a paradigm of the attributive use if there are any such uses at all. The conjunction reduction on (8) and (9) provides us with:

11. The Prince and Queen will die at dawn.

(11) strikes me as perfectly felicitous in this context, but it really shouldn't if Donnellan is right. If the sense of 'the prince' in (8) is retained in (11), and we can't imagine that the person even points to the purported referent while saying 'prince', then (11) seems to naturally blend a referential and attributive use through the use of a single determiner. It is *prima facie* odd that the word 'the' can distribute over the two distinct uses. [p 84-85]

That 'The Prince and Queen will die at dawn', given the backstory, is not overtly zeugmatic leads Sennet to argue that definite descriptions are not ambiguous between Russellianism and (something like) the Frege-Strawson Theory. However, it is important to recognise that Sennet is actually addressing Donnellan's distinction (discussed in Section 2.3.2) not the Ambiguity Thesis per se. The difference here is that, unlike the referential function discussed by Donnellan, the Frege-Strawson theory would not consider this use of 'the prince' to actually refer to the man in question. This is because when "The Prince will die at dawn" is uttered in Sennet's example, the speaker's intended referent is not necessarily a Prince and therefore does not necessarily fulfil the descriptive content of 'the F'. Thus Sennet's example is not a true counterexample to the Ambiguity Thesis. Moreover, there are other reasons to be skeptical of Sennet's conclusion; notably, Sennet attempts to address a number of these concerns.<sup>54</sup> However, probably the most pressing problem for Sennet, and an objection that he does not address, is that the conjunction reduction test is only a positive test for ambiguity. Zeugma in the reduced sentences is evidence of an ambiguity, but a lack of zeugma does not count as evidence to the contrary. Interestingly, as Sennet himself notes in 'Ambiguity' (2015), zeugma can be subtle. Especially in cases where the meanings are similar it might be that the faint zeugma is simply not striking enough to

<sup>&</sup>lt;sup>54</sup> Two such objections that Sennet addresses are the following: Firstly, it could be that in the reduced sentence, 'The Prince and Queen will die at dawn', 'The Prince' is given a Russellian reading instead of the referential one it had in the unreduced sentence. If this is the case then Sennet's example differs from other cases of zeugma [e.g., (8), (9), & (10)]. However, Sennet argues that in this case there would be a change of referent between each use of 'The Prince' (i.e., between the unreduced and the reduced sentences). Another objection is that there might be an unvoiced but implicitly understood second definite description in Sennet's test. According to this objection, a better interpretation of the reduced sentence is 'The Prince and the Queen will die at dawn' which is perfectly in accordance with an ambiguity in the definite article. Sennet agrees that this might be the case, but argues that if it is then there is a constraint on the unvoiced determiner in that it would need to operate in the same way as the voiced determiner to act in accordance with other examples of unvoiced determiners in English.

register. This is exactly what seems to be the case for Sennet (2002). At best it seems that he has merely failed to illustrate that definite descriptions are ambiguous.

So far the debate about whether simple sentences containing definite descriptions are semantically ambiguous has not appealed to experimental work. Indeed, while there are Kripke's two well-known arguments (the hypothetical language argument and the natural language argument) which rely on observations about language, and Sennet's appeal from the armchair to linguistic tests, there are no experiments that I know of that test the intuitions of ordinary language users on the matter. This is possibly because it is notoriously difficult to test for ambiguities in natural language (as discussed in Section 1.2). Thus, my aim here is to present two experiments designed to test for the presence of a genuine semantic ambiguity in definite descriptions. Specifically, I will test whether Russellianism can account for some further intuitions to see whether (3) holds. Indeed, I will argue that Russellianism cannot account for these results and therefore that the Ambiguity Thesis is the best explanation.

However, it needs to be noted here that these experiments are conducted in the context of Kripke's challenge. That is, they test whether the Frege-Strawson responses are indicative of the semantics of definite descriptions. Thus, they support the Ambiguity Thesis only insofar as Russell's Theory of Descriptions is already presupposed to give *a* semantics for definite descriptions in English. Presumably, philosophers who support a pure version of the Frege-Strawson Theory, like Elbourne (2013), will not consider this evidence as support for the Ambiguity Thesis but as evidence in favour of the pure Frege-Strawson Theory. Moreover, because this challenge says nothing about the conceptual reading, these experiments will not provide evidence that the conceptual reading is a genuine semantic reading. This is a consideration I will take up later in the chapter (see Section 2.5.2).

## 2.4.2 Study 3: The Implicature Survey

Study 3 was designed to test whether native English-language speakers consider part of the meaning of a definite description to be object-dependent. That is, whether the intended object of a definite description is a *semantic* constituent. Recall that the Frege-Strawson

Theory regards definite descriptions as referring expressions. Thus, on this view *the F is G* entails the proposition *that F is G*. A pragmatic explanation, on the other hand, holds that such a definite description merely implies or implicates the proposition *that F is G*. That is, although the speaker might have some entity in mind, it is merely implied that 'the *F*' refers object-dependably.

*Method*: A new sample of 167 New Zealand participants was collected (60.6% female). Like in Study 1, an additional sample of seventy-seven US participants was collected online (68.8% female; mean age of 40.1).

Before receiving the test question, participants received four paradigmatic case questions. These questions tested whether participants could correctly identify cases of entailment and cases of implicature in paradigm scenarios. First, participants received the following explanation:

The following questions are designed to test your ability to distinguish between entailment and **mere** implication (or implicature). Let us say that that something is <u>entailed</u> by an utterance of a sentence *S* when it is part of the meaning of *S* itself (or has exactly the same meaning as *S*) irrespective of who utters it. Let us say that that something is <u>merely implied</u> by an utterance of a sentence *S* when it can be inferred from the context but is **not** part of the meaning of the sentence itself.

Here is an example to illustrate the idea. Suppose you ask me if I am going to the party, and I answer as follows:

My Answer: 'I don't like any social events.'

In this case, the My Answer <u>entails</u> that *I don't like parties* because every party is (of necessity) a social event. But my answer <u>only implies</u> that I am not going to the party, because I could go to the party even though I don't like them (perhaps because my friends insist that I go).

Participants then received the following questions designed to test whether they could correctly utilise the distinction.

## **Question 1:**

Suppose I say the following to you:

#### (U1) I went to the top of the mountain

Circle the answer below you are most confident is correct:

- (a) U1 <u>entails</u> the following: *I went to the top of the mountain*.
- (b) U1 merely implies (without entailing) the following: *I went to the top of the mountain*.
- (c) U1 <u>does **not** entail or imply</u> the following: *I went to the top of the mountain*.

## **Question 2:**

Suppose I say the following to you:

#### (U2) John is still a bachelor

Circle the answer below you are most confident is correct:

- (a) U2 <u>entails</u> the following: *John is an unmarried man*.
- (b) U2 merely implies (without entailing) the following: John is an unmarried man.
- (c) U2 does **not** entail or imply the following: John is an unmarried man.

## **Question 3:**

Suppose I say the following to you:

#### (U3) I have not given up smoking yet

Circle the answer below you are most confident is correct:

- (a) U3 <u>entails</u> the following: *I am a smoker*.
- (b) U3 merely implies (without entailing) the following: *I am a smoker*.
- (c) U3 does **not** entail or imply the following: *I am a smoker*.

# Question 4:

Consider the following sentence:

#### (U4) Steve's wife is nice

Circle the answer below you are most confident is correct:

- (a) U4 <u>entails</u> the following: *Steve is nice*.
- (b) U4 merely implies (without entailing) the following: Steve is nice.
- (c) U4 does **not** entail or imply the following: *Steve is nice*.

For these questions the correct answers are, evidently, (a) for Questions 1 and 2, (b) for Question 3, and (c) for Question 4. Participants in the NZ sample received the paradigmatic questions in one of three different orders (Order A, Order B, Order C) to control for any potential order effects affecting participants' intuitions on the test question. Participants from the US sample all received the same order (Order A).<sup>55</sup>

All participants received the test question last.

## Test question:

Consider the following sentence:

#### (U5) The woman in the corner is happy

Circle the answer below you are most confident is correct:

- (a) U5 <u>entails</u> the following: *that woman in the corner is happy*.
- (b) U5 merely implies (without entailing) the following: *that woman in the corner is happy*.
- (c) U5 does **not** entail or imply the following: *that woman in the corner is happy*.

<sup>&</sup>lt;sup>55</sup> This was considered justified as only a very minimal effect of order was found for the NZ sample.

This question asks whether a sentence containing a definite description entails that *F* is *G*, as the Frege-Strawson Theory holds, or only implies it, as Kripke suggests is consistent with Russellianism. Therefore, on this question (a) responses were considered indicative of the Frege-Strawson Theory and (b) responses were considered indicative of a Kripkean type defence of Russellianism.

*Results and Discussion:* The proportions of participants who correctly applied the entailment/implicature distinction on the paradigmatic questions were: 77.9% for Question 1, 51.9% for Question 2, 53.5% on Question 3, and 76.8% on Question 4 (see Table 5 for the proportions for each sample). Mann-Whitney U tests indicate significant differences between the responses of the two samples on Question 1 and Question 3 (U = 4848.5, p < .001, U = 5274, p = .011).

Table 5. shows the proportions (%) of participants in each sample who answered correctly on each paradigmatic case.

61.0
57.1
41.6
75.3

The proportions of responses for participants on the test question were: 49.4%, 36.5%, and 14.1% for responses (a), (b), and (c) respectively. A Kolmogorov-Smirnov test indicated a significant difference between the proportions of responses [4.83, p < .001]. The proportion of responses to the test question was then calculated for only the participants who correctly answered the paradigmatic questions. Only 44 participants (18.0%) answered all four questions correctly; the proportion of responses was: 68.2%, 29.5%, and 2.3% for responses (a), (b), and (c) respectively. Again, a Kolmogorov-Smirnov test indicated a significant difference between the proportions of responses [2.81, p < .001]. The proportions for both calculations of the test question are shown in Figure 3.



*Figure 3.* Proportions of responses for all the participants and for only the participants who correctly answered the paradigmatic questions.

Only a minimal effect of order was found. On Question 4, a Kruskal-Wallis indicated a significant order effect [8.40, p = .015]. Follow up Mann-Whitney U tests indicated that the difference is between Order A and Order C, U = 1123.0, p = .006. Importantly, order had no effect on the test question [.342, p = .843].

These results give evidence for the Ambiguity Thesis against an alternative pragmatic explanation. This is observed from the finding that most participants read the description 'the woman in the corner' as though it was a complex demonstrative like 'that woman'. That is, participants tended to think that the definite description referred to a specific woman that was intended by the speaker, and that that woman contributed directly to the meaning of the sentence. Indeed, this finding is borne out by a binary chi-squared test which indicates that the proportion of (a) responses was significantly greater than chance (29.24, p< .001). Moreover, the proportion of participants that had this intuition, that is, the intuition (a), increased from 49.4% to 68.2% when the sample was limited to only the participants who demonstrated an aptitude for applying the entailment-implicature distinction on the paradigmatic questions. A Mann-Whitney U test indicated that this increase in (a) responses is significant (4305.00, p = .022). This finding suggests that the Frege-Strawson Theory does provide a semantic account of definite descriptions. A substantial proportion of participants did, however, consider the test sentence to only imply 'that woman'. This finding indicates that some participants had Kripke's intuition about definite descriptions — that while a speaker might have a specific woman in mind when they utter 'the woman in the corner is happy', that woman does not constitute part of the meaning of a sentence but is only conveyed pragmatically. Taken together, these two findings suggest that participants can read 'the F is G' as entailing or implying that woman, that is, in accordance with either the Frege-Strawson Theory or Russellianism. If this is the case, Study 3 gives further evidence in favour of the Ambiguity Thesis.

One potential worry is that there is a similar pattern of responses regardless of whether or not the sample is restricted to those who correctly answered the paradigmatic cases. If participants who demonstrate an aptitude for applying the entailment-implicature distinction have similar philosophical intuitions to participants who do not then there is reason to think that this aptitude does not translate to better philosophical intuitions. It is possible that we simply do not have the ability to apply the distinction in cases where the

answer cannot be verified. It is interesting to note, however, that the greater the number of paradigmatic questions that a participant answered correctly, the greater the likelihood that they answered (a) on the test question, and the less likely it was that they answered (c) — a response inconsistent with both the Frege-Strawson Theory and Russellianism.

Another potential worry is that participants did not read 'that woman' as meaning the intended woman but instead responded (a) for a different reason; perhaps because they read 'that woman in the corner, *whoever that woman might be*, is happy'. However, this reading seems fairly unusual; we would not generally use the phrase 'that woman' and then specify that the identity of the woman is not known.

### 2.4.3 Study 4: The Default Meaning Survey

Study 4 was designed to test what ordinary speakers consider to be the most natural interpretation of a definite description, that is, the *default meaning* of a definite description. The default meaning of a term or phrase is the meaning that is most commonly used, or the predictable meaning (see Jaszczolt, 2014). We get the semantic meaning of these terms and phrases from what most people in a community would say in instances in which no context is given. It is by introducing context that we get instances of speaker's meaning. If the Frege-Strawson Theory gives one (semantic) disambiguation of definite descriptions then, when presented out of context, we should expect participants to interpret 'the *F*' as 'that *F*' a substantial proportion of the time. This is because when no contextual information is given, the default meaning should be the literal meaning, not the speaker meaning. Likewise, and for similar reasons, if Russellianism offers a (semantic) disambiguation of definite descriptions as 'there is a unique *F*' when presented out of context.

It might be thought that it is impossible to strip statements of their context in this way, and that participants will always impose a hypothetical speaker into the vignette. If that is right, testing for default meaning in this way may still only test for *speaker meaning*. That is, it only tests what participants think a hypothetical speaker is trying to convey rather than what the sentence actually says. However, Reimer (1998) and Devitt (2004) give us a reason

to think that even if this worry were well founded, default meaning should still be an indicator of semantic meaning. They argue that any projected hypothetical speaker will use the expression in a way it is commonly used. Moreover, if an expression is frequently used in some manner, then that use introduces a semantic convention. That is, if speakers commonly use an expression with some intended meaning, then that intended meaning becomes an actual meaning.

As noted in 2.4.1, Neale (1990) argues against considering the referential use of definite description to be indicative of a semantic meaning. He notes that we use other quantifiers, such as 'every', as referring expressions but do not consider them to be a second semantic interpretation. However, Reimer addresses this argument as she notes that while Neale is right about these other quantifiers; the reason that they are not considered to have a semantic meaning in line with the Frege-Strawson Theory is because they are not used commonly as referential expressions. Definite descriptions, on the other hand, are. So we should consider definite descriptions to have this second meaning.

In Study 4, the default meaning of definite descriptions will be tested by investigating what participants take to be the meaning of a sentence containing a definite description when no contextual information is available. Furthermore, sentences containing indefinite descriptions will also be included to test whether there is a disanalogy between definite descriptions and other quantifiers as Reimer suggests. If it turns out that participants are inclined to say that definite descriptions are best interpreted as 'that *F*', while indefinite descriptions are best interpreted as 'there is an *F*', then there will be good evidence for Reimer's claim. That is, there will be evidence that definite descriptions constitute a special case in that they are used more frequently as referring expressions than other quantifiers like 'an', and so (*pace* Neale) it is not inconsistent to assign only definite descriptions a second alternative referential meaning.

*Method:* A new NZ sample of ninety-seven participants was collected (67% female) as well as a new online US sample of 347 participants (79.5% female; mean age of 35.20).

There were two versions of the final survey, the definite article version and the indefinite article version. The two surveys differed only in (S1); see the following:<sup>56</sup>

(1) Consider the following sentence:

## (S1) [The / A] woman in the corner is happy

Circle the answer below you are most confident is correct:

- (a) Someone who utters (S1) is most likely to mean *that* woman in the corner is happy.
- (b) Someone who utters (S1) is most likely to mean <u>there is exactly one</u> woman in the corner and she is happy.
- (c) Someone who utters (S1) is most likely to mean <u>there is at least one</u> woman in the corner and she is happy.
- (d) Someone who utters (S1) is equally likely to mean two (of the three) italicised (red) statements above.
- (e) Someone who utters (S1) is equally likely to mean <u>three</u> (of the three) italicised (red) statements above.
- (f) None of the above.

Each participant only received one of the two versions. For participants in the NZ sample it was noted that a possible flaw in the method was that response (d) did not force participants to specify which two statements are meant. This was amended for participants in the US sample. Participants from this sample who gave response (d) were then required to specify which two answers [of (a), (b), and (c)] they had in mind in a follow-up question.

*Results and Discussion:* The results for each sample are set out in Table 6. Overall, of participants who received the definite article version: 43.7% answered (a) [i.e., '<u>that</u> woman'], 14.3% answered (b) [i.e., '<u>there is exactly one</u> woman'], 7.4% answered (c) [i.e., '<u>there is at least one</u> woman'], 20.3% answered (d) [i.e., 'equally likely to mean <u>two</u> (of the three) italicised (red) statements'], 8.7% answered (e) [i.e., 'equally likely to mean <u>three</u> (of the three) italicised (red) statements'], and 5.6 answered (f) [i.e., 'none of the above'] (see Figure 4). A Chi-square indicated these responses significantly differed from chance [25.34, p < .001]. Of the 29 participants from the US sample who answered (d), 55.2% answered (a)

<sup>&</sup>lt;sup>56</sup> Note that italicised statements were also presented in red.

and (b) on the follow-up question, and 44.8% answered (a) and (c); again a Chi-square indicated these responses significantly differed from chance [14.97, p < .001]. A Mann-Whiney U test indicated no difference between the answers of the two samples [U = 4031.5, p = .163].

Of participants who received the indefinite article version: 22.5% answered (a) [i.e., '<u>that</u> woman'], 16.9% answered (b) [i.e., '<u>there is exactly one</u> woman'], 25.8% answered (c) [i.e., '<u>there is at least one</u> woman'], 12.2% answered (d) [i.e., 'equally likely to mean <u>two</u> (of the three) italicised (red) statements'], 10.3% answered (e) [i.e., 'equally likely to mean <u>three</u> (of the three) italicised (red) statements'], and 12.2 answered (f) [i.e., 'none of the above'] (see Figure 4). A Chi-square indicated these responses significantly differed from chance [141.91, p < .001]. Of the 29 participants from the US sample who answered (d), 21.4% answered (a) and (b) on the follow-up question, 64.3% answered (a) and (c), and 14.3% answered (b) and (c); again a Chi-square indicated these responses significantly differed from chance [6.14, p = .046]. Again, a Mann-Whiney U test indicated no difference between the answers of the two samples [U = 3316.5, p = .148]. A Mann-Whitney U test indicated a significant difference between the two survey versions [U = 19761.5, p < .001].
	definite	indefinite
NZ Sample		
(a)	31.4	6.5
(b)	19.6	17.4
(c)	3.9	37.0
(d)	35.3	26.1
(e)	9.8	8.7
(f)		4.3
US Sample		
(a)	47.2	26.9
(b)	12.8	16.8
(c)	8.3	22.8
(d)	16.1	8.4
(e)	8.3	10.8
(f)	7.2	14.4

Table 6. shows the proportions of responses for each sample on both the definite article version and the indefinite article version.



*Figure 4.* Proportions of responses for each of participants who received the definite article version and participants who received the indefinite article version.

Like Study 3, these results provide evidence for the Ambiguity Thesis against a pragmatic explanation.<sup>57</sup> Most participants who received the definite article version thought that 'that woman' was a best interpretation of the test sentence (S1), either by responding (a) or responding in favour of multiple interpretations [of which (a) was one alternative]. This suggests that ordinary language users tend to most naturally interpret the meaning of 'the F is G' as partly constituted by the referent and therefore indicates that the Frege-Strawson Theory offers the most natural analysis of definite descriptions. This gives us reason to reject an explanation of the data that explains away any responses consistent with the Frege-Strawson Theory as merely pragmatic because *most* responses were consistent with it. A pragmatic explanation should not need to explain away most of the data; it should only be used to account for small deviations. Indeed, the high endorsement of the Frege-Strawson Theory is good evidence that it provides a genuinely semantic interpretation. We can see that a small proportion of participants responded consistently with Russellianism. It is this finding that gives evidence for the Ambiguity Thesis. Moreover, evidence for the Ambiguity Thesis is further borne out in Study 4 by a substantial proportion of participants who considered there to be at least two equally likely interpretations.

Responses to the indefinite article version were substantially different. Like the definite article version, there is strong evidence in favour of an ambiguity in the meaning of a sentence containing an indefinite description. However, compared to the definite article version, a greater proportion of participants had intuitions consistent with Russellianism, and a smaller proportion had intuitions consistent with the Frege-Strawson Theory. It is this finding that is evidence for Reimer's claim that definite descriptions constitute a special case as they are used more frequently than other quantifiers as referring expressions. Indeed, this finding indicates that indefinite descriptions, at least, are much less frequently interpreted as referring expressions compared to definite descriptions. For this reason it seems that we should be able to consider definite descriptions to have adopted a semantic interpretation in line with the Frege-Strawson Theory without being required to think this extends to other quantifiers as well.

<sup>&</sup>lt;sup>57</sup> It is again worth pointing out that this is evidence for the Ambiguity Thesis only to the extent that we already accept Russellianism to be part of the semantics.

#### 2.5.1 General discussion

The Ambiguity Thesis holds that simple sentences containing definite descriptions can be disambiguated in multiple ways; they can have the semantic content suggested by Russellianism, the Frege-Strawson Theory, or a conceptual reading. The four experiments presented here provide a compelling case to accept the Ambiguity Thesis about definite descriptions. Studies 1 and 2 give initial evidence to this effect; they indicate that Russellianism is false because ordinary language users hold intuitions consistent with each of the Frege-Strawson Theory and Russellianism (Study 1), as well as a conceptual reading (Study 2). Studies 3 and 4 provide evidence against a possible pragmatic interpretation of the data. That is, they provide evidence that the variation uncovered in our intuitions is better accounted for by positing multiple semantic contents, as the Ambiguity Thesis holds, as opposed to a single content that is enriched with pragmatic considerations. Moreover, these results are also evidence against a pluralist explanation; that is, that simple sentences containing definite descriptions express multiple propositions (this will be discussed further in Section 2.5.2). In particular, these experiments are significant as they are some of the first to be conducted in this area of language. Mostly philosophers have only appealed to their own intuitions to inform on this debate; very few studies have actually employed the survey methodology of experimental philosophy. One notable exception is Abrusán and Szendrői (2012). However, the present set of experiments builds markedly on this study; most importantly, addressing concerns of the possibility of a pragmatic interpretation.

The Ambiguity Thesis, I contend, has the following advantages:

- 1. It can explain all the variation in participants' intuitions uncovered in Studies 1-4,
- 2. It does not require us to abandon our commitment to using intuitions as a useful source of evidence in this domain, and,
- 3. It provides a natural solution to the problems that face unitary theories

Each of these points will be discussed in turn. With regard to the first point, according to the Ambiguity Thesis, simple sentences containing non-referring definite descriptions can be disambiguated in multiple ways. Accordingly, the Ambiguity Thesis not only explains but

actually predicts variation in the intuitions people have about the truth value of such sentences. Notably, it predicts this variation should be observable in a number of ways. For one, we should expect to see variation within single cases from hearers and readers disambiguating the same sentences in different ways. This is seen in the pattern of endorsement of each of Russellianism, the Frege-Strawson Theory, and a conceptual reading in single cases. For example, take the single case of 'the Australian golden mountain is a mountain' from Study 2; in this case what we saw was that 23.1% of participants endorsed the Frege-Strawson Theory, 23.1% endorsed Russellianism, and 46.2 endorsed the conceptual reading. Secondly, we should also expect intuitions to vary across cases due to the different features of the subject matter in each case. This is observed in the present results by the differing proportions of intuitions to different cases. For example, in Study 1 there was a higher proportion of Russellian responses on the King of France case than on the President of England case. Lastly, we should expect individual participants to have different intuitions across similar cases. In Study 1, a substantial proportion of participants (approximately 74% and 55% of New Zealand and US participants respectively) did not respond consistently across cases. This indicates that individual participants were comfortable assigning the truth values predicted by different theories for different cases. That is, they were comfortable holding different intuitions for different cases.

One might wonder if the Ambiguity Thesis can account for why we see greater support for the Frege-Strawson Theory. If simple sentences containing definite descriptions were ambiguous we might expect a roughly equal split between the Frege- Strawson Theory and Russellianism (and perhaps the conceptual reading). While there is support for all three readings the responses favoured the Frege-Strawson Theory. The Ambiguity Thesis deals with this concern as it holds that the linguistic context and construction can play a pivotal role in determining the way an utterance is interpreted. In many cases, context might restrict what readings are available. It may be that only one interpretation is appropriate in that context, or that some interpretation is more likely due to context. The reason the Frege-Strawson Theory was found to have consistently greater support is not a challenge to the Ambiguity Thesis but simply shows that in most cases the context makes it more likely that a definite description is used as a referring expression.

Regarding the second point, by endorsing the Ambiguity Thesis we are able to respond to the claim made by some philosophers that the variation uncovered in experimental work suggests that we should not make philosophical claims on the basis of intuitions (cf., Machery et al., 2004). I discussed this point in Section 1.5. The rationale here is that if people hold disparate intuitions then, unless we have some principled reason to favour the intuitions of one group over another, it seems we cannot appeal to intuition to inform on philosophical theories.<sup>58</sup> The Ambiguity Thesis does not face this consequence; variation in intuitions is exactly what we should expect to see if the Ambiguity Thesis was true. Thus, the Ambiguity Thesis gives us another interpretation of varied intuitions, one, I believe, that has a better outcome for philosophy.

Regarding the last point, the Ambiguity Thesis gets around the problems that face each of the univocal theories of definite descriptions. Both Russellianism and the Frege-Strawson Theory face a number of difficulties when taken in isolation. In particular, each of the theories makes counterintuitive predictions about some cases. That is, for each theory there are counterexamples to the truth-conditions they assign. Proponents of each theory are therefore required to explain away these counterintuitive cases. I will give some examples of these kinds of cases to emphases the advantage of sidestepping them by appealing to the Ambiguity Thesis. The Ambiguity Thesis avoids such problems because for any seemingly problematic interpretation of a definite description, there is always another, non-problematic interpretation also available.

When Russell (1905) proposed his Theory of Descriptions one motivation was that, unlike some previous theories like Frege's, it could account for the problem of negative existentials; recall the following case:

## (4) The Present King of France does not exist.

Frege (and later Strawson) seem to be committed to say that negative existentials, such as (4) have no truth-value (see Neale, 1990; Russell, 1957). However, they strike us as plainly

<sup>&</sup>lt;sup>58</sup> For example, Machery et al. (2004) think this is the case for theories of reference about proper names. Their experimental work uncovered substantial variation in the intuitions held by participants suggesting that the causal historical theory of reference is not as 'obvious' or 'intuitive' as has been claimed.

true.<sup>59</sup> As I noted in Section 2.3.1, Russellianism does not face this counterintuitive outcome because according to Russell complex sentences containing definite descriptions entail a scopal ambiguity. Thus, negative existentials like (5) can be interpreted in two ways:

(4a) The Present King of France is such that he does not exist.

(4b) It is not the case that the Present King of France does exist.

If the operator is given wide-scope, as in (4b), then it renders the proposition true.

On the other hand, Russellianism was challenged for giving counterintuitive truth-conditions in other cases; in particular, in cases of simple sentences containing empty definite descriptions. Indeed, as noted earlier, Strawson (1950) argued that Russellian truthconditions were wrong in cases like:

(1) The present King of France is bald

Strawson argues that it is counterintuitive to consider (1) to be false, as Russellianism holds.<sup>60</sup> Because of these cases, the Frege-Strawson Theory is typically considered the more intuitive theory.

However, there are also counterexamples to the Frege-Strawson Theory. Indeed, Russell (1957), in a direct response to Strawson, maintained that cases like (1) were in fact false. To illustrate, he presented the following example:

<sup>&</sup>lt;sup>59</sup> But see Elbourne (2013) for an analysis that he believes overcomes this problem for the Frege-Strawson Theory (pg. 68).

<sup>&</sup>lt;sup>60</sup> Strawson (1950) also argues that Russellianism cannot account for our intuitions about non-unique definite descriptions (i.e., incomplete descriptions) like 'the table is covered with books' [p. 271] because Russell's theory always renders them false. According to Strawson, at least, his theory easily deals with incomplete descriptions as it only *presupposes* the existence of a unique table. To flesh this out further (especially because Frege it seems would not have agreed with Strawson in this respect), in setting out the Frege-Strawson Theory in this thesis I argued that context determines which entity is referred to. So in cases of incomplete descriptions the definite description refers to a contextually salient individual, or an individual who uniquely satisfies the predicate in a contextually restricted domain (see Section 2.2.1). However, as there is an ongoing debate as to whether incomplete descriptions are a genuine problem for Russellianism I will not discuss it further here (but see Ludlow, 2013).

Suppose, for example, that in some country there was a law that no person could hold public office if he considered it false that the Ruler of the Universe is wise. I think an avowed atheist who took advantage of Mr. Strawson's doctrine to say that he did not hold this proposition false would be regarded as a somewhat shifty character [p. 389 (1957)]

Since Russell's reply, a number of philosophers have presented other kinds of statements containing a definite description that don't elicit the Frege-Strawson intuition. Strawson (1971) himself came to see the force in these examples and moved away from his original position, although, as we will see, he faces residual concerns. Indeed, Strawson himself noted such cases, including the following:

(11)The exhibition was visited yesterday by the King of France.

Neale (1990), among others, has also offered similar examples, such as:

(12)My Mother is dating the present King of France.

(13)This morning my father had breakfast with the king of France.

In these cases we are much more inclined to say that the statements are false, just like Russellianism holds.

A number of philosophers have offered accounts to explain our divergent intuitions between cases like (1) and cases like (11)-(13).

For one, Strawson (1971) gave a modified version of his truth value gap account (also see Gundel, 1977; Horn, 1989; and Reinhart, 1981). According to this account, cases like (11)-(13), which strike us as unequivocally false, do not introduce a presupposition that there exists something that satisfies the definite description. The reason he suggests is that in these cases the non-referring definite description is not the topic of the statement; that is, when the sentence is *about* the entity described. To this end he makes the following two claims:

 If a non-referring definite description is the topic of a sentence, then the sentence is neither true nor false. 2. If a non-referring definite description is not the topic of a sentence, then the sentence is not neither true nor false.

Thus, according to Strawson's modified account, the Frege-Strawson Theory only accounts for sentences in which a definite description is in the topic position. Otherwise Russellianism gets it right. To show this he contrasts the following statement:

- (11)The exhibition was visited yesterday by the King of France.
- (14) The King of France visited the exhibition.

Strawson argues that (14) is neither true nor false while (11), as we saw above, is false. Atlas (2004) defends a slightly weaker version of Strawson's modified account in which *only* definite descriptions in the topic position of a sentence carry an existence presupposition, but not all definite descriptions in the topic position do. Thus, he maintains the second of Strawson's claims (above) but not the first. Atlas's version explains some of the further counterexamples to Strawson's modified theory, such as our intuitions that both of the following sentences are false despite differing in the topic of the sentence:

- (15) Among the bald people in the world is King of France.
- (16)The King of France is one of the bald people in the world.<sup>61</sup>

Although he leaves it open as to the conditions under which definite descriptions that are the topic of a sentence introduce an existence presupposition.

Schoubye (2010) gives an account that explains our divergent intuitions in these cases by emphasising the role of contextual features. He argues that in some cases we have different intuitions depending on the surrounding context; such as in the following case:

(17) The King of France always enjoys a croissant in the morning.

He argues that it is the contextual features around the utterance of (17) that determine whether it is considered false or truth valueless. Specifically, he argues that a sentence containing a non-referring definite description will be judged as false if it is considered to be

<sup>&</sup>lt;sup>61</sup> Example taken from 'Presupposition and Truth-value gap' in Reimer and Bezuidenhout (2004).

a consonant response to the 'Question Under Discussion'; that is, that in cases in which the use of a sentence containing a non-referring definite description is uttered as a legitimate attempt to answer a given question, it will be judged false if there are known true answers to the question. For example, if the question under discussion was 'which if any celebrity buys Froggy Croissants?' then (17) is judged false if it is taken as a legitimate attempt to answer the question because there are true answers. In contrast, if the question under discussion is 'what does the king of France always enjoy in the morning?' then (17) is judged neither true nor false.

Alternatively, some philosophers have explained our divergent intuitions by giving 'accommodation accounts' which aim to pragmatically accommodate any intuitions of falsity to sentences containing non-referring definite descriptions (e.g., Lasersohn, 1993; von Fintel, 2004). For example, according to von Fintel's (2004) account, which I mentioned briefly in Section 2.2.2, any sentence containing a non-referring definite description is truth valueless. The reason that some seem to be false is because we have an independent reason to reject them. Thus he offers a pragmatic account of our false judgements. The idea is as follows: if we are able to falsify a sentence containing a non-referring definite description independently of the presupposition failure, by challenging an entailment of the sentence, then we will. By appealing to these 'independent footholds for rejection', a speaker avoids needing to engage with the presupposition failure. For example, (12) can be rejected if you know that your mother is not dating the King of France, regardless of whether or not the King of France exists.

All of these accounts seek to explain our divergent intuitions. It is not important which, if any, is right because none of them are true on a pure univocal theory. What they all do is presuppose some kind of ambiguity (although this ambiguity is a pragmatic one on the accommodation accounts). They all assume that in some linguistic constructions a reading in the manner of the Frege-Strawson Theory is more natural, but in others a Russellian reading is most natural. While this is problematic for a pure univocal theory, it is not for the Ambiguity Thesis. If a term or phrase is ambiguous, you would expect that in some linguistic contexts or constructions it is given one reading, but in others it is given an alternative reading.

Although a conceptual reading has not been discussed in this literature, there are also some obvious counterexamples to the truth values assigned by this reading as well. These come from cases in which there is an actual *F*. For example, 'the present Queen of England has four children' is considered false on the conceptual reading because there is no conceptual link between the predicate being applied to the *F*. Some might hold a version of the conceptual reading in which it says nothing in cases of contingent sentences. However, this version is still problematic when we think that sentences like the one above are obviously true and sentences like 'the present Queen of England is a child' are false. Thus, if the conceptual reading is a semantic reading then it only seems to assign the right truth values in some instances, namely when the definite description is empty.

#### 2.5.2 Possible alternative explanations

I have argued that the Ambiguity Thesis explains the empirical evidence from the present studies. However, to show that this is the best explanation it is worth discussing some other possible explanations. I will discuss three alternative theories that could explain the present results and argue that the Ambiguity Thesis is a better explanation.

1) One might follow Kripke's (1977) lead and maintain that Russellianism supplemented with pragmatic consideration is still the best explanation. However, proponents of this view are now required to defend it in the face of Studies 3 and 4. These experiments give evidence that the Frege-Strawson Theory should be considered a semantic reading. This is exactly what the Ambiguity Thesis claims, thus, in the face of these experiments the Ambiguity Thesis is the better theory.

It is, however, still an open possibility that the reported intuitions in favour of the conceptual reading might be due to pragmatic factors; Studies 3 and 4 did not rule out this possibility. It might be argued that although many participants in Study 2 indicated that 'analytic' sentences were true (contra Russell and Strawson), this was because they adopted a principle of charity interpreting, where possible, the *speaker* as saying something true, even in cases where it is obvious that what is literally said is false. Further empirical research is needed to rule this interpretation out.

However, even if my hypothesis that there is a *third* semantic disambiguation of simple sentences containing a definite description proves false, this would not undermine the Ambiguity Thesis. It would merely show that there weren't quite as many possible disambiguations.

2) Alternatively, one might think these results are indicative of semantic pluralism rather than semantic ambiguity. As noted in Section 2.3.1, pluralism about definite descriptions holds that sentences of the form 'the F is G' express multiple propositions simultaneously (for example, see Murday, 2014). I will contend that the present empirical results better support the Ambiguity Thesis than a pluralist theory, although, as we will see, there is some reason to think that a pluralist theory is partially supported by the data. Firstly, as noted earlier, participants were reluctant to respond 'Both' in Studies 1 and 2. Indeed, in Study 1 there were no instances in which we see more than 10% of participants respond 'Both'. In Study 2 there is one instance where we see what might be taken as a substantial proportion of participants responding 'Both' (that is, over a noise threshold of 20%) which is 22.9% on the test sentence 'the King of France is a king'. In this instance it does appear that almost a guarter of participants thought that this sentence was both true and false. However, if pluralism were the correct theory, we should expect to see a greater proportion of participants making this response on these analytic truth sentences. It is worth noting that on contingent sentences like 'the King of France is bald', pluralism predicts that participants will respond both 'false' and 'neither true nor false'. However, the option 'Both' in this instance is best interpreted as both true and false (not both false and neither). This might explain why participants inclined towards pluralism did not opt for the 'Both' response on these cases.

The results of Study 4 might be taken as better evidence for the Ambiguity Theory against a pluralist theory. In this experiment participants tended to opt for singular meanings rather than responding that two or three of the options were equally likely. These results indicate that in any given instance participants tended to think that there was only a single meaning expressed by 'the *F* is *G*', but that the meaning expressed was one of multiple possibilities. Indeed, in ordinary language we do not

seem to consider a speaker to express multiple propositions with a single utterance; we take them to be saying only one thing at a time.

However, there is some evidence that pluralism may better capture the meaning of 'the *F* is *G*' in some instances. In Study 4 a small proportion of participants responded that two or three of the options were equally likely. This is consistent with the thought that these participants could discern different possible meanings of 'the *F* is *G*' were expressed simultaneously. But this is not the only reasonable explanation of this observation. As the cases were presented without a supplied context, participants might have considered all of the possible situations in which the utterance might be made and noted that in some cases a Russellian proposition would be more likely and in other cases a referential proposition would be more likely. If this is right, it does not mean that participants necessarily think that multiple propositions would be expressed in any context (as pluralism holds) just that either proposition could be expressed depending on the context (as the Ambiguity Thesis holds).

3) Finally, one might think that the variation in intuitions uncovered in these experiments simply indicates that intuition is not a reliable guide to truth. Many philosophers take their experimental work to be evidence of this (e.g., Machery et al., 2004).<sup>62</sup> Indeed, there are many examples from psychology and cognitive science which indicate that intuition is an unreliable source of evidence in some case; for example, we see this on the Cognitive Reflectivity Test (introduced in Section 1.5.3). Our unreliable intuitions on this test will be illustrated further in my next two chapters. To ask ordinary language users what they think is meant by certain utterances could, therefore, be seen as irrelevant. However, if this is the case then we should not think that it is restricted to ordinary language users. That is, there is no good reason to reject the intuitions of ordinary language users but not philosopher's intuitions on these kinds of cases; there is variation in the intuitions of philosophers as well (see Section 1.5.). If we take this concern seriously it seems we

<sup>&</sup>lt;sup>62</sup> Termed the 'negative program' of experimental philosophy by Alexander, Mallon & Weinberg (2010). This program uses empirical data to show that intuition is unreliable and should not be used as a source of evidence in specific philosophical domains in contrast to the better known 'positive program'.

would be required to stop relying on intuition altogether to answer philosophical questions. I do not think that this is something we should do. Here I have two responses. Firstly, as I argued in Section 1.4, at least with regard to the philosophy of language we should expect our intuitions to be insightful to extent that we consider ordinary language use to partly determine the meanings of terms and phrases in English. Secondly, there are some indications in the present data that intuitions in these cases are informative (though, of course, not completely error free). There is some evidence of systematicity in the responses to the present experiments; for one, in all the experiments we see something like a 60-40 split in intuitions between the Frege-Strawson Theory and Russellianism. This pattern suggests that the responses are not completely random or misinformed. For these reasons, I reject the claim that intuitions are not good sources of evidence. Despite this, I do take the worries some philosophers' have with using intuitions as data seriously. In Chapter 5 I will talk about the benefits of supplementing the evidence from surveys that use intuitions as data by conducting experiments that use other measures; for instance, experiments involving response times or brain imaging (c.f., Chemla & Bott, 2013; and Schwarz, 2014).

#### 2.5.3 Conclusion

In this Chapter I introduced and defended the Ambiguity Thesis as the best theory about definite descriptions as they occur in natural language—or at least in the English language. In the next two chapters I will investigate whether the evidence here in favour of the Ambiguity Thesis is also present for proper names and natural kind terms. Indeed, empirical evidence of a semantic ambiguity for these terms could have the same important implications for the philosophy of language as it has for definite descriptions.

# **Chapter 3: Revisiting Cross-cultural Semantics about Proper Names**

# 3.1 Introduction

One traditional project that philosophers of language have undertaken is to answer important questions about proper names. From the vast literature on this topic, two distinct questions regarding proper names have become prominent:

- (1) *The Reference Question*: How do individual tokens of a proper name come to refer to an individual?
- (2) *The Semantic Question*: What is the literal meaning or semantic value of an individual token of a proper name?

That there is a distinction between these two questions has not always been acknowledged. Thus it has also not consistently been recognised that these two questions need to be addressed in different ways. Two popular ways to answer the Reference Question are the following:

- **Descriptivism:** A token of a proper name comes to refer to an individual in virtue of having a meaning associated with it by the speaker that can be captured by a description. The description, in virtue of being a *definite* description (as opposed to an indefinite description), picks out a unique individual who possesses a salient set of properties (see, for example, Russell, 1905).
- **Causal-Historical Theory:** A token of a proper name comes to refer to an individual in virtue of a causal chain (of the right kind) existing between the tokening of the name and an initial baptism. (see, for example, Kripke, 1980).

In contrast, two of the most common answers to the *semantic* question are the following:<sup>63</sup>

Frege-Russell Theory: The meaning or semantic value of a proper name is captured

<sup>&</sup>lt;sup>63</sup> Another possibility is Devitt (1981). Devitt does not think the semantic value of a proper name is a definite description, nor does he think it is the entity the name refers to. Instead he thinks the semantic value is the causal chain itself. Devitt's theory will be discussed in Section 3.12.

by or is synonymous with a definite description.

**Millianism:** The meaning or semantic value of a proper name is the individual the name refers to.

Sometimes it is thought that there is a necessary connection between the answers we give to each question. It is often thought that Descriptivism is necessarily connected to the Frege-Russell Theory and likewise that the Causal-Historical Theory is ineluctably tied to Millianism. As a consequence, many philosophers have thought that Kripke's arguments against various forms of Descriptivism are also a general refutation of the Frege-Russell theory, but (as we shall see) they are not. Moreover, many philosophers have questioned the details of the Causal-Historical Theory of reference, and concluded on that basis that Millianism is false. In actual fact, although intimately related, the two questions, and their possible answers, should be considered independently. It is possible to mix and match the different theories of meaning and vice versa. For example, one could accept the Causal-Historical Theory of reference can be compatible with different theories of meaning and vice versa. For example, one could accept the Causal-Historical Theory of reference and be a Millian about the meaning of proper names. It is something like this pairing that Kripke has famously argued for; see the following:

My own view, on the other hand, regards Mill as more-or-less right about 'singular' names. [Kripke (1980) p. 127]

Rather, some picture like this is to be held: someone initially 'baptizes' the object, picking out the object perhaps by pointing to it, or perhaps by its properties, or perhaps by some other device. Then—I follow Mill here—speakers wish only to preserve the reference of the name, and as the name is passed from link to link, if one person wishes to use it in the same way as she heard, she uses it with the same reference as the speaker from whom she heard it. [Kripke (2013) p. 13]

But alternatively, one could accept the Causal-Historical Theory of reference and the Frege-Russell theory of meaning. Indeed, it is this pairing that causal descriptivists advocate (see, for example, Jackson, 1998b; Kroon, 1987; Nelson, 2002). To illustrate, consider the following:

Those who still think that descriptive fit determines reference have mostly retreated to a kind of *causal* descriptivism, claiming that what names refer to is determined by descriptions couched in

causal terms). Causal descriptivists thereby express large, perhaps even complete, agreement with the semantic intuitions that causal theorists of reference use to support causalism. They agree with the evidence, but still think that the best explanation for the evidence is a kind of descriptivism [i.e., the Frege-Russell Thesis], and not causalism [i.e., Millianism]. [Kroon (1987) p. 1]

Kripke's semantic argument runs as follows. There is no set of descriptions D such that any speaker competently employing [a name] n associates D with n and uses n to refer to whatever satisfies D. Thus, no set of descriptions determines the referent of n, and so no set of descriptions gives the content of n ... The [advocate of the Frege-Russell Thesis] can block the semantic argument if she can deliver some set of descriptive conditions D such that any speaker competently uses n only if she associates D with n, and such that D is guaranteed to be true of the intuitive referent of n. The following is a popular candidate: (CD) *the entity standing in relation R to my current use of the name n*, [where we] let 'relation R' ... stand for ... some sort of causal-historical chain relation. [Nelson (2002) p. 409-10]

As we can see, causal descriptivism, like the Frege-Russell Theory, holds that the semantic value of a proper name is synonymous with a definite description; however, the kind of definite description is a causal one (see Nelson, 2002; Raatikainen, 2006). Thus, causal descriptivism can be regarded as a version of the Frege-Russell Theory that takes the reference of proper names to behave as the Causal-Historical Theory states, that is, as a causal relation between the utterance of the proper name and the referent. We can see Nelson give an example of what this causal description might be, namely something like the expression 'the entity standing in relation *R* to my current use of the name *n*'.<sup>64</sup> It follows from this kind of description that the content of a proper name from a given utterance is indexed to the specific user of that proper name.<sup>65</sup> Kroon notes that in fact most contemporary supporters of the Frege-Russell Theory are causal descriptivists.

<sup>&</sup>lt;sup>64</sup> However, note that some variants of causal descriptivism do not use indexicals in their analysis. For example, a metalinguistic description theory associates a proper name 'N' with a description like the following 'the thing which is a bearer of 'N'' (e.g., Bach, 1987; Katz, 1990, 1994; Searle, 1983). These theories might be taken as a form of causal descriptivism (Kroon, 1987). However, Raatikainen (2006) in his attack of causal descriptivism does not think that moving away from including indexicals in the description strengthens the theory against his objections.

<sup>&</sup>lt;sup>65</sup> Kroon (1987) presents a case to demonstrate that the meaning of a name depends on the user. In his case, Speaker A makes the statement 'Dumas was the best friend I ever had' which is overheard by Speaker B. Speaker B subsequently asks 'who was Dumas?' Given that Speaker A is referring to his dog who was named after a famous 19<sup>th</sup> century writer, Kroon argues that while a causalist (I will discuss causalism in Section 3.12) is forced to conclude that Speaker B also refers to Speaker A's dog. However, as Kroon notes, we consider Speaker B to be able to use 'Dumas' in another way, such as to enquire about the famous 19<sup>th</sup> century writer.

Indeed, even Kripke, an apparent follower of Millianism, recognises the possibility that the Frege-Russell Thesis might be attached to the Causal-Historical Theory. In this context, he says:<sup>66</sup>

[T]here is a sense in which a description [Frege-Russell] theory must be trivially true if any theory of the reference of names, spelled out in terms independent of the notion of reference, is available. For if such a theory gives conditions under which an object is to be the referent of a name, then it of course uniquely satisfies these conditions. Since I am not pretending to give any theory which eliminates the notion of reference in this sense, I am not aware of any such trivial fulfilment of the description [Frege-Russell] theory ... If any such trivial fulfilment were available, however, the arguments I have given show that the description must be one of a completely different sort from that supposed by Frege, Russell, Searle, Strawson and others. [Kripke (1980) *fn*. 38]

As we can see, the Semantic Question and the Reference Question are two distinct but intricately connected questions. In particular the two questions are connected because the meaning of a term determines its referent. My focus in this thesis is on the Semantic Question. However, because of the intricate connection the answer to this question will have some bearing on the Reference Question as well. So far all of the previous experimental work investigating proper names has been, at least implicitly, trying to answer the Reference Question (I will discuss this experimental work in the next section). However, nothing in this particular experimental literature says anything about reference that would indicate whether the Millianism or the Frege-Russell Theory is right. My aim in this chapter is to address this debate between Millianism and the Frege-Russell Theory by employing the survey methodology. Thus, I will be revisiting previous experimental work has investigated, but with a different goal in mind. What I will argue is that, like the previous chapter, the best way to account for the results of the experiment conducted in this chapter is by appealing to the Ambiguity Thesis. In order to make this claim, I extend the Ambiguity Thesis to include proper names as well as definite descriptions. This I consider to be a very natural extension.

The remainder of this chapter will be set as follows. In Section 3.2 I will describe the experimental work that has been conducted in this area. Most of the experimental work conducted in the domain of the philosophy of language has focused on proper names.

<sup>&</sup>lt;sup>66</sup> It might also be possible to pair Millianism with Descriptivism. However, as I note below, I will assume a Causal-Historical Theory of reference and so am concerned with pairings that include this theory.

However, as we will see, of these experiments most have only shed light on the Reference question. In Section 3.3 I will introduce the present studies, Studies 5 and 6. In doing so I will also introduce the Cognitive Reflexivity Test, a test I will employ in the Study 5, in Section 3.4. In Section 3.5 and Section 3.6 I will describe the experimental method employed and the experimental results respectively. In Section 3.7 I will follow up these results with a US sample of participants in Study 6. Section 3.8 will include a discussion of the present results. I will then go on to discuss why we should defend the informativeness of intuitions in Section 3.9. It is in Section 3.10 that I will argue that the Ambiguity Thesis is the best explanation of the data. In Section 3.11 I will discuss some of the problems often associated with the traditional univocal theories of meaning, Millianism and the Frege-Russell Theory, to show how the Ambiguity Thesis avoids these problems. Section 3.12 will consider some other theories of meaning and discuss the implications they have for the present study. Lastly, in Section 3.13 I will consider the possibility of a pragmatic explanation of the data before giving a brief conclusion in Section 3.14.

## 3.2 Experimental philosophy

Following an early study by Machery, Mallon, Nichols, and Stich (2004) there have been a number of experimental studies aimed at investigating the intuitions of ordinary language users with regard to the Reference Question.<sup>67</sup> This type of research has been driven by the claim that there is uniformity in the intuitions held by various populations (dubbed the 'uniformity conjecture'; see Sytsma & Livengood, 2011). In particular, Kripke (1980), when he proposed the Causal-Historical Theory, made this claim as he assumed that everyone would find it an intuitive account; indeed, the most influential of Kripke's arguments rely heavily on the intuitions elicited from thought experiments. Consequently, these experiments have typically tested whether the intuitions of non-philosophers are in fact consistent with the Causal-Historical Theory as opposed to Descriptivism. In the first of these experiments, Machery et al. presented participants with short vignettes that were based on Kripke's thought experiments. For example, their Gödel case went as follows:

<sup>&</sup>lt;sup>67</sup> Note that although the research I will present does focus on the Reference Question, it is not always explicitly recognised by the philosophers who conduct this research that this is the question being investigated.

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. [Machery et al. (2004) p. 6-7]

They then asked participants to make a judgement about who John is talking about when he uses the name 'Gödel' (i.e., who does the name *refer* to?) providing a Descriptivist option ("the person who really discovered the incompleteness of arithmetic") and a Causal-Historical option ("the person who got hold of the manuscript and claimed credit for the work") to choose between. What they found was a large amount of variation in the responses given by participants. First of all there was a significant difference between the responses given by Western participants and the responses given by East Asian participants (from the University of Hong Kong) wherein the Western participants were more likely to give Causal-Historical responses (56.5%). Besides the cross-cultural variation there was a large amount of intra-cultural variation; that is, within each cultural sample there was a lot of disagreement among participants. Because of the large amounts of variation uncovered in this experiment, Machery et al. concluded that there is no uniformity in the intuitions held by ordinary language users about theories of reference and therefore that we should not rely on intuition as a source of evidence to answer the Reference Question.

While Machery et al.'s (2004) study was one of the first to demonstrate how the survey methodology could be used to inform on questions relevant to the philosophy of language, their study has met with a number of criticisms. Most of these criticisms are aimed at the specific surveys used and so are examples of the objection from the general methodology problems of surveys (introduced in Chapter 1, Section 1.4). These criticisms hold that Machery et al. (2004) fail to make their argument against the informativeness of intuitions

to answer the Reference Question because of confuses in the presentation of the Gödel case. These objections are the following:

 The (Meta)-Linguistic Objection: Martí (2009) claimed that Machery et al. tested the wrong kind of intuition with their probe. By asking who a referring term picks out Machery et al. tested a meta-linguistic intuition whereas Martí argues that a linguistic intuition, about how a referring term is used, would be more probative. For example, Martí suggests that a more appropriate question would be something like:

> One day, the fraud is exposed, and John exclaims: 'Today is a sad day: we have found out that Gödel was a thief and a liar'. What do you think about John's reaction? [Martí (2009) p. 47]

In this case what Martí seems to claim is that the rebuff of John's reaction would illustrate support for the Frege-Russell Theory and the acceptance of it would illustrate support for Millianism. Petho (2005) has made a similar objection arguing that instead of using phrases like "use a name" and "talking about" in the question, the question should *actually* use a name and then talk about it (effectively making it a linguistic question although Petho does not use this term). In response to this objection, I would suggest that the Reference Question and the Semantic Question are being conflated. Because Machery et al. are primarily concerned with the Reference Question, their meta-linguistic probe is entirely appropriate. To change the probe in the way Marti and Petho (independently) suggest could be motivated only if we change the research question to focus on the Semantic Question rather than the Reference Question (but see Machery, Olivola & De Blanc, 2009, for an alternative response).

2. Linguistic Competence Objection: Lam (2011) has argued that the cross-cultural variation in Machery et al.'s (2004) study could be based on a difference in language competency. Specifically, Lam argues that deviations in the East Asian participants' responses might be due to a lack of competency with the English language rather than any real intuitional difference. Notably, Lam went on to empirically test this hypothesis; he presented Gödel type cases to both Cantonese and English speaking participants in each of their native languages. What he found was that Cantonese

speaking participants were just as likely (if not more so) to give Causal-Historical responses as the English speaking participants were, supporting his objection. However, instead of using Machery et al.'s (2004) original probe, Lam used novel probes that were only loosely based on the originals. For example, Lam used a more well-known subject, namely Shakespeare, and included the proper name in the answer choices instead of using definite descriptions as the original probes did (Machery et al., 2010). This change of probe prompted Machery et al. (2010) to respond directly to Lam by translating their original Gödel case for Chinese participants. They found results consistent with their original study, with the Chinese participants less likely to give Causal-Historical responses than the Western participants, suggesting that the original study had not been affected by a lack of linguistic competence (also see Sytsma, Livengood, Sato & Mineki, 2015).

The following two objections are particularly important as they hinge on the semanticpragmatic distinction discussed in Chapter 1, Section 1.2. That is, on the observation that features other than the semantics of language can affect what information is conveyed by a statement. As a consequence participants may interpret survey questions in an unintended way and respond to what the question is taken to ask (pragmatically) rather than what it literally asks (semantically; Horvath, 2010; Kauppinen, 2007). It was objections of this kind that I tried to explicitly rule out in the previous experiment on definite descriptions.

3. The Speaker's Reference Objection: According to this objection there is a difference between whom John is literally referring to (semantic reference) and who the reader might interpret he is intending to refer to (speaker's reference). As such, it has been argued that the variation Machery et al. found in participants' intuitions could be based on the way they interpreted the question (Deustch, 2009; Ludwig, 2007). Machery, Sytsma, and Deutsch (2015) have since clarified the question to make clear that it is the semantic *reference* of the name they are concerned about:

When John uses the name "Gödel," *regardless of who he might intend to be talking about*, he is *actually* talking about: (p.69)

They found similar results to the original (2004) study, suggesting that there is genuine semantic cultural variation in intuitions about the reference of names.

4. The Perspectival Ambiguity Objection: Like the previous objection, this objection posits an ambiguity in the presentation of the Gödel case. Sytsma and Livengood (2011) argued that either the narrator's perspective or John's perspective can be adopted by a participant when answering the question; thus, there is an ambiguity as to who's epistemic perspective is taken. This is problematic as the narrator has more information than John in that the narrator knows that Schmidt discovered the theorem while Gödel merely claimed credit for it, whereas, as far as John knows Gödel discovered the theorem. Machery et al. (2004) assumed that the narrator's perspective would be adopted by participants, but the question asks about who John is talking about. As such, Sytsma and Livengood argued that some participants might interpret the question in such a way that it should be answered from John's perspective. This could explain the variation in responses. Sytsma and Livengood empirically evaluated their objection by clarifying whose perspective is to be taken in the answer choices. For example, to emphasise the narrator's perspective they gave the answer choices:

(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?

They found evidence that Machery et al.'s original probe was ambiguous for Western participants. For example, they found that participants who had the narrator's perspective emphasised were more likely to give Causal-Historical responses (74%) and participants who had John's perspective emphasised were less likely to give Causal-Historical responses (22%) compared to participants who received the original probe (39%).<sup>68</sup> However, more recently Sytsma, Livengood, Sato and Oguchi, (2015) tested for this perspective ambiguity cross-culturally in Japanese populations and found that, while the original question was ambiguous for Western participants, it was not ambiguous (or not nearly as ambiguous) for Japanese participants.

<sup>&</sup>lt;sup>68</sup> Interestingly responses to the original probe do not replicate Machery et al.'s (2004) original results as only 39% (compared to the original 56.5%) gave Causal-Historical responses. This means that a greater proportion of participants from this Western sample actually had Descriptivist intuitions.

Moreover, once this ambiguity was controlled for there were still significant crosscultural differences.

So far all of the objections discussed have focused on the actual methodology employed by Machery et al. in their original experiment. These objections have all been undermined in follow-up experiments by adapting the surveys used. A different kind of objection can be made to Machery et al.'s experiment that does not object to the set-up of the survey but to the informativeness of non-philosophers intuitions.

5. The Expertise Objection: this objection was discussed at length in Chapter 1, Section 1.5 as a general objection to experimental philosophy. It is, however, worth mentioning again here because it is an objection that Devitt (2011) has applied directly to experiments in the philosophy of language, and specifically to experiments like Machery et al.'s (2004) experiment. In this context Devitt has argued that it is experts (i.e., philosophers of language and specialised linguists; see Genone, 2012) who are more likely to have reliable intuitions to Kripke type thought experiments so we should not be concerned with the intuitions of ordinary language users in these cases. Earlier I responded to Devitt and advocates of the Expertise Objection (e.g., in Chapter 1, Section 1.5.5). Here I want to simply sidestep Devitt's objection and say that even if he is right in claiming that we should not rely on the intuitions of ordinary language users when considering Krikpean thought experiments, this would only go to show that we cannot use surveys to inform on the Reference Question. It might be entirely reasonably to still look to intuitions to inform on the Semantic Question.<sup>69</sup>

Indeed, a lot of the criticisms that philosophers have made to the surveying of nonphilosophers' intuitions is only confined to the Reference Question. For example, Jackson (2011) has argued that surveys in this area are uninformative for the reason that they cannot distinguish between some theories of reference; see Jackson the following passage:

There is, however, something that surveys cannot show. They cannot show that, for some

<sup>&</sup>lt;sup>69</sup> See Genone and Lombrozo (2012) for an alternative response. They propose that a hybrid theory of reference, in which both causal and descriptive factors play a role in determining reference, is a better explanation of the variation found in folk intuitions. I discuss this study at length in the following chapter.

group, the description theory of reference for proper names is true whereas, for some other group, the causal theory is true. The surveys we are talking about consist of soliciting responses to vignettes. Subjects are presented with short, or sometimes not so short, descriptions of possible cases and invited to say who if anyone *N* is, or sometimes who if anyone does "*N*" refer to. However, because the cases are *described*, there is no question of their delivering ammunition against the description theory of reference for proper names. What they may do, and in my view do do, is provide evidence that for some groups the key descriptions are certain causal descriptions. [Jackson (2011) p. 114]

Here we see Jackson argue that any survey response thought to provide support for the Causal-Historical Theory could also be explained on a descriptive account; thus, surveys cannot distinguish some types of descriptivism from the Causal-Historical Theory. Ichikawa et al. (2012) make a similar point specifically with regard to Machery et al.'s experiment; consider the following:

Note that a causal descriptivist about names will also say that the correct answer to this question is (B), as may a descriptivist who rejects semantic individualism. So the experiment isn't really testing descriptivism as such versus Kripke's causal historical theory, but only some particular versions of descriptivism against Kripke's theory. These versions of descriptivism say that names refer to the satisfiers of (generally non-linguistic) descriptions that the name's user associates with the name. One such version is 'famous deeds' descriptivism, and the descriptions MMNS use are typically famous deeds. [pg. 57]

However, if this is right, that surveying intuitions cannot distinguish between different theories, I would argue that it only applies to the Reference Question. Surveys that probe for non-philosophers intuitions can inform on the Semantic Question. This is because surveys can be formulated to distinguish between semantic theories. Indeed, this is what the present studies will be shown to do. Moreover, the intuitions of non-philosophers are entirely relevant when considering the Semantic Question; it is the ordinary users of language who determine what the meanings of the words in their language are and moreover their intuitions are reliable about these cases (this point I pushed in Chapter 1, Section 1.3). This is true regardless of whether the intuitions of ordinary languages users can inform on the Reference Question.

The last experiment that I will discuss here is a recent study by Bourget and Chalmers (2013). Unlike all the other experimental work discussed so far Bourget and Chalmers

investigated the intuitions of professional philosophers. Moreover, although they do not explicitly state it, the question being investigated is the Semantic Question, not the Reference Question. In this experiment they surveyed almost 2000 professional philosophers about their views on various philosophical debates. With regard to proper names their question simply read:

#### Proper names: Fregean or Millian?

What Bourget and Chalmers found on this question was that 34.5% of those surveyed identified as a Millian (i.e., in favour of Millianism, 28.7% identified as a Fregean (i.e., in favour of the Frege-Russell Theory), and 36.8% identified as 'other'. What this study indicates is that, at least among philosophers, there is no clear consensus as to the answer to the Semantic Question.

## 3.3 Present studies: Studies 5 and 6

The purpose of the present studies is to investigate the intuitions of *non-philosophers* with regard to the Semantic Question. In a sense this research is motivated by Bourget and Chalmers' experiment (at least with regard to their question on proper names); however, the focus here is squarely on the intuitions of ordinary language users. To this end Studies 5 and 6 will be conducted more similarly to the Machery et al. (2004) study and follow up studies regarding the Reference Question; the main difference will be that the intuitions elicited will be about the semantic content of a proper name rather than the reference. Unsurprisingly perhaps, I uncover a similar lack of consensus among non-philosophers. It is this lack of consensus that I will argue is evidence for the Ambiguity Thesis.

In order to address the Semantic Question I will make an assumption about the answer to the Reference Question. Specifically, I will assume that the Causal-Historical Theory is the correct theory of reference. While the experimental work has uncovered a lot of variation in the intuitions of ordinary language users on this question, at least among philosophers there is something of a general consensus in favour of the Causal-Historical Theory. As I have noted, both philosophers who support Millianism and philosophers who support the FregeRussell Theory tend to take the Causal-Historical Theory to be the correct theory of reference (e.g. causal descriptivists). We can see this in the results of Machery's (2012) study on language experts (discussed in Chapter 1, Section 1.5.5) in which he found that although there was some variation among experts; most had intuitions consistent with the Causal-Historical Theory. Moreover, Machery found that the experts most likely to have read Kripke's Naming and Necessity were even more likely to report this Kripkean intuition (88.6%). Indeed, that the Causal-Historical Theory is considered to be the orthodoxy among philosophers is due largely to Kripke's own arguments. In particular his arguments from ignorance and error have persuaded many philosophers to endorse the Causal-Historical Theory; for one, Descriptivism seems to mistakenly predict that a token use of a proper name will not refer if the speaker is ignorant of any unique information that could pick out the referent. An example Kripke gives is of a speaker who means to refer to Feynman, a well-known theoretical physicist, but might not know anything about Feynman that could distinguish him from any other theoretical physicist such as Gell-Mann. In this instance Kripke argues that intuitively it seems like the speaker does manage to refer to Feynman. Furthermore, Kripke argues that even when something is uniquely picked out by a name Descriptivism sometimes predicts that we refer to the wrong entity. To illustrate this point Kripke gives the original presentation of the Gödel case to argue that while Descriptivism predicts that the utterance of 'Gödel' actually refers to Schmidt, because Schmidt uniquely satisfies the description associated with 'Gödel', this is not the case:

So, since the man who really discovered the incompleteness of arithmetic is in fact Schmidt, we, when we talk about 'Gödel', are in fact always referring to Schmidt. But it seems to me that we are not. We simply are not. [Kripke (1980) p. 84]

Thus three claims I want to make so far are the following:

- Experts are in (general) agreement about the Causal-Historical Theory based on the arguments from ignorance and error.
- With respect to the Semantic Question there is no agreement among experts (see Bourget & Chalmers, 2013).
- 3. With respect to the Semantic Question the intuitions of non-philosophers are

entirely relevant because it is their language that experts have intuitions about.

By assuming the Causal-Historical Theory of reference, I will be testing semantic accounts that are compatible with this picture of reference. In particular I will be testing Millianism against causal descriptivism. Note that this is because causal descriptivism holds that both the Frege-Russell Theory of meaning and the Causal-Historical Theory of reference are correct. Recall that each of these theories of meaning gives a different answer to what the semantic content of a proper name is. Causal descriptivism holds that it is synonymous with a description indexed to the speaker and is therefore determined by the mental state of the speaker. Millianism, in contrast, regards the semantic value of a proper name as the actual referent itself and is therefore independent of the mental state of the user. Thus, the intuition that will be tested in the present studies is whether two agents who have in mind different content (with different descriptive information) can mean the same thing if the actual referent is the same. In particular, the intuitions of non-philosophers will be of significance given the divergence of expert opinions on the Semantic Question (Bourget & Chalmers, 2013).

Another way in which Studies 5 and 6 will be like Machery et al. and other experiments is that the experiment will conduct a cross-cultural comparison between Western participants (from New Zealand and the United States) and East Asian participants (from China) to identify any differences on the Semantic Question. To my knowledge this will be the first experiment to test for cultural differences with specific focus on the Semantic Question.

It is expected that Study 5 will uncover significant variation in the intuitions of ordinary language users. This prediction is based on three observations:

1) The large amount of variation found about ordinary languages users' intuitions about the reference of proper names is an indication that the same might be true about the meaning of proper names given the close connection between the Semantic Question and the Reference Question. Moreover, if we consider the reason that Machery et al. (2004) initially predicted a possible cross-cultural difference (regarding the Reference Question) then it is reasonable to expect the same reason to hold in this instance too (regarding the Semantic Question). Machery et al.'s prediction was based on a large amount of research in cognitive psychology which has found evidence of cultural differences for a number of cognitive processes (see Nisbett, Peng, Choi & Norenzayan, 2001, for an overview). In particular, systematic differences in the way East Asians and Westerners make categorical judgements led Machery et al. to predict that Westerners were more likely to rely on a causal based account of reference (i.e., the Causal-Historical Theory) while East Asians were more likely to rely on a similarity based account (i.e., Descriptivism). Moreover, this body of research has suggested that Westerners tend to attribute causality to an object while East Asians tend to attribute causality to contextual features (Nisbett et al., 2001). What this research suggests is that when it comes to the Semantic Question NZ and US participants will be likely to consider meaning to be constituted by an actual object whereas Chinese participants will be more likely to think that the meaning is determined by features other than simply the object itself but is related to context such as what beliefs are held about the actual referent.

- 2) Bourget and Chalmers' (2013) evidence of variation among philosophers suggests that there will be similar variation among non-philosophers. In this thesis I have tried to show that the expertise argument does not get hold, at least in this domain (e.g., Chapter 1, Section 1.5). If this is right then we should not expect major differences in the intuitions of philosophers and non-philosophers on the Semantic Question.
- 3) Chapter 2 found evidence that simple sentences containing definite descriptions are ambiguous. If proper names are similarly ambiguous we should expect there to be variation in the intuitions reported. Indeed, this is what will be taken as evidence of an ambiguity in proper names in Study 5.

Variation in the reported intuitions will be seen by any significant differences between the cultural groups, significant differences within each cultural group (at least 20% of participants making each of the possible responses), and significant differences within individual participants (i.e., changes in intuition across different cases).

## 3.4 Cognitive reflexivity

As I discussed in Chapter 1, Section 1.5, the Expertise Objection holds that the intuitions of non-philosophers are less reliable than the intuitions of philosophers. One reason this is thought to be the case is because philosophers have more reflective intuitions than non-philosophers (e.g., Devitt, 2011; see Section 1.5.3 in particular). As a further measure in the Study 5 I will also test the reflectiveness of participants to see whether reflection is in fact related to philosophical intuitions. That is, I will test whether people who display more reflective thinking tend to have different intuitions regarding the meaning of proper names than people who display little reflection. If it turns out that philosophically relevant intuitions do not vary as a function of reflective thinking, then there will be good reason to reject this version of the Expertise Objection, that is, the claim that non-philosophers have less reliable intuitions than philosophers because they do not reflect as much on their intuitions.

Reflective thinking will be measured by using Frederick's (2005) Cognitive Reflection Test (CRT), introduced in Chapter 1, Section 1.5.3, which measures one's tendency to challenge one's own intuitions by reflecting upon an intuition rather than accept it as obviously correct (see also Livengood, Sytsma, Feltz, Scheines & Machery, 2010). Previous research using the CRT has indicated that males tend to score higher than females and that scores are positively related to intelligence scores (Frederik, 2005), and that both greater level of education and greater philosophical training are related to higher scores (Livengood et al., 2010).<sup>70</sup>

As a further manipulation, three levels of the CRT will be included: no version of the CRT (control), an identical version of the CRT to previous research (non-reflective version), and a version that emphasises that the questions should be reflected on and thus encourages reflective thinking (reflective version). This additional manipulation will be used to examine whether philosophically relevant intuitions can be primed to be reflective. If a difference is found between intuitions regarding proper names based on CRT version then it will be likely that the different versions primed different levels of reflective thinking.

<sup>&</sup>lt;sup>70</sup> Livengood et al. (2010) in particular were interested in the how philosophers differ in temperament to nonphilosophers and suggested that, given the higher CRT scores exhibited by philosophers, a possibility is that people with more reflective temperaments are attracted towards studying philosophy.

#### 3.5 Study 5: Experimental method

*Participants*: 393 undergraduate students from Victoria University of Wellington, New Zealand and 113 undergraduate students from Nanjing, China (35 from Nanjing University, 67 from Nanjing Normal University, 11 from Nanjing University of Science and Technology) participated voluntarily. Of the New Zealand participants, 193 were approached in a firstyear philosophy course during normal lecture time (the other 200 were approached in a first-year political science lecture) and only 44 had taken any philosophy courses previously. 295 participants in the New Zealand sample identified as being of NZ-European descent, 39 as Maori, 22 as Pacific Islander, and 25 as Asian (some participants identified with more than one ethnic group). Gender was only recorded for the New Zealand sample with 195 males and 194 females. The probes were presented in English to the New Zealand participants and Chinese to the Chinese participants. The original English versions were first translated into Chinese and then back translated and compared for fidelity to ensure no information was lost (see the Appendix A for the Chinese versions).

*Survey and Procedure*: The survey administered was part of a wider research project investigating non-philosophers' intuitions concerning language use in general. Before answering questions about language use participants from the New Zealand sample were presented with one of three versions of the Cognitive Reflexivity Test (CRT). In the "no CRT" version the CRT was not included at all (control). In the other two versions participants received the same presentation of the CRT as in previous studies but with one minor change to Question 2 which originally read "If it takes 5 machines 5 minutes to make 5 widgets". It was thought that this question could be interpreted in two ways, either as "If it takes 5 machines 5 minutes to make 5 widgets each" or "If it takes 5 machines 5 minutes to make 5 widgets in total". As the correct answer (as reported by Frederick, 2005) reflects an interpretation of the question in the latter way, the question was adapted to read "If it takes 5 machines 5 minutes to make a total of 5 widgets". The final presentation of the CRT was as follows:

(1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball.

How much does the ball cost? \_\_\_\_\_ cents

- (2) If it takes 5 machines 5 minutes to make a total of 5 widgets, how long would it take 100 machines to make 100 widgets? \_\_\_\_\_ minutes
- (3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size.

If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? \_\_\_\_\_ days

The two survey versions that included the CRT differed in the statements that preceded and immediately followed the presentation of the CRT. In the "non-reflective" version the preceding statement read:

Below are several problems that vary in difficulty. Try to answer as many as you can:

And the statement following the CRT read simply:

Now turn over and answer the English language questions.

In the "reflective" version the preceding statement was designed to encourage reflective thinking and therefore read:

The three problems below are all difficult (more difficult than they might first seem). <u>Think</u> <u>carefully before you answer</u>.

Reflective thinking was also encouraged across the rest of the survey as the statement following the CRT read:

Now turn over and answer the English language questions. <u>These questions may also be difficult</u>, so again reflect carefully before you answer.

After completing the CRT (all versions) NZ participants moved on to the next part of the survey.

Participants from both samples were presented with two key cases to test their intuitions about proper names. Each key case included the description of a scenario followed by a question designed to elicit relevant intuitions. The first case (Frege's case) was directly modelled on Frege's Puzzle (1892) in which two names are used independently to refer to the same actual entity: In ancient times, Greeks gave the name 'Hesperus' to the bright celestial body first visible to the naked eye in the evenings, and completely independently gave the name 'Phosphorus' to the bright celestial body last visible to the naked eye in the mornings. Indeed, the ancient Greeks believed that Hesperus and Phosphorus were different stars. It was not until many years later that astronomers discovered that Hesperus and Phosphorus were not stars at all, but were the very same planet, namely the planet Venus. Suppose Argos, an ancient Greek, says that "Hesperus is always bright" and Adonia, another ancient Greek, says that "Phosphorus is always bright".

Do you think Argos and Adonia mean the same thing by each of their statements? Choose the answer below you think is most likely to be correct.

- (A) Yes, they do mean the same thing
- (B) No, they don't mean the same thing

The second case (Cat case) followed a similar structure but in this case the same proper name (type) was coincidently used in two independent instances to refer to the same actual entity:

Suppose that one day John finds a stray cat which he decides to adopt. John gives the cat the name 'Kitty' because it looks so cute. John tells his friends that he has adopted a pet named Kitty and says "Kitty is a cat". Now suppose that, when John isn't home, John's cat tends to wander off. John's cat often makes its way into Jane's house and Jane soon becomes fond of the cat. She has no idea that the cat belongs to John. One winter, Jane decides independently to give John's cat the name 'Kitty' just because she likes the name. She then tells her family about Kitty and says "Kitty is a cat".

Do John and Jane mean the same thing by their utterance "Kitty is a cat"? Choose the answer below you think is most likely to be correct.

- (A) Yes, they do mean the same thing
- (B) No, they don't mean the same thing

For both cases, (B) indicated support for causal descriptivism and (A) indicated support for Millianism.

Participants from the NZ sample were also presented with a control case (Aristotle case) in which it was made clear that two independent uses of the same proper name referred to different actual entities:

Aristotle was a philosopher in ancient Greece. He was the student of Plato and the teacher of Alexander the Great. He is considered one of the most important philosophers in history. Aristotle Onassis was a well-known Greek shipping magnate in the 1900's. He married Jackie Kennedy after the death of her first husband, John F. Kennedy. Suppose Plato, in ancient Greece, says "Aristotle is a great man" and Jackie Kennedy, just before her second marriage, says "Aristotle is a great man".

Do you think Plato and Jackie Kennedy mean the same thing when they each say "Aristotle is a great man"? Choose the answer below you think is most likely to be correct.

- (A) Yes, they do mean the same thing
- (B) No, they don't mean the same thing

The Aristotle case was included to deter participants from changing their answers to one of the key cases based on their answer to the other key case. For the New Zealand sample the order in which the two key cases were presented was reversed so approximately half of the participants received Frege's case first and the Cat case second (Order-a) while the remaining participants got the reverse order (Order-b). The Aristotle case was always presented second (i.e., in between the two key cases). Participants in the Chinese sample always received Frege's case first.

#### **3.6 Experimental results**

*CRT*: Participants who received the CRT were given a score from 0-3 based on how many of the items they correctly answered. The mean score was 1.16 (*SD*= 1.06). Although participants given the "reflective" version (M= 1.22, *SD*= 1.09) of the CRT scored higher than participants given the "non-reflective" version (M= 1.09, *SD*= 1.03) as expected, this difference was not found to be significant [U = 7072.00, p = .36].

*Proper name cases*: Answers were scored binomially. For each probe, (A) answers (i.e., answers indicative of Millian intuitions) were given a score of 0 and (B) answers (i.e., answers indicative of causal descriptivist intuitions) were given a score of 1. Scores were then also summed for the key cases (Frege's case and the Cat case) so the total range was 0-2. Of the accumulated scores, the mean for the NZ sample was 0.77 (*SD*= .68), and the mean for the Chinese sample was 1.37 (*SD*= .70) indicating that participants in the NZ sample were more likely to endorse Millianism and participants in the Chinese sample were more likely to endorse causal descriptivism. This difference between the samples was found to be significant [*U* = 11077, *p* < .001]. Furthermore, chi-square tests found significant differences between the samples for each of the key cases (see Figure 1) [ $\chi^2$  (1, 486) = 45.30, *p* < .001 and  $\chi^2$  (1, 483) = 7.88, *p* = .005 for Frege's case and the Cat case respectively].

Binomial tests also indicated significant variation in responses within each cultural sample. In each of the Chinese and the NZ samples at least 20% of the responses were (A) and at least 20% were (B) (for each of Frege's case and the Cat case; see Figure 1).<sup>71</sup> This indicates that there was substantial support for each of Millianism and the Frege-Russell Theory in each sample that cannot be explained as noise.

<sup>&</sup>lt;sup>71</sup> In all binomial tests p < .003.



*Figure 1.* Percentages of participants who responded in favour of answer (A) indicating support for Millianism as opposed to causal descriptivism for the NZ sample and the Chinese sample..

Within each sample the consistency in responses to each of the cases was also examined (see Table 1 below for means and standard deviations). For the Chinese sample no difference was found between answers to Frege's case and answers to the Cat case (Z = .31, p = .76). Because the NZ sample received the Aristotle case as well as the key cases all three cases were compared for these samples. A Friedman test indicated a significant difference between the probes ( $\chi 2 = 166.71$ , p < .001). Follow-up Wilcoxon signed ranks tests indicated significant differences between Frege's case and the Cat case (Z = 6.04, p < .001), Frege's case and the Aristotle case (Z = 12.18, p < .001).
In both samples a considerable proportion of participants (48.6% of the NZ sample, 37.8% of the Chinese sample) answered differently to the key cases (i.e., answered (A) to Frege's case and (B) to the Cat case or vice versa) indicating that within individual participants intuitions about proper names were often not stable across scenarios. Because a lower proportion of the NZ sample (77%) correctly answered (B) on the Aristotle case than expected, the key cases were also examined when participants who incorrectly responded to the Aristotle case were excluded. Similar responses to the key cases were found after exclusion based on response to the Aristotle case as before the exclusion (M= .50, SD= .50, and M= .28, SD= .45 for Frege's case and the Cat case respectively). Furthermore, a similar proportion (49.1%) of participants answered differently to the key cases as before the exclusion.

The following results focus exclusively on the NZ sample for which extra data was collected. Firstly, a significant effect of order was found for both the Cat case (U = 15539.00, p = .031) and the Aristotle case (U = 16279.00, p = .032) where participants in Order-a (Cat case: M = .22, SD = .42; Aristotle case: M = .72, SD = .45) were more likely to answer (A) than participants in Order-b (Cat case: M = .32, SD = .47; Aristotle case: M = .81, SD = .39). No significant order effect was found for Frege's case. No significant differences were found between classes (philosophy vs. non-philosophy). The only significant gender difference found was on the Aristotle case with females (M = .81, SD = .40) more likely to correctly answer (B) than males (M = .72, SD = .45) [U = 16013.50, p = .042].

Although it was expected that there would be a difference in responses based on CRT version, the only difference found between the three CRT versions (none, reflective, non-reflective) was on the Cat case ( $\chi$ 2= 7.27, p = .026). Tukey post-hoc comparisons of the three CRT versions indicated that participants who received the 'no CRT' version (M = .20, SD= .40) were significantly more likely to endorse (A) answers than participants who received the 'reflective' version (M = .33, SD= .47). An examination of participants who received a version of the CRT (the reflective version and the non-reflective version) indicated no relationship between CRT score and answers to the proper name cases.

## 3.7 Study 6: US sample follow up

A follow up study with a US sample was also conducted. Fifty-seven United States residents participated voluntarily as part of an online free personality test. In this sample the mean age of participants was 36.63 (*SD* = 14.04) and included 14 males and 43 females. All US participants indicated that they were competent with English. All participants received the probes in the following order: Frege's case, the Aristotle case, the Cat case (i.e., Order-a).

The data was analysed in the same manner as the other two samples. The mean accumulated score of Frege's case and the Cat case was 0.44 (*SD*= .60), indicating that, like the NZ sample, participants were more likely to endorse Millianism (see Table 1 for proportions of responses to each case). . A Friedman test indicated a significant difference between the probes ( $\chi 2 = 34.60$ , p < .001). Follow-up Wilcoxon signed ranks tests indicated significant differences between the Aristotle case and Frege's case [Z = 4.64, p < .001] and the Aristotle case and the Cat case [Z = 4.67, p < .001]. However, as with the Chinese sample, there was no significant difference between answers to Frege's case and answers to the Cat case. Although a direct comparison to the other samples is not possible, the responses of the US sample appear to be similar to the NZ sample (see Table 1 for a comparison). The main difference between the US and NZ samples seem to be that there is greater support for Millianism on Frege's case from the US sample. Indeed, the NZ sample shows about equal support between each theory on this case.<sup>72</sup>

Like in the Chinese and the NZ sample, there was significant variation in the responses of the US sample as at least 20% of responses were (A) and at least 20% of responses were (B) for each of Frege's case and the Cat case. This indicates that the variation cannot be explained as noise.<sup>73</sup>

<sup>&</sup>lt;sup>72</sup> Although a direct statistical comparison cannot be made between the US sample and the other two samples, because of relevant differences in the way the study was conducted between the samples, we can note that a statistical comparison using Kruskal-Wallis tests would find significant differences between all three samples [ $\chi 2$  (2, 543) = 34.23, p < .001, and  $\chi 2$  (2, 540) = 63.76, p < .001 for Frege's case and the Cat case respectively] and in particular, between the US sample and both the other samples on Frege's case, and between the US sample and the Chinese sample only on the Cat case (using Mann Whitney comparisons). <sup>73</sup> Note that binomial tests indicate that there was not significantly greater than 20% of (A) responses on either Frege's case or the Cat case, p = .473 and p = .347.

Only 72% of participants correctly answered (B) on the Aristotle case. As with the NZ sample, even when excluding participants who did not answer the Aristotle case correctly, there were similar responses to the key cases (M= .17, SD= .38, and M= .22, SD= .42 for Frege's case and the Cat case respectively). Furthermore, a similar proportion of participants answered differently to the key cases as before the exclusion (29.3%).

Table 1. shows the mean scores to each probe for each of the three samples when (A) responses were scored as 0 and (B) responses were scored as 1.

	Score (SD)
Frege's Case	
NZ participants	.49 (.50)
Chinese participants	.68 (.47)
US participants	.21 (.41)
Cat Case	
NZ participants	.28 (.45)
Chinese participants	.68 (.47)
US participants	.23 (.42)
Aristotle Case	
NZ participants	.77 (.42)
US participants	.72 (.45)

Table 1. Mean scores (SD in parentheses)

#### 3.8 Discussion

The primary objective of the present research was to inform on the Semantic Question by investigating whether the intuitions of ordinary language users align with Millianism or the Frege-Russell Theory. The results of Studies 5 and 6 indicate substantial variation between the intuitions held by participants. Indeed, there were discrepancies in the way participants responded at all levels: between cultural samples, within cultural samples, and within individual participants. The most notable findings are the cross-cultural differences which suggest that Chinese participants tend to hold intuitions consistent with the Frege-Russell Theory whereas NZ and US participants, in comparison, tend to hold intuitions more consistent with Millianism. Although there was some variation between the responses of the NZ participants and the US participants, the difference appears to be driven almost exclusively by the responses to Frege's case by the NZ sample. Indeed, there was substantial variation in the response of NZ participants to this case with fairly equal support for the two alternatives. However, despite this one discrepancy the NZ and US samples responded very similarly. Thus the main contrast found between the cultural samples is between the Chinese sample and the two Western samples.

Although the focus of Studies 5 and 6 is on the Semantic Question, it is easy to see some similarities between these results and the results of previous studies that have focused on the Reference Question. In particular, the present cross-cultural differences are very like the differences between Western and East Asian participants reported by Machery et al. (2004) in which they found that Western participants tended to hold more Causal-Historical intuitions and East Asian participants tended to hold more Descriptivist intuitions.

Even when cross-cultural differences are taken into account, the present results still indicate a large amount of variation in responses. This is especially evident for NZ participants in Frege's case for which the division between Millian and Frege-Russell responses is almost a 50-50 split. It is interesting to note that there is not a similar split in the Cat case in which the NZ sample indicates strong support in favour of Millianism. This is especially curious considering that Frege's case and the Cat case are analogous with the exception that in the Cat case the same proper name is used in two instances independently. In Frege's case, by contrast, different proper names are used in each token use. The difference in intuitions for

Frege's case and the Cat case suggest that participants in the NZ sample were influenced by the irrelevant factor of whether or not the proper names used were of the same type. The results from the Chinese and US samples, on the other hand, did not exhibit any significant difference between the two cases; although, there was still variation in the responses with approximately 30% of Chinese participants for each case supporting Millianism, unlike the majority of the sample, and approximately 20% of US participants supporting the Frege-Russell Theory, unlike the majority of that sample. Interestingly, this intra-cultural variation is similar to the variation Bourget and Chalmers (2013) found among professional philosophers. In their study there was almost equal divides between support for Millianism, support for Frege (i.e., the Frege-Russell Theory) and support for 'other'.

# **3.9 Defence of intuitions**

I have argued that when considering the Semantic Question the intuitions of ordinary language users are entirely relevant because it is their language and they in part determine the semantic value of the terms. This claim challenges the Expertise Objection, which regards the intuitions of non-philosophers as essentially irrelevant to philosophical questions. In this section I want to defend the probity intuitions. I will argue that the present studies provide evidence that the intuitions of ordinary language users about the Semantic Question are not less relevant than the intuitions of professional philosophers (or not substantially so). Then I will explain the variation uncovered in the present studies by appealing again to the Ambiguity Thesis (Section 3.10).

The similarity between the responses found in the present studies and in Bourget and Chalmers' (2013) study suggests that there is not an inherent difference between the intuitions of philosophers and non-philosophers with regard to the Semantic Question. Although we cannot do a direct comparison between Bourget and Chalmers' experiment and Studies 5 and 6, each suggests there is no consensus among their respective participants as to the answer to this question; and, moreover, each study indicates that both Millianism and the Frege-Russell Theory are considered viable alternatives. This is noteworthy given that the Expertise Objection assumes that there *will* be differences

between the intuitions that philosophers hold compared to the intuitions of nonphilosophers. Furthermore, the inclusion of the CRT in Study 5 provides evidence that having more reflective intuitions in general does not impact philosophically relevant intuitions. There was no evidence that participants who scored highly on the CRT responded differently on the key cases to participants who received low CRT scores. Furthermore, there was only minimal evidence that CRT group impacted the way participants responded to the key cases as the only significant difference found was that participants in the 'no CRT' group were more likely to endorse (A) on the Cat case than the 'reflective' group. These results appear to suggest that general cognitive reflexivity did not affect philosophically relevant intuitions (or did so minimally) in the present study. What this might suggest is that reflection does not improve philosophically relevant intuitions. It might be that it is impossible to externally corroborate the truth of philosophically intuitions; that is, there is no way to improve our intuitions because there is no source to show us when we are right. Therefore reflection does not mean that better intuitions are generated in these cases (this point was discussed in Chapter 1, Section 1.5). Importantly, these findings are indicative that at least one version of the Expertise Objection does not get hold on the Semantic Question. This is because if philosophical intuitions do not vary as a function of how reflective someone is then it seems unlikely that philosophers will have significantly better intuitions simply because they have been trained to think reflectively.

One present concern if we are defending the probity of intuitions is the number of participants who did not answer the Aristotle case as expected. Unlike the two key cases, the Aristotle case had a correct answer (B) and an incorrect answer (A). The reason the Aristotle case was included was to distract participants from changing their answer to one of the key cases based on their answer to the other key case; it was not expected that there would be a substantial error in responses. As it turned out, although the Aristotle case had the least amount of variation in responses, there was more variation than might be considered desirable for an account that wants to defend intuitions. This is problematic as the Aristotle case can serve to indicate a base error rate of ordinary language users' intuitions. If error rates are higher than desired at a base level then we can infer that the key cases have considerable error rates. However, after excluding participants who did not answer the Aristotle case correctly, there was very little difference in the proportion of

responses to the key cases. Furthermore, in the NZ sample there was no statistical difference between the intuitions of males and females on the key cases even though females were significantly more likely to respond to the Aristotle case correctly. This suggests that having more reliable intuitions in some instances does not necessarily impact philosophical intuitions. Interestingly, it is by including these kinds of control cases that mean we might be able to respond to the 'The Objection from the General Methodological Problems of Surveys' (Chapter 1, Section 1.4) as they give us a way to assess what kind of error rates we are dealing with in a given survey.

## 3.10 The Ambiguity Thesis

In the previous chapter I argued that the best explanation for the data concerning definite descriptions was the Ambiguity Thesis. Here I want to propose that the Ambiguity Thesis can also be applied to the case of proper names. In this context the Ambiguity Thesis might be revised as:

**The Ambiguity Thesis:** proper names are ambiguous and have multiple possible meanings associated with them. A native speaker can use a definite description either with the content suggested by Millianism or as suggested by the Frege-Russell Theory depending on their communicative intention; no one disambiguation is correct outside of a context of utterance.

To encapsulate definite descriptions and proper names, as well as any other parts of language that might be similarly captured by the Ambiguity Thesis, it might also be worth specifying a general version of the Ambiguity Thesis:

**The Ambiguity Thesis (general):** some terms are ambiguous and have multiple possible meanings associated with them. A native speaker can use these terms and phrases with any of the possible contents depending on their communicative intention; no one disambiguation is correct outside of a context of utterance.

With regard to the present experiment, the Ambiguity Thesis explains the variation in responses as the meaning of proper names in some instances being disambiguated in the

manner of Millianism and in other instances in the manner of the Frege-Russell Theory. The advantage of the Ambiguity Thesis is that it allows for such variation between different participants as well the variation within individual participants across the different cases. In the first instance the variation is due to different participants disambiguating the same case in different ways. In the second instance the variation is due to the same individual disambiguating a proper name in different ways in different contexts. This is consistent with the present results in which it was found that individual participants did not have robust intuitions across the two cases. Indeed, approximately 40% of participants appeared to have Millian intuitions on one case and causal descriptivist intuitions on the other. Consequently, all the variation in ordinary language users' intuitions, uncovered in the present studies, as well as the variation in philosophers' intuitions, uncovered in Bourget and Chalmers' study, is allowed for by the Ambiguity Thesis.

Moreover, in this context we can extend the explanatory power of the Ambiguity Thesis to explain the cross-cultural differences, which were not looked at in the last chapter.<sup>74</sup> It explains cross-cultural differences as cultural language norms dictating which disambiguation is considered more probable. That is, it holds that context plays different roles in different languages. So in English context might make a Millian disambiguation the more likely interpretation, whereas in Chinese the same context might make a Frege-Russell disambiguation the more likely interpretation. This account of cross-cultural differences in intuitions is one previously unconsidered.

Overall, the Ambiguity Thesis, I argue, has the following important advantages:

- It can explain the variation in non-philosopher's intuitions uncovered in Studies 5 and 6,
- It can explain the variation in philosopher's intuitions reported in Bourget & Chalmers (2013),
- 3. It does not require us to abandon our commitment to using intuitions as a useful source of evidence in this domain, and,

<sup>&</sup>lt;sup>74</sup> Indeed, many East Asian languages do not distinguish between definite and indefinite descriptions e.g. Chinese.

4. Postulating this ambiguity allows us to retain all of the benefits of both Frege-Russell theory and Millianism without inheriting the costs associated with either.

Indeed, the Ambiguity Thesis allows the experimental philosopher to respond to the commonly made criticisms to the programme. First, we can respond to the objection that intuition is not a good source of evidence, at least for determining the contents of a proper name because there is too much intuitional variation to account for. The objection is misguided because variation in intuitions is exactly what we should expect to see if the Ambiguity Thesis is true. Second, we can respond to at least one version of the Expertise Objection because we have shown that the variation in intuitions does not come about because of a difference in reflectiveness. In the remaining sections, I will discuss in more detail why the Ambiguity Thesis has the other advantages outlined above.

One might argue at this point that while this experiment is evidence for a hybrid theory of meaning, it is not specifically indicative of an ambiguity. Recall from the previous chapter that a hybrid theory of meaning holds that there are two or more semantic elements associated with an term A, which work together in some way to determine A's referent or extension in any possible world. Recall also that there are multiple views consistent with this theory, such as the following three:

- 4. Ambiguity View. If a term A is ambiguous, then there are straightforwardly at least two different meanings associated with A in a language. On any occasion of utterance, A will have just one of those meanings associated with it, and which one will be determined by the context. This is the view about definite descriptions, proper names and natural kind terms advocated in this dissertation.
- 5. **Pluralist View.** On the pluralist view, a term A has multiple meanings associated with it on any occasion of utterance. The term is not (or at least need not be) ambiguous on this conception because the single term can consistently have the same contents associated with it on every occasion utterance. This kind of view is advocated by—or at least consistent with—David Chalmers' two-dimensional semantics.
- 6. **Component View.** Unlike on the ambiguity and pluralist views, the mixed view tells us that a term A might have a *single* meaning or semantic value associated with it,

but that different components work together to determine what that meaning is. If you think the (single) meaning of an expression is captured by a conceptual analysis, you will accept that there are simpler concepts that can be thought of as *components* of the meaning that when conjoined make up the meaning of the expression.

The results of Studies 5 and 6 look like they are best taken as evidence of an ambiguity in proper names. However, it might be that these experiments do not rule out the possibility of another hybrid theory. Although it does not look like the present results support a component theory, because the evidence favours the view that there is more than one meaning associated with a proper name, it might be that there is evidence of a pluralist view. In particular, because participants were given a forced binary choice between the responses (A) and (B), there was no response available that would have served as evidence for a pluralist view, that is, the option to respond (A) *and* (B) on an individual case. A further experiment could allow participants to select both responses. If it was found that a majority of participants thought that both options were correct in every instance, then there would be strong evidence for a pluralist view over an ambiguity. In the next chapter this possibility is allowed for in an experiment concerning natural kind terms.

## 3.11 Problems with traditional theories

This section aims to discuss the problems that different theories of meaning face with the intent of showing that by endorsing the Ambiguity Thesis we avoid all these problems. Firstly, Millianism faces the following problems:

1. The Problem of Empty Names – Millianism gives no explanation for what the meaning of an empty name is. For example, it appears that the names like "Pegasus" and "Zeus", which have no referent, are nonetheless meaningful and can be used informatively in sentences to tell us something. Furthermore, negative existentials, such as "Pegasus does not exist," appear to be both meaningful and true when containing an empty name. However, according to Millianism, empty names are meaningless as they have no semantic value and so sentences containing them can

be neither true nor false (see Braun, 1993).

2. Frege's Puzzle – Millianism makes counterintuitive predictions about identity statements as it predicts that names that refer to the same thing are interchangeable as the semantic value of each is identical. However, in some cases there seems to be a significant difference between what the statements containing co-referential proper names mean. Specifically, "Frege's Puzzle" shows that while the names "Hesperus" and "Phosphorus" both refer to Venus, they are not used equivalently. To an ancient astronomer "Hesperus is Phosphorus" would appear to be a false statement and one they would find very informative if confirmed. "Hesperus is Hesperus," on the other hand, would appear to be obviously true and completely uninformative (Frege, 1892).

It is commonly assumed that these two problems are problems for Millianism specifically, and that descriptivists face no difficulty in this domain (but cf. Brock 2004). Advocates of Millianism have attempted to address these two problems, but there are residual difficulties. For example, one proposed solution to 'The Problem of Empty Names' goes as follows: sentences that contain a proper name express a singular proposition with an ordered pair of individual and property. Sentences that contain an empty name express a proposition that closely resembles a singular proposition in structure with the exception that there is a missing semantic value. That is, there is a gap in the individual position rendering it a 'gappy proposition'.<sup>75</sup> According to this view, a gappy proposition is false because the actual subject of the proposition does not have the property expressed by the predicate. For example, the sentence 'Pegasus is a winged horse' expresses the proposition <GAP, is a winged horse> which is false in virtue of there being no actual referent. Furthermore, because simple gappy propositions turn out to be false, complex gappy propositions can be true. For example, the negation of the same sentence expresses the proposition <NOT <GAP, is a winged horse>>, which is true. Thus, gappy propositions give us a way to respond to 'The Problem of Empty Names' that corresponds with our intuitions regarding the use of empty names. However, Brock (2004) has argued that this solution does not adequately resolve the problem because we are left with the counterintuitive

<sup>&</sup>lt;sup>75</sup> See, for example, Adams, Fuller & Stecker (1997), Braun (1993), Oppy (1997); see Brock (2004).

consequence that any sentence that differs only by what empty name is contained therein will express the same proposition. Consider the following sentences:

- (1) According to Greek legend, Pegasus is a winged horse
- (2) According to Greek legend, Zeus is a winged horse

The advocate of the gappy proposition analysis needs to regard the embedded sentences in (1) and (2)—that is, the simple sentences that fall within the scope of the operator 'according to Greek legend'—as expressing the same proposition. Consequently, we should expect that (1) and (2) have exactly the same truth value. Intuitively, they do not. So the proposed solution does not solve 'The Problem of Empty Names' for the Millian.

As a proposed solution to 'Frege's puzzle', Salmon (1988) has argued that it is a mistake to think that Millianism is committed to the view that someone who believes that "Hesperus is Hesperus" also believes that "Hesperus is Phosphorus". Salmon argues the reason for this is because we can make a distinction between the following assumptions:

- (1) For every x, A does not believe that  $x \neq x$
- (2) For every x, A does not believe that x is self-distinct

The reason it is thought that Millianism is committed to think that "Hesperus is Phosphorus" is true is because (1) is assumed. However, Salmon argues that while (2) is likely true because we do not expect things to be distinct from themselves, it does not mean that (1) is also true. A speaker may, to the contrary of (1), believe that  $x \neq x$  without believing that x is self-distinct. Thus by stating "Hesperus is not Phosphorus" only a belief about the relationship between x and x is asserted; it is not asserted (or likely believed) that Hesperus is not self-distinct. However, as with the proposed solution to 'The Problem of Empty Names', Salmon's solution to 'Frege's Puzzle' has not been widely accepted.<sup>76</sup>

The Frege-Russell Theory of meaning was initially proposed because it avoided these two problems. However, it is worth noting two things; firstly, any advantage that the Frege-Russell Theory has over Millianism regarding the 'The Problem of Empty Names' and

<sup>&</sup>lt;sup>76</sup> There are other proposed solutions to Frege's Puzzle in line with Millianism that have also been put forth e.g., Richard (1987), Salmon (1986), and Soames (1989); See McKay and Nelson (2014) for an overview.

'Frege's Puzzle' is equally shared by the Ambiguity Thesis. Secondly, the Frege-Russell Theory inherits problems of its own that are *not* also shared by the Ambiguity Thesis. For one, Kripke (1980) famously attacked the theory in a number of arguments including his modal argument. The modal argument challenges the claim that proper names are synonymous with a definite description. Kripke famously puts the idea as follows:

Suppose the reference of a name is given by a description or a cluster of descriptions. If the name *means the same* as that description or cluster of descriptions, it will not be a rigid designator. It will not necessarily designate the same object in all possible worlds, since other objects might have had the given properties in other possible worlds, unless (of course) we happened to use essential properties in our description. So suppose we say, 'Aristotle is the greatest man who studied with Plato'. If we used that as a *definition*, the name 'Aristotle' is to mean 'the greatest man who studied with Plato'. Then of course in some possible world that man might not have studied with Plato and some other man would have been Aristotle. If, on the other hand, we merely use the description to *fix the referent* then that man will be the referent of 'Aristotle' in all possible worlds. The only use of the description will have been to pick out to which man we mean to refer. But then, when we say counterfactually 'suppose Aristotle had never gone into philosophy at all', we need not mean 'suppose a man who studied with Plato, and taught Alexander the Great, and wrote this and that, and so on, had never gone into philosophy at all' [Kripke (1980) p. 57].

What Kripke's modal argument comes down to is:

- (1) 'Aristotle might not have taught Alexander' is unambiguously true.
- (2) 'The teacher of Alexander might not have taught Alexander' is ambiguous.
- (3) Therefore, 'Aristotle' does not mean 'the teacher of Alexander'.

The modal argument shows that the meaning of a proper name is not captured by a definite description because definite descriptions, unlike proper names, are not rigid designators.

There are two common lines of response to the modal argument that have been given by defenders of the Frege-Russell Theory. The standard response holds that the descriptions associated to proper names can be rigidified in such a way that the modal argument does not get hold. For one, a description could be rigidified by being indexed to the actual world,

for example, 'the *actual* teacher of Alexander'. According to this response if (2) is modified to contain a rigidified description then it will be false because the description will be unambiguous. However, this response has been challenged. Brock (2004), for one, argues that if proper names are synonymous with rigidified descriptions then we run into problems when considering sentences like:

- (4) Pegasus might have existed.
- (5) The actual mythical winged horse might have existed.

Brock argues that (5) is false because if there is nothing that fits the description then it should be obvious that it is not possible that something fit the description. The problem is that we typically consider sentences like (4) to express true propositions. However, according to the rigidified version of the Frege-Russell Theory, 'Pegasus' means 'the *actual* mythical winged horse' so (4) is also false.

The second line of response to the modal argument is Dummett's wide-scope descriptivist response.<sup>77</sup> In essence Dummett argues that, contrary to the modal argument, (1) and (2) do not entail (3) because proper names are synonymous with descriptions that are required to be read as taking wide-scope over modal operators. To see this consider the two ways that (2) can be interpreted:

- (2.1) It is possible that the teacher of Alexander did not teach Alexander.
- (2.2) The teacher of Alexander is such that *he*—that man—might not have taught Alexander.

The difference between (2.1) and (2.2) is a difference in the scope taken by the description. In (2.1) the description is given narrow-scope as the modal operator extends over the entire proposition. Thus, if we assume Russell's theory of descriptions, (2.1) can be formulated in logical notation as:

(2.1)\* ◊∃*x* (*F*!*x* & ~*Fx*)

<sup>&</sup>lt;sup>77</sup> Also see Sosa (2001).

What  $(2.1)^*$  expresses is the false proposition that it is possible that there is an x such that x is uniquely F and that x is not F. In (2.2), on the other hand, the description is given widescope over the modal operator. Thus (2.2) can be formulated in logical notation as:

 $(2.2)^* \exists x (F!x \& \diamond Fx)$ 

What  $(2.2)^*$  expresses is the true proposition that there is an x such that x is uniquely F and it is possible that x is not F. Thus, wide-scope descriptivism accounts for the apparent ambiguity of sentences like 'the teacher of Alexander might not have taught Alexander' as an ambiguity between the possible logical notations available. The wide-scope descriptivist's claim is that if names are required by linguistic convention to take wide-scope over modal operators then the correct way to interpret the equivalent descriptive sentence is captured by  $(2.2)^*$ . On this reading both 'Aristotle might not have taught Alexander' and 'the teacher of Alexander might not have taught Alexander' have the same truth values and the modal argument fails. Kripke has, however, maintained that the modal argument can be reformulated in a way that scope distinctions do not apply. Consider the following:

- (1.1) The simple sentence 'Aristotle taught Alexander' is not necessary.
- (2.1) The simple sentence 'The teacher of Alexander taught Alexander' is necessary.
- (3.1) Therefore, 'Aristotle' does not mean 'the teacher of Alexander'.

The sentences evaluated in this modified argument are simple sentences. The simple sentences do not include modal operators. Thus, on this reading, scope ambiguities do not arise.

My view is that wide-scope descriptivism goes both wrong and right. Where I agree is that the key is positing an ambiguity. However, while wide-scope descriptivism posits that proper names are equivalent to a description that is ambiguous, the Ambiguity Thesis postulates that the ambiguity is in the proper name itself. Here it should be noted that causal descriptivism goes some way to overcome the problems associated with the Frege-Russell Theory as it makes a very specific claim about what kind of description is associated with a

proper name.<sup>78</sup> However, in doing so causal descriptivism inherits its own set of problems.<sup>79</sup> The advantage of postulating that proper names are ambiguous is that the Ambiguity Thesis is in a position to answer all of the challenges that have faced both Millianism and the Frege-Russell Theory. This is because the advocate of the Ambiguity Thesis is able to disambiguate the semantic content of a proper name in at least two different ways, as the actual object or as a definite description. Thus, in any given instance a non-problematic interpretation of a proper name is available. Consider the two problems associated with Millianism. The Ambiguity Thesis can avoid 'the problem of empty names' by disambiguating an empty name in the same way as the Frege-Russell Theory, that is, it can hold that the semantic content of a proper name is synonymous with a definite description. In the same way 'Frege's puzzle' is avoided as the Ambiguity Thesis can hold that proper names that appear in identity statements need to be disambiguated consistently with the Frege-Russell Theory. Thus, it can account for the apparent disparity between the sentences 'Hesperus is Phosphorus' and 'Hesperus is Hesperus' by associating different descriptions with 'Hesperus' and 'Phosphorus'. Furthermore, Kripke's modal argument positioned against the Frege-Russell Theory does not arise because the Ambiguity Thesis is comfortable with the claim that a proper name is not synonymous with a definite description in cases when it is problematic to do so. Thus, in the case of Kripke's (1980) problematic examples like "suppose Aristotle had never gone into philosophy at all" [p. 57] the Ambiguity Thesis can agree that 'Aristotle' does not mean 'the teacher of Alexander' but means instead that man, as Millianism asserts. In summary, common semantic theories of proper names face fundamental problems that, despite considerable debate, tend to be viewed as drawbacks

(a) *a* believes that *N* is *G* 

<sup>&</sup>lt;sup>78</sup> As opposed to the usual 'famous deeds' descriptions associated traditionally. Famous deeds descriptivism as a variant of the Frege-Strawson Theory will discussed in the next section. Here it is simply worth noting that generally when we talk about the Frege-Strawson Theory it is this kind of description that is assigned.

<sup>&</sup>lt;sup>79</sup> For example, some philosophers have argued that because meaning is conceptualised as the mental state of the user, causal descriptivism is unusable as a theory of shared linguistic meaning (Raatikainen, 2006). Soames (1998) argues that this is particularly problematic when a speaker attributes a belief to someone other than them-self. For example:

<sup>(</sup>b) *a* believes [the actual entity standing in the appropriate causal-historical chain to my use of 'N'] is a *G* 

While causal descriptivism predicts that (a) and (b) should be interchangeable, Soames argues that because the meaning of N is indexed to the person who utters it, (b) entail that a has a belief about N that relies on a causal relation between the speaker and N. Consequently, Soames argues that causal descriptivism results in bizarre belief attributions. But see Nelson (2002) for a defence of causal descriptivism.

for each theory. The relative merit of the Ambiguity Thesis is that it avoids all these common problems.

## 3.12 More theories of meaning

One problem with the way I have presented Studies 5 and 6 is as tests exclusively between Millianism and causal descriptivism. However, we cannot consider the responses gathered as strictly supporting one of these theories as there are other theories that might be compatible with each of the survey responses. That is, (A) responses may not indicate clear support for Millianism and (B) responses may not indicate clear support for causal descriptivism because each response might be consistent with some other theory of meaning.

I think this is particularly true of (B) responses (i.e., "No, they don't mean the same thing"). For one, Devitt's causal chain theory would also predict these responses. Devitt holds that underlying the use of a name is a network of semantically significant "designating chains" which are grounded in an actual object. See the following:

Under each token of a semantic name type (1.4) lies a d[esignation]-chain. These chains are linked together to form the *causal network* for the name type. [Devitt (1981) p. 31]

The name is introduced at a formal or informal dubbing. This dubbing is in the presence of the object that will from then on be the bearer of the name. The event is perceived by the dubber and probably others. To perceive something is to be causally affected by it. As a result of this causal action, a witness to the dubbing, if of suitable linguistic sophistication, will gain an ability to use the name to designate the object. Any use of the name exercising that ability designates the object in virtue of the use's causal link to the object: perception of the object prompted the thoughts which led to the use of the name. In short, those present at the dubbing acquire a semantic ability that is causally grounded in the object....people not at the dubbing acquire the semantic ability from those at the dubbing. [Devitt & Sterelny (1999) p. 66-67]

Thus, according to Devitt's theory the semantic content of a proper name is simply the external designating chain itself (so not reliant on the mental state of the speaker). For this reason two agents who use the same proper name will necessarily mean different things by their respective utterances because they will have different causal chains that led to their

usage. The main problem here is that (B) responses were taken to be indicative of the Frege-Russell Theory. However, Devitt's causal chain theory also predicts (B) responses without also endorsing the Frege-Russell Theory of meaning.

Furthermore it should be noted that while I use causal descriptivism as indicative that the Frege-Russell Theory is correct, this was based purely on the assumption that the Causal-Historical Theory of reference is correct. A philosopher who does not accept this assumption would not take (B) responses to be indicative of the Frege-Russell Theory. Indeed, traditionally most supporters of the Frege-Strawson Theory took something more like famous deeds descriptivism to be true. Famous deeds descriptivism takes the meaning of a proper name to be synonymous with a description where the description describes a historical deed; for example, "Aristotle" would mean something like 'the teacher of Alexander the Great'. From this theory of meaning a Descriptivist theory of reference comes too as it is in virtue of this description that a unique individual who possesses the salient set of properties is picked out. In Studies 5 and 6 we should expect a famous deeds descriptivist to also give (B) responses because the descriptive content the agents would assign to each utterance of a proper name is different. For instance, in Frege's Puzzle we might expect Argos to mean 'the first celestial body visible in the evening' with his use of "Hesperus", and Adonia to mean 'the last celestial body visible in the morning' by her use of "Phosphorus". Moreover, in the Cat Case we should expect John and Jane to assign different descriptions to Kitty even though they both use the name "Kitty".<sup>80</sup>

To this end it might be better to conceptualise Studies 5 and 6 as testing between two groups of theories; those that predict (A) responses and those that predict (B) responses. This does not pose a problem for the Ambiguity Thesis in general; however, it does mean that we cannot say for certain that, if the Ambiguity Thesis is correct, then the ambiguity of proper names is between a causal descriptivist interpretation and a Millian interpretation specifically. Future research interested in the Semantic Question could think about formulating a test to differentiate more precisely between different theories of meaning. For example, it might be possible to tease apart causal descriptivism from famous deeds

<sup>&</sup>lt;sup>80</sup> Although this is less clear in the Cat case as it is implied that both John and Jane see the cat as possessed of a certain distinctive appearance, in which case they could assign the very same description to "Kitty" and thus a famous deeds descriptivist might respond (A) to this case.

descriptivism by constructing a case in which two agents assigned the same descriptive content to a proper name, but came to the description in different ways. In this instance famous deeds descriptivism would predict that the two agents mean the same thing, whereas causal descriptivism would predict that the two agents mean something different.

## 3.13 The semantic pragmatic distinction

I have argued that proper names are semantically ambiguous. As noted in Chapter 1, Section 1.2, one objection to posited ambiguities in philosophy is that the ambiguity can be otherwise explained by appealing to pragmatic considerations. That is, it is possible to explain away apparent ambiguities by a univocal theory by distinguishing between semantic meaning and a pragmatic interpretation. In the last chapter I specifically tried to rule out this objection with regard to definite descriptions by arguing that sentences containing definite descriptions are semantically ambiguous. However, could the same objection be made to Studies 5 and 6? Machery et al.'s (2004) experiment about the reference of proper names faced at least two different objections that hinge on the pragmatic semantic distinction (see Section 3.2).<sup>81</sup> Each of these objections held that the variation they uncovered in the intuitions of ordinary language users might be due to some participants giving a semantic response and other giving a pragmatic response. For example, it was suggested that some participants might give a response based on who the agent (John) literally refers to and some participants might give a response based on who John intended to refer to. Follow up experiments ruled out this explanation of Machery et al.'s data (e.g., Machery, Sytsma, and Deutsch, 2015). It might be that follow up experiments would also be required to rule out this kind of objection to Studies 5 and 6.

Studies 5 and 6 were considered to rule out the possibility of a pragmatic reading (or reading for speaker's meaning) with the way the question was posed. The question asks whether two agents "mean the same thing" by their utterances of a proper name. The choice of the term 'mean' was an attempt to get at semantic content, that is, what the proper name actually means. On way to object to Studies 5 and 6 might be to point out that

<sup>&</sup>lt;sup>81</sup> Note that Machery et al. (2004) are aware that they do not rule pragmatic explanations.

'mean' is commonly used as synonymous with 'intend'. Thus, the problem with the present questions are that it is not clear whether participants were responding to the question of whether the agents actually meant the same thing by their utterances, or whether they were responding to the question of whether the agents intended to say the same thing by their utterances. I think it is unlikely that participants would be responding to the latter question because there is no indication that the two agents are interacting. It might, therefore, be considered unusual to assume that the agents were intending to say the same thing. However, future experiments considering the same question might nonetheless benefit from more explicitly ruling out this possibility. For example, the answer choices might be revised as:

- (A) Yes, they do actually mean the same thing (regardless of whether they think they do).
- (B) No, they don't actually mean the same thing (regardless of whether they think they do).

Alternatively, the question could be posed to emphasise the semantic meaning. For example, "do they *actually* mean the same thing by their utterances regardless of whether they would consider themselves to?" In the next chapter I will present an experiment that offers a greater range of responses so that any possible ambiguity in the interpretation of the participants' responses is ruled out.

## 3.14 Conclusion

The main objective Studies 5 and 6 has been to motivate the Ambiguity Thesis as a way to explain the variation in intuitions about the meaning of proper names. By arguing that proper names are ambiguous and have multiple possible semantic contents, the Ambiguity Thesis provides a way to answer the Semantic Question and maintain that intuitions, at least in this instance, are informative. While there are many benefits of this explanation compared to traditional theories of meaning about proper names, it is also beneficial as it begins to link various terms and phrases; that is, it extends the explanation of definite

descriptions to proper names thus strengthening the explanatory power of the Ambiguity Thesis. In the next chapter I will consider whether the Ambiguity Thesis should also be applied to natural kind terms.

# Chapter 4: An Experimental Perspective on Semantic Externalism about Natural Kind Terms

# 4.1 Introduction

In this chapter I will investigate natural kind terms. Natural kind terms denote sets of things that structure the natural world (as opposed to artificial kinds which reflect the actions of humans);<sup>82</sup> these terms include count terms like 'tigers', mass terms like 'water' or 'gold', and natural phenomena like 'heat' (Kripke, 1980).<sup>83</sup> Natural kinds are of interest to philosophers of language because of problems that are raised regarding the semantics of natural kind terms (Bird & Tobin, 2015). Consequently, philosophers of language have been concerned with the following question: what is the content of a natural kind term?

Philosophy of language takes an interest in natural kinds because basic issues are raised by the semantics of natural kind terms. For example, if we think of naming an entity and describing it as, semantically speaking, fundamentally different ways of talking about it, should we think of natural kind terms as functioning like names or like descriptions?

The philosophical literature can be divided into two categories depending on which of the following two theses are defended:

**Externalism:** the semantic content of a natural kind term is in part determined (or constituted) by the actual state of the world. That is, by factors external to the speaker (see, for example, Kripke, 1980; Putnam, 1975).

**Internalism:** the semantic content of a natural kind term is solely determined (or constituted) by the mental state of the speaker. That is, by factors internal to the speaker (see, for example, Besson, 2012; Segal, 2000).

Accordingly, a striking difference between these two theses is that Externalism entails that two people with *different intrinsic mental* states could nonetheless mean the *same* thing

<sup>&</sup>lt;sup>82</sup> It is the sciences that are often considered to reveal natural kinds (see Bird & Tobin, 2015).

<sup>&</sup>lt;sup>83</sup> But see Gray (2006) for an argument against the treatment of natural phenomena as a natural kind.

with their utterance of a natural kind term whereas Internalism entails that they would necessarily mean *different* things. Likewise, according to Externalism, two people with the same intrinsic mental state could mean different things whereas according to Internalism they would mean the same thing (see Jyllkä, Railo & Haukioja, 2009). As we can see, the debate between these two theses looks remarkably similar to the debate between Millanism and the Frege-Russell Theory about proper names. For this reason it looks like a good place to see if the Ambiguity Thesis extends to natural kind terms as well as proper names and definite descriptions. Indeed, Kripke (1980) saw a connection between the reference of proper names and natural kind terms. When he presented his causal-historical theory it was firstly proposed as a theory about the reference of proper names, however, he went on to argue that also provided an accurate account of the reference of natural kind terms.<sup>84</sup> As we will see, Kripke has a number of thought experiments to show that the causal-historical theory of reference captures our intuitions about natural kind terms that mirror his thought experiments about proper names.<sup>85</sup> It is worth noting that while many philosophers were concerned with natural kind terms, it was not always the case that they saw a connection between proper names and natural kind terms and treated the two types of terms very differently. As we will see in the next section, Mill, for one, gives markedly different analyses for the two language terms.

With this in mind the aim of the present chapter is to conduct a similar study to the study about proper names in the last chapter, this time about natural kind terms. I will present two studies that test the semantic intuitions of ordinary language users, from New Zealand, China, and the US, and argue that the experimental results indicate that the Ambiguity Thesis provides the best account of the meaning of natural kind terms.

The rest of the chapter will be set out as follows. In Section 4.2 I will re-introduce the debate at the meta-level and motivate Meta-Internalism as a way of settling the debate between Internalism and Externalism. Meta-Internalism gives us good reason to take the intuitions of ordinary language users seriously and so motivates experimental work in this area. In Section 4.3 I will discuss the traditional and more recent literature that looks at this debate

<sup>&</sup>lt;sup>84</sup> While Kripke (1980) did not consider himself to be proposing a theory, the way that I have set out the Causal-historical theory (Chapter 3) is as a theory of reference.

<sup>&</sup>lt;sup>85</sup> Although, as we saw in the last chapter, experimental philosophy has shown these intuitions, at least about proper names, are not as universal as Kripke assumed.

to show that there is no consensus among philosophers and, moreover, that it is unclear what we should expect the intuitions of ordinary language users to be. In Section 4.4 I will introduce the possible phenomenon of philosophical bullying and filtering. This is a phenomenon I have not discussed in the previous chapters but one which I think requires attention in this area. In Section 4.5 I will discuss some of the relevant empirical work on natural kind terms. I will then introduce Study 7 in Section 4.6. Section 4.7 and Section 4.8 will contain the experimental method and results; followed by a follow-up study, Study 8, in Section 4.9. Section 4.10 will contain a discussion of the results in which I will argue that they provide evidence for the Ambiguity Thesis. Section 4.11 will discuss whether the results support a philosophical bullying phenomenon. Section 4.12 will discuss and reject some possible alternative explanations. Finally, Section 4.13 will contain a brief conclusion.

#### 4.2 Meta Debate

To settle the debate between Internalism and Externalism a further question arises regarding *how* the debate should be resolved, that is, what kind of evidence is that should be considered relevant. Recently, Cohnitz and Haukioja (2013) have tried to address this question by taking the distinction between Internalism and Externalism to the meta-level. They introduce following two possibilities for determining whether Internalism or Externalism is the correct theory:

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

Meta-Externalism: How a linguistic expression E in an utterance U by a speaker S refers and which theory of reference is true of E is not determined by the individual psychological states of S at the time of U [Cohnitz & Haukioja (2013) p. 482].

According to Meta-Internalism, the debate is settled by appealing to the psychological states of a speaker; for example, the intuitions or mental dispositions one has about their utterance. It is these psychological states that determine whether Internalism or Externalism is correct. Thus, if a speaker regards the meaning of their utterance to be partly determined by the actual state of the world (and possibly by facts unbeknownst to themselves) then, according to Meta-Internalism, Externalism would be correct. If a speaker regards the meaning of their utterance to be determined only by their own mental state, then Internalism would be correct. Meta-Externalism, in contrast, holds that the debate between Externalism and Internalism is not settled by appealing to language users but that it is also determined externally.

Note that as we can see in these characterisations of Meta-Internalism and Meta-Externalism, Cohnitz and Haukioja (2013) are primarily concerned with the *reference* of natural kind terms rather than the meaning *per se*. However, the same distinction applies to the content of natural kind terms; the meaning and the reference seem to be very closely linked. I have noted that while the meaning of a term in a language need not be the same as its reference or extension there is, nonetheless, an important connection between the two. Here I will presuppose an intensional semantics which makes explicit this connection.

**Intensional Semantics:** The <u>meaning</u> of a referential or denoting term determines its extension in all possible worlds.

A consequence of this assumption is the following:

**Strong Dependence Thesis:** For any two referential or denoting terms A and B, A and B have the <u>same meaning</u> *only* if A and B have the same extension in all possible worlds; equivalently, A and B have <u>different meanings</u> *if* A and B have a different extension in some possible worlds.

Note that an intensional semantics of this kind is compatible with a *hyper*intensional semantics, according to which (for example) terms like 'trilateral' and 'triangle' differ in meaning despite having exactly the same extension in all possible worlds. For nothing in the Intensional Semantics assumption (or the Strong Dependence Thesis) rules out the possibility that two terms with different meanings can have that same extension across worlds, so long as each meaning determines that same extension.

In uncovering the *meaning* of natural kind terms, that is, whether Internalism or Externalism is the correct theory, it is necessary to decide which *way* of settling the debate is correct. I

will adopt a Meta-Internalist approach to settle the debate. Cohnitz and Haukioja (2013) support this approach as they argue in favour of Meta-Internalism (about reference). Firstly, they argue that Meta-Externalism is problematic. To do so they consider a world in which Meta-Externalism is true. The speaker's in this world speak Frenglish, that is, Fregean-English about proper names (i.e., the language is specified to be Descriptivist). So the speaker's consider their use of a proper name to refer to the entity that fits the description associated to that name. Thus, they would respond to Kripke type thought experiments (e.g., the Gödel case) accordingly. However, it is also the case that in this world the Causalhistorical Theory is actually true according to the external facts about reference (i.e., according to Meta-Externalism). So in this world the true theory of reference has nothing to do with the way that speaker's use proper names. Therefore, Cohnitz and Haukioja argue that in this scenario it is irrelevant that Meta-Externalism is true. For our intents we can adapt this example to be explicitly about meaning. Again we can consider that Meta-Externalism is true of some world and that in this world the speakers speak according to the Frege-Russell Theory. Thus, they would respond to my proper names cases, Frege's Case and the Cat Case (from Chapter 3) according to this theory (i.e., that the agents do not mean the same thing by their utterances). However, according to the external facts of the world (i.e., according to Meta-Externalism) unbeknown to anyone in this world Millianism is actually the correct theory of meaning. In this instance, again, we see that it is irrelevant that Meta-Externalism is true; it has no impact on the way the speakers use proper names.

Cohnitz and Haukioja instead consider a Meta-Internalist view in which the correct theory of reference is determined by the speaker's dispositions towards the application of the relevant expressions. So if Externalism is correct it is true in virtue of ordinary language users having the disposition to put the burden of reference determination on external factors. That is, the disposition to consider the correct application of a term in actual and counterfactual scenarios to be based on an *external facts*, regardless of whether they know what these facts are. In contrast, if Internalism is true it is only in virtue of ordinary language users having the disposition to put the burden of reference determination on facts internal to themselves.

In my present investigation of Internalism and Externalism, adopting Meta-Internalism gives a natural motivation for experimental philosophy because it holds that it is the dispositions

(or intuitions) of ordinary language users that are relevant for determining the correct theory. Thus, it is entirely relevant to empirically test the intuitions of ordinary language users. Moreover, by accepting Meta-Internalism we might have good reason to think that the Ambiguity Thesis is correct if there is variation in the intuitions of ordinary language users about the correct application conditions for natural kind terms.

#### 4.3 Intuitions about natural kind terms

There is a long tradition of philosophers of language discussing the nature of natural kind terms. As it very often is in the philosophy of language tradition, this literature relies very heavily on intuitions; either the intuitions of specific philosophers who assumed their intuitions would be shared, or the intuitions gathered in empirical investigation. (In the next section I will discuss the relevant experimental philosophy literature; as we will see in this section, however, there are other empirical facts relevant here). In this section I will first show that there is not a consensus among philosophers about this debate. I will then show that there is mixed evidence about whether ordinary language users are likely to have Externalist or Internalist intuitions.

Starting with philosophers intuitions: it appears that early ideas about the semantic content of natural kind terms tended to align with Internalism (although the term 'Internalism' was not used); many well-known philosophers (for example, Mill, Locke, and Frege) held that members of a natural kind shared some set of properties that were discernible to the speaker in a given instance. For one, Mill, in *A System of Logic* (1843), gives an Internalist account of natural kind terms that differs hugely from his account of proper names (cf. Chapter 3). While Mill considered the semantic content of a proper name to be constituted by the actual referent without conveying any other information (we might think of this as an Externalist account about proper names), he thought that general terms (such as natural kind terms) had connotations which constituted the meaning and determined the extension of the term. According to this view, in order to be included as a member of a natural kind term the actual entities needed to possess the properties that were connoted by the general term.

Locke (1894) too had similar views about general terms (such as natural kind terms). According to Locke, the meaning of these terms is not located in the world but is a concept held by the given speaker, what he called the 'nominal essence'. The nominal essence represents a collection of the observable qualities that the speaker associates with the term (3.3). Although Locke conceded that the nominal essence might be generated by a 'real essence', that is, by some deep constitution shared by members of the kind, he maintained that because we do not have direct access to the real essence, ordinary uses of kind terms are based on the nominal essences (3.3-3.5).

We also see Frege endorse an Internalist theory, as he defends a similar descriptivist account to his semantic theory about proper names; that is, that the semantic content of a natural kind term is captured by a definite description of the form 'the *F*'. According to this account it is via these descriptions that a speaker associates with a natural kind term that determines what actual entities are included as the content of that term. Lastly, we see that Kripke (1980), at least, interprets Kant to be endorsing an Internalist account in the following passage:

Here is what Immanuel Kant says about gold...Kant is introducing the distinction between analytic and synthetic judgements, and he says: 'All analytic judgements depend wholly on the law of contradiction, and are in their nature *a priori* cognitions, whether the concepts that supply them with matter be empirical or not. For the predicate of an affirmative analytic judgement is already contained in the concept of the subject, of which it cannot be denied without contradiction....For this very reason all analytic judgements are *a priori* even when the concepts are empirical, as, for example, "Gold is a yellow metal"; for to know this I require no experience beyond my concept of gold as a yellow metal. It is, in fact, the very concept, and I need only analyze it without looking beyond it.'... 'It is in fact the very concept' sounds as if Kant is saying here that 'gold' just *means* 'yellow metal'. [Kripke (1980) p. 117]

Indeed, Kripke takes Kant to be simply saying that the meaning of a term like 'gold' is captured by an indefinite description such as 'a yellow metal' and so no further experience beyond understanding the term is needed to ascertain that truth of analytic statements like 'gold is a yellow metal'.

These early accounts suggest that Internalism will better capture the semantic intuitions of ordinary language users about natural kind terms. This is because these philosophers had

the intuition that we use natural kind terms as descriptive terms and so the terms refer to anything that is captured by the descriptions, or connotation, that the speaker has in mind when they use the term. However, more recently there has been a tendency for philosophers to espouse Externalist accounts. In a famous challenge to Internalism, both Kripke (1980) and Putnam (1975) have challenged Internalist intuitions by arguing that descriptions about superficial properties do not determine the extension of a natural kind term across all possible worlds. Indeed, Kripke continues his discussion of Kant as follows:

Kant is saying here that 'gold' just *means* 'yellow metal'. If he says that, then it's especially strange, so let's suppose that that is not what he's saying. At least Kant thinks it's a *part* of the concept that gold is to be a yellow metal. He thinks we know this *a priori*, and that we could not possibly discover this to be empirically false. [Kripke (1980) p. 117]

Kant (someone just pointed out to me) gives the example 'gold is a yellow metal', which seems to me an extraordinary one, because it's something I think that can turn out to be false. [Kripke (1980) p. 39]

As we can see, Kripke relies heavily on his own intuitions to claim that the meaning of a natural kind term cannot be constituted (even partly) by a description because that description could turn out to be false. Both Kripke and Putnam go on to give Externalist accounts in which they argue that what counts as a member of a natural kind term relies on the actual state of the world and not the mental states of the speaker; we can see this view expressed in the following passages:

[W]e see that the psychological state of the speaker does not determine the extension (or the "meaning," speaking preanalytically) of the word. [Putnam (1973) p. 703]

We have now seen that the extension of a term is not fixed by a concept that the individual speaker has in his head, and this is true both because extension is, in general, determined *socially*— there is division of linguistic labor as much as of "real" labor— and because extension is, in part, determined indexically. The extension of our terms depends upon the actual nature of the particular things that serve as paradigms, and this actual nature is not, in general, fully known to the speaker. Traditional semantic theory leaves out two contributions to the determination of reference-the contribution of society and the contribution of the real world; a better semantic theory must encompass both. [Putnam (1973) p. 710-711]

According to this view, natural kind terms constitute a kind of thing in which members share

the same deep constitution or essential properties, that is, something like the same biological structure, and do so even when we do not know what the deep or essential properties of a kind are. Consider the following:

We use 'gold' as a term for a certain *kind* of thing. Others have discovered this kind of thing and we have heard of it. We thus as part of a community of speakers have a certain connection between ourselves and a certain kind of thing. The kind of thing is *thought* to have certain identifying marks. Some of these marks may not really be true of gold. We might discover that we are wrong about them. Further, there might be a substance which has all the identifying marks we commonly attributed to gold and used to identify it in the first place, but which is not the same kind of thing, which is not the same substance. We would say of such a thing that though it has all the appearances we initially used to identify gold, it is not gold. Such a thing is, for example, as we well know, iron pyrites or fool's gold. This is not another kind of gold. It's a completely different thing which to the uninitiated person looks just like the substance which we discovered and called gold. We can say this not because we have changed the *meaning* of the term gold, and thrown in some other criteria which distinguished gold from pyrites. It seems to me that that's not true. On the contrary, we *discovered* that certain properties were true of gold in addition to the initial identifying marks by which we identified it. [Kripke (1980) p. 119]

Here Kripke asserts that although we may use certain descriptions to pick out the extension of a natural kind term, those descriptions themselves do not constitute part of the meaning of the term. Instead, the meaning is constituted by the actual members of the kind which share some kind of deep constitution or essential properties that are empirically discoverable.

To motivate this account of meaning, both Kripke and Putnam offer a number of thought experiments designed to elicit Externalist intuitions. Two kinds of arguments that Kripke gives are arguments that involve abnormal members and arguments that involve misperception (see Smith, 2005). Firstly, he shows that, although some entity might not fit any of the descriptions typically associated to a natural kind term, they may still be considered a genuine member of that kind. For example, a three-legged albino tiger would not fit the descriptions we usually associate with 'tigers'. Moreover, appealing to a clusterconcepts version, in which a prospective member need only satisfy a portion of all the possible descriptions associated to a kind, does not help the Internalist in this case as we can conceive of very abnormal members (Smith, 2005). For his next kind of argument Kripke shows that we might be mistaken about the descriptions we typically associate to a kind, but that discovering these misperceptions would not have an effect on the kind itself; it would only amend our beliefs about the kind. For example, if it was discovered that we had been deceived in thinking that gold was yellow and that in actual fact it was blue, Kripke argues that it would not be correct to say that gold does not exist. Instead, he argues that we would simply revise our belief that gold is yellow.

Similarly, Putnam (1973) gives a famous thought experiment designed to show that Externalism better accounts for our intuitions regarding Twin Earth. Twin Earth, we are asked to imagine, is identical to Earth but with one difference: the clear liquid on Twin Earth is not H<sub>2</sub>O as it is on Earth but is made up of a different chemical compound, namely XYZ. The intuition Putnam appeals to is whether an earthling who has travelled to Twin Earth would mean the same thing as a twin-earthling would by their utterance of 'water'. Given that both the earthling and a twin-earthling would have the same mental states by the utterance of 'water', Internalism predicts that they would mean the same thing. However, Putnam argues that intuitively it seems that 'water' would mean something different when used by the earthling compared to a twin-earthling; they just would not know it. He concludes that 'meanings just ain't in the head,' a phrase that has become closely associated with Externalism.

Since Kripke and Putnam's arguments it appears that Externalism has become widely accepted by philosophers.<sup>86</sup> Indeed, Bourget and Chalmers (2013) study of professional philosophers' supports this claim. Specifically, one question they asked was:

#### Mental content: internalism or externalism?

What they found was that 51.1% of philosophers surveyed identified as an Externalist, 20% identified as an Internalist, and 28.9% identified as 'other'. Although the question is not exclusively about natural kind terms, the results reinforce the general impression that philosophers tend to espouse Externalist type accounts. However, as indicated by Bourget and Chalmers' study, as well as by the state of the literature, there is still some support for Internalism among philosophers (e.g., Besson, 2012; Crane, 1991; Segal, 2000; Smith, 2005).

<sup>&</sup>lt;sup>86</sup> Both Jyllka et al. (2009) and Wikforss (2008) make this claim.

Smith (2005), for one, defends what he describes as a neo-Lockean view about natural kind terms in which the reference of a natural kind term is determined by the 'nominal essence' associated with the term, that is, its discernible properties. He does, however, reject individualism, the view that the meaning of a language term depends completely on the nominal essence associated to the term by the individual speaker. Instead, Smith argues in favour of a division of linguistic labour in which the meaning of language term depends on the nominal essence that the relevant experts on the kind in question in a linguistic community associate with the term (p. 70). Moreover, Smith goes on argue that real-world examples show that in the past we have not used language in the way that Externalism predicts. For one, when it was discovered that 'jade' referred to both jadeite and nephrite we continued to call both of them 'jade'. Thus, he claims we have an example in which even after it is discovered that that sample members do not share the same deep constitution, we continue to think the samples are legitimate members of the same kind. Although it is worth noting that Smith does not consider the alternative that this example merely illustrates that 'jade' (unlike 'gold') is not a natural kind term.

Other contemporary philosophers have argued that Externalists have a problem explaining the meaning of empty natural kind terms. For this argument an appeal to our intuitions about Dry Earth has been made (see Boghossian, 1997; Segal, 2000). On Dry Earth there is no liquid similar to the watery stuff on Earth in any way although we can imagine that the tenants of Dry Earth are under the impression that there is. In this case, the problem for advocates of Externalism is to explain what the term 'water' means on Dry Earth. Besson (2012) argues that the only option for the Externalist is to concede that the term 'water' is meaningless on Dry Earth.<sup>87</sup> Internalists, on the other hand, have available the option of arguing that 'water' on Dry Earth means the same thing as it does on Earth (Besson, 2012; Segal, 2000), a more satisfactory response than that available to the Externalist.

What I have tried to illustrate here is that among philosophers there does not seem to be a consensus about what the semantic content of natural kind terms are. Although there does seem to be a general tendency for philosophers to have Externalist intuitions now, as we have seen there are still a faction who continue to think there are good reasons to reject

<sup>&</sup>lt;sup>87</sup> Note Segal (2000) does not think the Externalist can take this route.

these Externalist intuitions and, moreover, appeal to other intuitions that undermine Externalism. Moving on to non-philosophers, or ordinary language users' intuitions: there appear to be mixed reports about what we should expect the intuitions of non-philosophers to be on the matter. A number of psychology experiments seem to support Externalism as they present evidence for the thesis 'psychological essentialism'. Psychological essentialism holds that people believe or behave as if members of a kind (generally a natural kind) have in common some underlying essence even if they do not know what that essence is (Gelman, 2004; Medin, 1989; Medin & Ortony, 1989).<sup>88</sup> Evidence of psychological essentialism comes from a number of studies that show that even young children rely heavily on biological categories rather than perceptual similarities when making inferences about natural kinds (e.g., Keil, 1986; Gelman, 1988; Gelman & Markman, 1986, 1987). For example, four-year olds who are told that fish breathe under water but that dolphins surface for air (accompanied by pictures) are more likely to infer that a shark (categorised as a fish) can breathe under water, even though the shark looks much more similar to the dolphin than to the fish (Gelman & Markman, 1986).<sup>89</sup> Although psychological essentialism does not make any claim about the actual state of the world, it does tell us something about the way people approach the world and therefore what a speaker intends to mean by their utterance of a natural kind term (Medin, 1989). If we take Meta-Internalism seriously then if ordinary language users intend the meaning of their natural kind terms to be determined by the actual state of the world, as psychological essentialism suggests they do, then there is good reason to think that Externalism is true.

In favour of Internalism, on the other hand, there appears to be a general belief held by philosophers that non-philosophers hold Internalist intuitions about natural kind terms. Smith (2005) has remarked that overwhelmingly his students seem not to hold Twin Earth intuitions and in fact seem to think it is obviously the case that Twin Earth XYZ is water (p.81). Frank Jackson too has noted:

In the mouths and from the pens of the folk it is indeterminate whether it is H2O or the watery stuff on Twin Earth that counts as water on Twin Earth. [Jackson (1998a) p. 38 n. 12]

 <sup>&</sup>lt;sup>88</sup> More specifically this characterisation is known as the 'placeholder' version of psychological essentialism.
<sup>89</sup> But see Olivola & Machery (2014) for a discussion indicating that psychological essentialism is not a consistent feature of human cognition cross-culturally.

Putnam even seems to have supported this notion by noting that undergraduate students tend to think that a term refers to whatever passes the operational definition at the time. That is, they tend to think that 'gold' could correctly refer to substances without the atomic number 79 as long as the substance passed the operational definition at the time it was uttered (p. 153). These verbal accounts suggest that Externalism is not as self-evident as some philosophers have claimed it to be (e.g., by Kripke and Putnam, see Seegar, 2010 p. 239) and that pre-theoretical intuitions actually align better with Internalism.

The literature discussed so far does not seem to settle the debate between Internalism and Externalism, especially with regards to the intuitions of ordinary language users. Indeed, there seems to be good reasons to predict that ordinary language users could have either kind of intuition. Thus, the best way to settle these contrary reports is to empirically determine what the intuitions ordinary language users are. As noted earlier, using the intuitions of ordinary language users is highly controversial because many philosophers take the intuitions of philosophers to be superior (see Chapter 1, Section 1.5). I have argued that I think there are good reasons to think this is not the case. With explicit reference to this debate Seegar (2010) has argued that the truth of some intuitions, and specifically our intuitions about Twin Earth, is determined simply by whether most people share the intuition or not. This makes the intuitions of ordinary language users entirely relevant. However, I think there is a further reason to consider the intuitions of ordinary language users to be important when we are considering this particular debate about natural kind terms which I will discuss in the next section.

# 4.4 Philosophical bullying and philosophical filtering

An additional reason that uncovering the intuitions of ordinary language users is important in this debate is because it might reveal an underlying bullying or filtering phenomenon occurring in our practice of philosophy. By philosophical bullying what I mean is that in the course of studying philosophy, students might get certain views pushed on them. So although they might not initially take some theory to be intuitive or rational, over time they might be coerced into thinking that it is, in fact, the best theory. The idea here would be that students did not come to accept the theory for good reasons, that is, because they changed their minds based on good arguments in favour of the theory. Instead, the idea would be that they would actually find the theory to be more intuitive simply because it was presented to them as if it was clearly the better theory and that everyone has intuitions in line with it. If this was the case students might consciously or unconsciously come to think that the theory was intuitive. We can imagine this scenario plays out as follows: a student is given a philosophical problem and then taught about two theories that try to solve the problem. Theory A predicts that we will have x intuition and Theory B predicts that we will have y intuition to a given though experiment. The student has the intuition y when presented with the thought experiment. However, they are told that intuition that is elicited is x and therefore assumes that everyone actually has the intuition x. To avoid being wrong or doing poorly in the class the student decides that Theory A is the correct theory and that they have an incorrect intuition. They then go on to espouse Theory A in assessments and discussions. We see this kind of social conformity observed in psychology experiments about perception. In one famous experiment, Asch (1951) asked a group to each say aloud which of three lines was the longest as part of a vision test. Each group consisted of one participant and seven confederates (who were only pretending to be participants unbeknownst to the actual participant). While it was obvious what the correct answer was to the question, all the confederates gave the same wrong answer. The participant gave their answer last. Asch found that 75% of participants conformed and gave the wrong answer at least once. These kinds of studies suggest that social pressure can change the judgement a person gives, even if they have a different intuition. Indeed, if we can observe this conformity in something as obvious as perception then it seems very likely that the same will be true of philosophical judgements. Philosophy students who conform to this kind of social pressure might then end up actually believing that they have the intuition x. Indeed, when presented with the same thought experiment at a later time they respond immediately in favour of intuition x.

By philosophical filtering, on the other hand, I mean a similar kind of phenomenon whereby the intuitions reported by undergraduate students and the intuitions reported by professional philosophers are at a discord. In this instance, however, the discord would come about by a systematic drop out of students who do not share the intuitions of the
philosophers who teach them. Indeed, it might be that only undergraduates who initially have the 'right' intuition progress on in philosophy. Students who do not share the intuition may assume they must be 'bad' at philosophy and therefore drop out at a higher rate than students who do have the intuition. If this is the case there would be a kind of biased philosophical filtering.

There are a number of places in philosophy in which these kinds of phenomena could be occurring; <sup>90</sup> for example, in Gettier cases about knowledge we are told quite explicitly that they are cases in which the agent does not know that *p*. In Trolley cases we are told that under certain circumstances the correct action is to switch. And in Twin Earth we are told that the liquid on Twin Earth is not water. This case of Twin Earth looks particularly like it could turn out to be an instance of philosophical bullying because it looks like there is a discrepancy between the intuitions of philosophers and non-philosophers. That is, it at least looks like philosophers tend to be Externalists while verbal reports suggest that non-philosophers have Internalist intuitions. Indeed, the verbal reports suggest that when given the Twin Earth thought experiment, undergraduates want to say that the liquid on Twin Earth *is* water.

Philosophical bullying/filtering offers an alternative explanation to the Expertise Objection. Both the Expertise Objection and the philosophical bullying/filtering phenomenon suggest that there will be a difference in the intuitions of philosophers compared to nonphilosophers. The Expertise Objection holds that any discord observed between the intuitions of professional philosophers and non-philosophers is due to the expert status of the philosophers' intuitions. Thus, according to the Expertise Objection, such a discord gives us reason to give preference to philosophers' intuitions. Philosophical bullying/filtering also offers an explanation for the discord between philosophers and non-philosophers. However, in this explanation claims that rather than being more reliable, philosopher intuitions might actually be unduly biased. If this is the case then the argument that reflective thinking aids philosophers' intuitions is challenged as it would more likely be the case that reflective thinking actually serves to reinforce such a bias as by reaffirming any biased beliefs (Kornblith, 2010; Alexander, 2012, p. 96). Non-philosophers, on the other

<sup>&</sup>lt;sup>90</sup> Although I take philosophical bullying and philosophical filtering to be distinct phenomena, I will hereby treat them as a single phenomenon for ease of explanation.

hand, are free of any theoretical affiliations that might bias their judgments and would, according to this explanation, be more reliable (Machery, 2012).

Studies 7 and 8 I will test for a difference between philosophers and non-philosophers intuitions about natural kind terms by comparing non-philosophers intuitions uncovered in these studies to the philosophers' intuitions uncovered by Bourget and Chalmers (2013). If any difference is indicated I will argue that this difference is better accounted for by a philosophical bullying/ filtering phenomenon than by holding that the intuitions of philosophers are more expert. In particular this is because of the global divergence uncovered by Bourget and Chalmers in philosophers' intuitions. That is, they found that many of the philosophers intuitions were influenced by geographical factors, age, specialisation, etc. which suggests that these intuitions are not unbiased.<sup>91</sup> Before outlining the present studies however, I will discuss some of the relevant experimental research that has been conducted in the area.

# 4.5 Experimental Philosophy

As I have noted, within the philosophy of language domain most of the experimental work has focused on the debate between descriptivism and the causal-historical theory of reference for proper names following Machery et al. (2004). In light of this, it has been noted in a number of places that the literature would benefit from examining natural kind terms in the same sort of way:

In addition, our focus in this paper has been on intuitions about proper names, since proper names have been at the center of debates about semantics. However, it will be important to examine whether intuitions about the reference of other sorts of terms, for example natural kind terms (see, e.g., Putnam 1975), also exhibit systematic cross-cultural differences. We hope that future work will begin to address these questions. [Machery et al. (2004) p. 9]

A somewhat different empirical challenge to the externalist thought experiments may be brewing in the new "experimental philosophy" movement. It is possible, as experimental

<sup>&</sup>lt;sup>91</sup> Bourget and Chalmers webpage shows the up-to-date effects. These indicate that with regard to Internalism and Externalism about mental content, geographical area, age, and specialisation all affect philosophers' intuitions.

philosophers appear to have shown regarding intuitions about other prominent philosophical thought experiments, that externalist intuitions are a culturally local product. Weinberg et al. (2001) report that the Gettier intuition tends not to be had by East Asian subjects, and Machery et al. (2004) say that a majority of the East Asian subjects they surveyed had intuitions that favor the descriptivist theory of reference supposedly refuted by Kripke (1972). Surely, there will soon be cross-cultural empirical data on externalist intuitions. If it is found that only Westerners, for example, tend to have externalist intuitions, that could form the basis of a new kind of empirical critique of externalism. [Lau & Deutsch (2014) "Externalism About Mental Content" *The Stanford Encyclopedia of Philosophy*]

Here we see a call for the need of experimental work to test for Externalist intuitions among ordinary language users, and specifically for cross-cultural studies. Notably, however, there are some experiments of relevance here. In this section I will discuss three experimental studies which explicitly consider a hybrid theory and contrast it with univocal theories of meaning, that is, Externalism and Internalism. I will then discuss two further experimental studies which do not discuss a hybrid theory, but whose results are consistent with such a theory. Recall that a Hybrid Theory makes the following claim:

**Hybrid Theory of Meaning:** There are two or more semantic elements associated with a term A, which work together in some way to determine A's referent or extension in any possible world.

Genone & Lombrozo (2012), for one, consider and defend a hybrid theory of natural kind terms (as well as other general terms). They explain the hybrid theory thus: "[it is a theory] that incorporates both descriptive and causal factors as relevant to determining reference" (p. 721). Of course, a hybrid theory need not incorporate these *particular* elements. What is important is that there is not a single 'component of' the meaning of a term that by itself, and in all contexts, determines the reference of that term.

In their study Genone and Lombrozo presented participants with vignettes about one of four concepts (diseases, minerals, artefacts, legal documents) and asked whether two agents referred to the same thing with the use of a name. Crucially, the two agents' concepts of the name differed either in descriptive information or by the source of the concept (that is, the causal-chain from which they acquired the term). To simulate shared or unshared causal-chains the agents were described to be from either the same island or separate islands. To simulate a difference in descriptive information the number of beliefs the two agents shared about a concept was altered. What they found was that when agents shared a causal-chain and descriptive beliefs then participants tended to give 'same reference' responses and when both causal origin and descriptive beliefs differed between agents then participants tended to give "different reference" responses. However, when agents a shared causal-chain but had different descriptive beliefs or when agents had a different causal origin but shared descriptive beliefs then "same reference" and "different reference" responses were mixed at roughly 50%. It is this finding that Genone and Lombrozo considered as evidence for a hybrid theory. This conceptualization of a hybrid theory, though, is underspecified. Genone and Lombrozo (2012) recognize as much, suggesting that "there are many ways in which a hybrid theory of reference might be formulated, and in particular different ways in which causal relations and descriptive information associated with a referring term might work together or independently in order to pick out a referent" (p. 721). Genone and Lombrozo do not worry too much about how the hybrid theory should best be explicated, though, claiming that experimental results support the idea of a hybrid theory, no matter how it is unpacked. It is my contention though that this is not the case. Again recall that a hybrid theory is compatible with the following three views:

- 1. **Ambiguity View.** If a term A is ambiguous, then there are straightforwardly at least two different meanings associated with a word in a language. On any occasion of utterance, A will have just one of those meanings associated with it, and which one will be determined by the context.
- 2. **Pluralist View.** On the pluralist view, a term A has multiple meanings associated with it on any occasion of utterance. The term is not (or at least need not be) ambiguous on this conception because the single term can consistently have the same contents associated with it on every occasion utterance.
- 3. **Component View.** Unlike on the ambiguity and pluralist views, according to the mixed view a term A might have a *single* meaning or semantic value associated with it, but that different components work together to determine what that meaning is.

While Genone and Lombrozo leave open what the best characterisation of the hybrid theory is, I will argue is that not all of these views are supported. Specifically, I will claim that only the ambiguity view is supported by the present results.

There are two further studies which explicitly consider a hybrid theory of natural kind terms. Braisby, Franks and Hampton (1996) in an early study looked at whether ordinary language users have essentialist beliefs about natural kinds; that is, the beliefs that members of a natural kind share essential properties and do so regardless of any beliefs about which properties determine reference. Consequently, Essentialism is something akin to Externalism.<sup>92</sup> To find evidence of these essentialist beliefs, they presented participants with scenarios based on seven natural kinds (cat, water, tiger, gold, bronze, lemon, and oak). Each kind was presented in three scenarios: one in which something is discovered about a non-essential property (cats do not mew), one in which something is discovered about an essential property of a previously considered member of the kind (Tibby the cat-like creature turns out to be a robot), and one in which something is discovered about an essential property of all category members (all cats turn out to be robots). Following each scenario, participants were asked whether they agreed or disagreed with the following kinds of statements:

Existential (+): Cats do exist

Existential (-): Cats do not exist

*Qualified* (+): Cats do exist, and people's beliefs concerning cats have changed *Qualified* (-): There are no such things as cats, only robots controlled from Mars *Membership* (+): Tibby is a cat, though we were wrong about her being a mammal *Membership* (-): Tibby is not a cat, though she is a robot controlled from Mars

Braisby et al. found that participants generally made essentialist responses (i.e., responses predicted by Externalism) but did so more often in the non-essential property scenarios.<sup>93</sup>

<sup>&</sup>lt;sup>92</sup> Braisby et al. (1996) state that essentialism is the view held by Kripke (1980) and Putnam (1975).

<sup>&</sup>lt;sup>93</sup> What counted as an Externalist response depended on the scenario presented. For example, in the scenario that something is discovered about a non-essential property (cats do not mew), the Externalist responses were endorsement of Existential (+), Qualified (+), and Membership (+). In the scenario that something is discovered about an essential property of one previously considered member of the kind (Tibby the cat-like creature turns out to be a robot), the Externalist responses were endorsement of Existential (+) and Membership (-). In the scenario that something is discovered about all category members (all cats turn out to be robots), the Externalist responses were endorsement of Existential (+), and Membership (-). In the scenario that something is discovered about all category members (all cats turn out to be robots), the Externalist responses were endorsement of (+), and Membership (+) indicates an

Responses in the two essential property scenarios widely varied (46-95% agreement with Externalism across two experiments) depending on the scenario and the type of question. Furthermore, a number of contradictions were apparent within individuals; for example, in the same scenario Tibby was considered both a cat and a non-cat by some. What Braisby et al. conclude is that the meaning of natural kind terms is sensitive to context and does not strictly follow an essentialist pattern as Putnam and Kripke have claimed. Instead, they argue for a representation change theory of natural kind terms. According to this theory, ordinary language users make use of at least two independent contents of natural kind terms depending on the classification perspective used. That is, from one perspective the content of a term is determined by the properties initially associated to the term, and from another perspective the content is derived from the newly discovered actual essence (p. 269). While their results appear to support a hybrid theory, it is again left underspecified as to the exact characterisation. Braisby et al.'s representation change theory does not distinguish between the different versions of the hybrid theory. I will argue that there data is best understood as initial evidence that natural kind terms are ambiguous.

While Braisby et al. argue for a hybrid theory (and therefore give some indication that natural kind terms might be ambiguous), Jyllkä, Railo and Haukioja (2009) have more recently objected to this experiment and argued that the results can be explained on a strict Externalist view. Jyllkä et al. claimed that Braisby et al.'s scenarios and questions were ambiguous and so the disagreements found could be based on the way the questions were interpreted. For example, the biggest disagreements Braisby et al. found towards Externalism were in the essential individual scenarios in which a discovery is made about an essential property of an individual (e.g., Tibby the cat-like creature is discovered to be a robot) in which Externalism predicts that sentences such as "Tibby is a cat, although we were wrong about her being a mammal" should be considered false. However, Jyllkä et al. argue that the formulation of the scenario is faulty as it opens by stating that Tibby *is* a cat. This formulation may bias the reader to interpret "cat" in such a way that "Tibby is a cat" is true. In turn, Jyllkä et al. conducted probably the most relevant experiment to the present studies by testing the intuitions ordinary language users hold about natural kind terms.

Externalist response. Braisby et al. set out the predictions of Externalism for each of the six statements in each of the three scenarios in their Table 1 (p. 257).

Along with Externalism and Internalism, they also introduced a hybrid account which holds that natural kind terms can be used with either an 'externalistic sense' or an 'epistemic sense' in which the epistemic sense determines membership via the observable properties a speaker associates with a term (i.e., as Internalism holds). To illustrate, Jyllkä et al.'s hybrid account holds that in a Twin Earth type scenario an earthling on Twin Earth that utters "there is water" would be both correct (in the epistemic sense) and incorrect (in the externalistic sense). Overall the hybrid account makes the following claims:

1. Speakers believe that samples falling under natural kind concept *C* share some hidden, empirically discoverable, essence *E*.

2. The speakers hold that, in one (externalistic) sense, C applies to whatever possesses E.

3. The speakers hold that, in another (epistemic) sense, *C* applies to whatever fits the identificatory knowledge associated with *C* [Jyllkä et al. (2009) p. 42]

Jyllkä et al. presented Finnish participants with six scenarios about different fictitious natural kinds; for example:<sup>94</sup>

A yellowish, bitter-smelling, fragile mineral called zircaum occurs widely in Mid-Siberian soil. Scientists generally believe that zircaum is the compound ACB. In Northern Norway, a deposit of a substance just like zircaum is found—it is yellowish, bitter-smelling, fragile, etc. When scientists examine its deep structure, they conclude that it is ACB, just like zircaum is believed to be.

A few weeks after the discovery in Northern Norway, scientists examine the Mid-Siberian substance more closely. Using methods and instruments more exact than previously available, they find out that they were wrong about the deep structure of the substance: the substance is KML instead of ACB. However, the substance found in Northern Norway was indeed ACB, just as the scientists thought it was. [Jyllkä et al. (2009) p. 49]

The clarifying statement "in other words, it turns out that the substance found in Norway does not share molecular structure with the substance found in Siberia" was also added in

<sup>&</sup>lt;sup>94</sup> In one experiment Jyllkä et al. included positive and negative formulations of the scenarios in first and third person perspectives. In the positive scenarios a substance that is thought not to belong to a natural kind turns out to actually be of the same structure; in the negative scenarios a substance that is thought to belong to a natural kind turns out to have a different structure (e.g., the example given about zircaum above is an example of the negative scenario). However, no differences were found between the different formulations or perspectives of the scenarios.

the second experiment to aid comprehension. Participants were then asked whether the earlier judgement (in this case that the substance found in Norway is zircaum) was correct. Four answer choices were provided: 'yes', 'no', 'on the one hand yes, on the other no', and 'cannot say'. 'Yes' responses were considered indicative of Internalist intuitions (Internalists would consider the earlier judgements true) and accounted for 22% of the final responses. 'No' responses were considered indicative of Externalism (Externalists would consider the earlier judgements false) and accounted for 48% of the final responses.<sup>95</sup> 'On the one hand yes, on the other no' responses were considered indicative of their hybrid account because the hybrid account holds that are two ways to consider such judgements; these made up 17% of the responses.<sup>96</sup> While the responses were very varied, Jyllkä et al. considered the results to support Externalism. This is because the number of Externalist responses was significantly higher than of Internalist and hybrid account responses. Since participants were actively given the opportunity to give hybrid account responses, Jyllkä et al. argued that the variation in responses cannot be taken to support a hybrid account. Instead, in order to explain the high variation in responses Jyllkä et al. considered that the intuitions of ordinary language users may not be precise in this domain as they did not have a chance to reflect on their responses and so we should not expect clear intuitions. Nevertheless, they maintained that the higher endorsement of Externalism is evidence that ordinary language users have Externalist predispositions.

The problem with Jyllkä et al.'s (2009) study is that their characterisation of the hybrid theory should not be taken as evidence against all views consistent with hybrid theories. Without distinguishing between them, a pure externalist theory can seem more attractive. However, once they are properly distinguished we can see that an ambiguity theory is to be preferred. Indeed, their study does not rule out the ambiguity view, and therefore does not rule out the Ambiguity Thesis. The hybrid theory they have in mind, and they find evidence against, looks like it is the pluralist view.<sup>97</sup> This is because their hybrid theory states that there are two senses associated with a natural kind term, both of which are expressed with a given utterance of that term. According to this theory a

<sup>&</sup>lt;sup>95</sup> Note "yes" and "no" responses indicated the opposite theories in the positive scenarios.

<sup>&</sup>lt;sup>96</sup> 12% of participants responded 'cannot say'.

<sup>&</sup>lt;sup>97</sup> Although it is worth noting that in Section 4.12 I will give some indication from Jyllkä et al. that their hybrid theory is best considered as a component theory. This exemplifies the confusion among philosophers regarding hybrid theories.

speaker holds that a natural kind term has two senses or semantic values associated with it, namely whatever possesses as essential, empirically discoverable property, and whatever fits the descriptive content associated with the term. If this theory was characterised as different components working together to determine a single meaning of a natural kind term, this would look more like a component view. This hybrid theory as they conceive of it is also not compatible with an ambiguity view, because someone who thought that a natural kind term was ambiguous would presumably not give the hybrid theory response. It is more natural to give just one disambiguation. Sometimes the disambiguation will vary between participants. But even that need not be the case. To illustrate, consider the following example:

(S1) 'Bank' means the side of a river.

If asked whether S1 is correct, presumably most participants would respond 'Yes' rather than 'On the one hand yes, on the other no', because the sentence in which the term is embedded, combined with a background charitable disposition to interpret sentences as true, force one disambiguation. Given that 'bank' is a paradigmatic ambiguous term, we can see that this kind of test does not give us a reason against an ambiguity theory. Indeed, if an ambiguity theory were true, we would not expect there to be a great proportion of 'on the one hand yes, on the other no' responses because ambiguity theories hold that in a specific instance the utterance of an ambiguous term will only have one semantic content, it is just ambiguous between multiple contents which one it is. Accordingly, we would expect to see a mix of Externalist and Internalist responses. This is exactly what Jyllkä et al. do find. Furthermore, an extremely high proportion of participants gave mixed responses (81%), that is, did not respond consistently to different scenarios. This is again variation predicted by an ambiguity thesis. While Jyllkä et al. did not consider this variation to be indicative of the hybrid theory, because participants could explicitly give a hybrid response; I contend that their results do look like evidence of the Ambiguity Thesis.

Furthermore, I think there was a methodological problem in the simplistic answers choices available ('yes', 'no', 'on the one hand yes, on the other no'). This problem is not unlike a possible objection that faced Studies 5 and 6 discussed in Chapter 3, Section 3.12. The problem is these responses may not have adequately represented the actual intuitions that

participants held. For instance, participants who made 'no' responses could in fact do so for completely Internalist reasons if what they meant was that a natural kind term means the same thing that experts mean by the term. That is, when we use a term 'gold' what we mean is *the same thing that scientists, i.e. metallurgists, mean by 'gold' at the time of the utterance*. So Internalists could consider the earlier judgements incorrect if scientists at the time would have considered the judgment incorrect. Similarly, participants who made 'yes' responses could do so for Externalist reasons if what they meant was that the natural kind term in the scenario had a different meaning than the meaning we would have if we used the term. A key feature in the present studies will be to provide answer choices that cannot be interpreted in multiple ways and thus provide clear support for either Internalism or Externalism.

Some recent experiments by Buckwalter and Stich (2010) and, as a replication, Adleberg, Thompson and Nahmias (Forthcoming) give some evidence towards Externalism, although these studies too do not distinguish between the different hybrid theories. In these studies the authors were interested solely in any gender differences in philosophical intuitions. One vignette in each study described the planet Twin Earth, which differs to Earth only in that the watery stuff is XYZ rather than H<sub>2</sub>O. Participants in each study (undergraduate students) were then asked whether Oscar from Earth and Twin-Oscar from Twin Earth mean the same thing by their utterance of 'water'. Participants responded on a scale from 1 ('they mean different things') to 7 ('they mean the same thing'). What they found was that participants tended slightly toward Externalism. Buckwalter and Stich found that the mean response for males and females respectively were 5.63 and 4.49, and Adleberg et al. found the mean responses 5.04 and 5.31 (keeping in mind that 4 is the midpoint between an Internalist and Externalist response). While these studies give some evidence about the intuitions of ordinary language users, just by reporting the means they do not give us an indication about whether there is evidence of an ambiguity. Evidence for the Ambiguity Thesis would be indicated by participants giving both very high (Externalist) and very low (Internalist) responses (which together give us a mid-point means). However, it might be that all of the participants gave mid-point responses which would be evidence against an ambiguity and evidence for a difference hybrid theory. Thus, the results presented by these experiments do not tell us how we should interpret the mixed results, although they do give evidence for

some hybrid theory.

# 4.6 Present studies: Studies 7 and 8

The aim of Studies 7 and 8 is to extend the data collected so far in Chapter 2 and Chapter 3 with experiments to empirically uncover the intuitions of ordinary language users about the content of natural kind terms. However, this study will also aim to answer some further questions that have not yet been considered. All together there are a number of questions under investigation in this chapter:

- 1. Is there evidence for the Ambiguity Thesis?
- 2. Is there evidence of a philosophical bullying and/or filtering phenomenon?
- 3. Are there cross-cultural differences?
- 4. Does reflectiveness affect intuitions?
- 5. Are there differences between considering Earth and Twin-Earth cases?

Firstly, the most important question under consideration is whether the Ambiguity Thesis extends to natural kind terms. It is predicted that it will, due the connection often noted between natural kind terms and proper names, as well as from what is suggested by the previous experimental work (e.g., Jyllkä et al., 2009). Evidence in favour of the Ambiguity Thesis will come from variation in the responses between participants. This variation may occur at many different levels: between any groups of participants [based on culture, gender, reflectiveness priming, the type of cases received (I will talk about this variable shortly)], within groups, or within individual participants across different cases.

A secondary question I will look at is whether there is evidence of a philosophical bullying/filtering phenomenon. This question has not been investigated in the previous two chapters. Both this phenomenon and the Expertise Objection would be borne out by the same results; namely, a difference between the intuitions of philosopher and nonphilosophers. Specifically, in this study, we would expect the difference to be borne out by participants overwhelmingly reporting Internalist intuitions (as verbal reports suggest they will) in contrast to professional philosophers who tend to espouse Externalism. However, given that Bourget and Chalmers found that philosophers intuitions are influenced by extraneous factors, such geographical area, there will be initial reason to suspect that philosophers are bullied into Externalism (or filtered out based on Externalist intuitions) through the course of studying philosophy. Indeed, if it is the case that there is not global divergence among non-philosophers, that is, if there is not cross-cultural variation among non-philosophers, but instead a global convergence toward Internalism, this will be evidence of a bullying/filtering effect over the Expertise Objection. Moreover, if it turns out that reflective thinking does not make a difference to philosophical intuitions (as we saw in Chapter 3 for proper names at least) this would also be evidence against the Expertise Objection's explanation. Not only would such a finding be reason to rethink the way philosophy is presently taught, it would also justify more experimental work to determine what our pre-theoretical intuitions are in other philosophical domains and would therefore be a good reason to reject the Expertise Objection.

Study 7 will investigate cross-cultural differences and whether reflectiveness affects our intuitions similarly to Chapter 3. Firstly, Studies 7 and 8 will include New Zealand, Chinese, and US participants to investigate any cross-cultural differences in intuitions. It is expected that there will be similar variation between the cultural groups as there was in the proper names studies (Studies 5and 6). In this case this variation would likely be borne out by a greater proportion of Externalist responses among New Zealand and US participants than Chinese participants. Reflectiveness will be tested again using the Cognitive Reflexivity Test (CRT). Participants will receive one of three versions of the CRT before they are given the natural kinds cases. This will determine both whether being more reflective (i.e., scoring higher on the CRT) is related to intuitions regarding natural kind terms, and whether priming for reflective thinking (through the three different versions of the CRT) is related to intuitions is not specific to the terms.

Finally a new question will be introduced here that I have not yet discussed (i.e., Question 5 from the list above). This question will investigate whether ordinary language users have different intuitions depending on whether they are considering actual cases of natural kinds. Participants will either receive vignettes about the actual Earth or they will receive vignettes

about a different planet, namely Twin Earth. Previous research in cognitive psychology suggests that people reason differently when presented with an abstract problem compared to problems that they are familiar with in real life (e.g., Johnson-Laird, Legrenzi & Legrenzi, 1972; Kaiser, Jonides & Alexander, 1986). Furthermore, experimental research in other areas of philosophy, such as metaphysics, epistemology, and ethics, have already found some evidence that people reason differently in abstract versus concrete scenarios (Carlsmith, Darley & Robertson, 2002; Roskies & Nichols, 2008; see Sinnott-Armstrong, 2008, for an overview). In particular, Roskies and Nichols (2008) found that participants held different intuitions depending on whether they were asked to reason about the actual world or about an alternate world. When told that the world had been discovered to be deterministic, participants in the actual condition were more likely to give compatibilist responses to questions about the moral responsibility, free choice, and blame worthiness of an agent compared to participants in the alternate world condition. These findings suggest that participants might have different intuitions about the semantic content of a natural kind term depending on whether they are thinking about the actual world or about the alternate world Twin Earth.

In Study 7 and 8 participants will be presented with Kripke and Putnam type scenarios in a similar manner to Jyllkä et al.'s (2009) experiment. In these scenarios an agent will make some statement about a natural kind. Participants will be asked whether their statement was literally true. This way of phrasing the question will avoid the problem faced in the previous chapter (about proper names) in which it was not clear whether participants were responding to the literal or the intended meaning of an agent's statement. Following this question participants' will also be required to give the reason(s) for their response. There will be four reasons provided: two that correspond with their 'yes' responses (i.e., the agents' statement was false). One reason for each answer will be an Externalist reason and one reason for each answer will be an Internalist answer. Moreover, participants will be given the opportunity to provide their own reason. This will resolve any possible ambiguity between Internalism and Externalism that comes from simple yes/no responses, such as in Jyllkä et al.'s study.

#### 4.7 Study 7: Experimental Method

*Participants*: 393 undergraduate students from Victoria University of Wellington, New Zealand and 113 undergraduate students from Nanjing, China (35 from Nanjing University, 67 from Nanjing Normal University, 11 from Nanjing University of Science and Technology) participated voluntarily. Of the New Zealand participants, 193 were approached in a firstyear philosophy course during normal lecture time (the other 200 were approached in a first-year political science lecture) and only 44 had taken any philosophy courses previously. 295 participants in the New Zealand sample identified as being of NZ-European descent, 39 as Maori, 22 as Pacific Islander, and 25 as Asian (some participants identified with more than one ethnic group). Gender was only recorded for the New Zealand sample with 195 males and 194 females. The probes were presented in English to the New Zealand participants and Chinese to the Chinese participants. The original English versions were first translated into Chinese and then back translated and compared for fidelity to ensure no information was lost (see Appendix B for the Chinese versions).

*Survey and Procedure*: Before answering questions about language use participants from the New Zealand sample were presented with one of three versions of the Cognitive Reflexivity Test (CRT). Like in Study 5 (in Chapter 3), there were three versions of the CRT: the 'no CRT' version (control version), the 'non-reflective' version (included the CRT), and the 'reflective' version (included the CRT and emphasised that the questions should be considered carefully). Again, the 'reflective' version also included instructions to consider the questions in the rest of the survey carefully.

Following the presentation of the CRT (all versions), participants received either two 'Twin Earth' cases or two 'Earth' cases to test their intuitions about natural kind terms. Each case included the description of a scenario followed by a question designed to elicit the relevant intuitions. The 'Twin Earth' cases described a scenario in which the alternate world Twin Earth is compared to Earth. The two cases were the 'Water' case:

Suppose that somewhere there is a planet called Twin Earth. Twin Earth is almost exactly like Earth; however, there is a tiny difference between the two planets: the colourless, odourless liquid that falls from the sky and fills the oceans, lakes and seas on Twin Earth is not H2O. Instead, it has a different chemical compound, XYZ. Indeed, XYZ is

indistinguishable from H2O at normal temperatures and pressures. Suppose now that a visitor, Oscar, to Twin Earth from Earth happily exclaims when he gets there "there is water on Twin Earth." He then proceeds to quench his thirst by drinking a glass of XYZ. When Oscar made the statement "there is water on Twin Earth" was his statement literally true? Choose the option that seems most correct:

- (1) Yes
- (2) No
- (3) Not sure

If you circled (1) or (2) above, choose the reasons that best support your answer. (You can circle more than one.)

(A) **-Yes-** because Oscar means the same thing as Twin Earthians do when they use the word 'water'; they all mean something like 'the colourless, odourless liquid that falls from the sky and fills the oceans, lakes and seas.'

(B) -No- because when Oscar uses word 'water' he means something like
"whatever scientists on Earth refer to when they use the word 'water'".

(C) **-Yes**- because Oscar is speaking the local language of the people on Twin Earth when he uses the word 'water', and what they mean (unbeknownst to Oscar) is 'XYZ.'

(D) **–No–** because Oscar means something different from Twin Earthians when they use the word 'water'; Oscar means 'H2O' rather than 'XYZ.'

(E) Because of some other consideration. (Please write your answer/s and reason/s below.)

And the 'Metal' case:

The metal with the atomic number 13 is referred to as "aluminium" by experts on Earth. The metal with the atomic number 42 is referred to as "molybdenum" by experts on Earth. Both metals are hard and gray to silver in colour; however, the metal with the atomic number 42 is rarely found on Earth. Suppose that the metals are indistinguishable except by scientific experts. Now suppose that somewhere there is a planet called Twin Earth. Twin Earth is almost exactly like Earth; however, there is a tiny difference between the two planets: on Twin Earth the metal with the atomic number 42 is one of the most commonly found chemical elements while the metal with the atomic number 13 is rarely found. As such, most pots and pans on Twin Earth are made out the metal with the atomic number 42. Furthermore, on Twin Earth the metal with the atomic number 42 is called "molybdenum" while the metal with the atomic number 42 is called "aluminium". Suppose now that a

visitor, Carl, to Twin Earth from Earth happily exclaims when he gets there "there are aluminium pots and pans on Twin Earth." He then proceeds to pick up a pot made out of the metal with the atomic number 42.

When Carl made the statement "there are aluminium pots and pans on Twin Earth" was his statement literally true? Choose the option that seems most correct:

- (1) Yes
- (2) No
- (3) Not sure

If you circled (1) or (2) above, choose the reasons that best support your answer. (You can circle more than one.)

(A) -Yes- because Carl means the same thing as Twin Earthians do when they use the word 'aluminium'; they all mean something like 'hard, silvery metal that pots and pans are made out of.'

(B) -No- because when Carl uses word 'aluminium' he means something like
"whatever experts on Earth refer to when they use the word 'aluminium'".

(C) **-Yes-** because Carl is speaking the local language of the people on Twin Earth when he uses the word 'aluminium' and what they mean (unbeknownst to Carl) is 'has the atomic number 42.'

(D) **-No-** because Carl means something different from Twin Earthians when they use the word 'aluminium'; Carl means 'has the atomic number 13' rather than 'has the atomic number 42.'

(E) Because of some other consideration. (Please write your answer/s and reason/s below.)

The 'Earth' cases described scenarios set on Earth about actual natural kinds. These cases were the 'Gold' case:

Suppose the year is 1860, well before Henry Moseley developed the concept of an atomic number. William is an American prospector, looking for precious metals in the San Bernardino Mountains in Southern California. While looking in a valley, William discovers a large nugget of shiny heavy yellowish metal. William exclaims happily "I have found gold." The nugget William discovers happens to have the atomic number 34. The metal we call 'gold' today has atomic number 79.

When William makes the statement "I have found gold" is his statement literally true? Choose the option that seems most correct:

(1) Yes

(2) No

(3) Not Sure

If you circled (1) or (2) above, choose the reasons that best support your answer. (You can circle more than one.)

(A) -Yes- because William means the same thing as we do when we use the word 'gold'; we all mean something like 'shiny heavy yellow metal'.

(B) -No- because William means the same thing as we do when we use the word 'gold', but what we all mean (unbeknownst to William) is 'has the atomic number 79'.

(C) **-Yes-** because William means something different to what we do when we use the word 'gold'; William means 'something with the atomic number 34' whereas we mean 'something with the atomic number 79.'

(D) **-No-** because William means the same thing we do when we use the word 'gold', we all mean "whatever scientists—chemists and geologists—refer to when they use the word 'gold'".

(E) Because of some other consideration. (Please write your answer/s and reason/s below.)

And the 'Tree' case:

Suppose that beech trees can't be distinguished from elm trees save by an expert. Both beech and elm trees share important features: they are most common in North America, they are tall, they have light bark and they have small oval leaves which they shed in the winter. Nonetheless, beech trees belong to the plant family *Fagaceae* whereas elm trees belong to the plant family *Ulmaceae*. Suppose Jane (like most people) can't tell the difference between trees from the *Fagaceae* and *Ulmaceae* plant families. One hot day Jane stumbles into a forest of *Ulmaceae* plant family trees and says excitedly, pointing to them, "they are beech trees."

When Jane makes the statement "they are beech trees" is her judgment correct? Choose the option that seems most correct:

(1) Yes

(2) No

(3) Not Sure

If you circled (1) or (2) above, choose the reasons that best support your answer. (You can circle more than one.)

(A) **-Yes-** because Jane means the same thing that most people mean when they use the word 'beech'; they all mean something like 'the tall deciduous tree commonly found in North America with light bark, small oval leaves, etc.'

(B) **–No–** because Jane means the same thing that most people mean when they use the word 'beech', but what they all mean (unbeknownst to Jane) is something like 'from the *Fagaceae* plant family'.

(C) **-Yes-** because Jane means something different to what the experts mean when she uses the word 'beech'; Jane means something like 'from the *Ulmaceae* plant family' whereas the experts mean something like 'from the *Fagaceae* plant family.'

(D) **–No–** because Jane means the same thing that most people do when they use the word 'beech', they all mean "whatever the experts refer to when they use the word 'beech'".

(E) Because of some other consideration. (Please write your answer/s and reason/s below.)

For each case, participants first answered whether or not they thought the judgement made by the agent was correct and then gave a reason for their answer. Both an Internalist and an Externalist reason were available for both of the 'yes' and the 'no' answers to the initial judgement question. Participants were also able to give their own reason, i.e., (E) responses. For participants who provided their own reason, the reason was categorised as an Externalist reason or an Internalist reason. The order in which the cases were presented was reversed so approximately half of the participants got either the Water case or the Gold case first and the Metal case or the Tree second (Order-a) while the remaining participants got the reverse order (Order-b).

### 4.8 Experimental results

*CRT*: Participants who received the CRT were given a score from 0-3. The mean score was 1.16 (*SD*= 1.06). Although participants given the 'reflective' version (M= 1.22, *SD*= 1.09) of the CRT scored higher than participants given the 'non-reflective' version (M= 1.09, *SD*= 1.03) as expected, this difference was not found to be significant [U = 7072.00, p = .36].

		Internalists	Externalists	Mixed
NZ sample				
	Gold Case	45.5	50.6	3.9
	Tree Case	31.6	66.2	2.2
	Water Case	32.9	54.9	12.1
	Metal Case	32.9	56.3	10.8
Chinese sample				
	Gold Case	41.7	54.2	4.2
	Tree Case	54.3	41.3	4.3
	Water Case	38.3	46.8	14.9
	Metal Case	34.7	55.1	10.2

Table 1. show the categorisation of participants (%)

Natural kind cases: Participants who initially answered "Yes" to the judgement question and then endorsed a 'No' reason and participants who initially answered "No" to the judgement question and then endorsed a 'Yes' reason were excluded from analyses. Participants who initially answered "Not sure" to the judgement question were also excluded from analyses. Participants were categorised as 'Externalists', 'Internalists', or 'Mixed' for each of the two natural kinds cases that they received based on their responses. The 'Mixed' category was included for participants who endorsed more than one consistent reason (i.e., either both the 'Yes' reasons or both the 'No' reasons). There were 40 instances (31 from the Earth cases and 9 from the Twin Earth cases) in which participants responded (E) and were categorised based on their own reason. Of these 40 responses, 23 were successfully categorised as either an Internalist or an Externalist response (the other 17 responses were unclear and omitted from further analyses). Of these 23 categorised responses there were 12 Internalist responses (9 from the Earth cases and 3 from the Twin Earth cases) and 11 Externalist responses (8 from the Earth cases and 3 from the Twin Earth cases). Table 1 shows the percentage of participants classified in each category for each sample. Each participant was also assigned a total rating score between -2 (Strong Internalist) and 2 (Strong Externalist) based on their responses across the two natural kinds cases they received (see Table 2). Participants whose total rating score was 0 (neutral) were either categorised as 'Mixed' for both the natural kinds cases they received or because they were categorised as an 'Internalist' on one case and an 'Externalist' on the other case.

Table 2. shows the proportions of participant assigned each Total Rating Score from -2 (Strong Internalist) to 2 (Strong Externalist) for both the NZ and the Chinese samples.

Earth		Tv	win Earth	
-2	12.2	-2	11.1	
-1	14.5	-1	14.3	
0	29.1	0	29.1	
1	17.4	1	14.8	
2	26.7	2	30.7	
Chinese sample				
Earth		Τι	win Earth	
-2	22.2	-2	27.3	
-1	16.7	-1	00.0	
0	22.2	0	20.0	
1	16.7	1	16.4	
2	22.2	2	36.4	

# NZ sample

*New Zealand sample*. For participants in the NZ sample, a majority of the participants gave Externalist responses to each case (see Table 1). However, the greatest endorsement of Externalist responses was only 66.2% (on the Tree case) indicating a substantial proportion of participants still gave Internalist responses. Indeed, only on the Tree case did a Binomial test indicate a significantly greater proportion of Externalist responses (than Internalist and Mixed, p < .001). On the Gold case, Water case, and Metal case there is not significantly greater proportions of Externalist responses. The responses of participants who received the Twin Earth cases and participants who received the Earth cases were also compared to see if participants responded differently to each of these types of cases. This was tested by comparing the Total Rating Scores of participants which combined their responses to the two cases received.<sup>98</sup> However, a Mann-Whitney test indicates no difference between the Total Rating Scores of participants in the Earth group and the Twin Earth group (U =15728.00, p = .584). One possible difference between the two groups is a greater proportion of 'Mixed' responses to the Twin Earth cases than to the Earth cases (see Table 1). That is, it appears that participants were more likely to give both an Internalist and an Externalist response to the Twin Earth cases than to the Earth cases. A Mann-Whitney test confirmed that this difference is significant (U = 43974.50, p < .001). There were no differences found in responses between CRT groups, gender, class, or order.

*Chinese sample*. For the Chinese sample it is less clear whether there is greater support for Externalism or Internalism (see Table 1). Indeed, there are greater proportions of Externalist responses on three of the cases (the Gold case, the Water case, the Metal case); however, these proportions of Externalist responses are not significant on any of the three cases. On the Tree case there is a greater proportion of Internalist responses; but, again, this proportion is not significantly greater than the Externalist and Mixed responses. Like in the NZ sample, a Mann-Whitney test indicates no difference between the Total Rating Scores of participants in the Earth group and the Twin Earth group, U = 1285.50, p = .214. Again, it appeared that there is a greater proportion of 'Mixed' responses to the Twin Earth cases than to the Earth cases similar to the NZ sample (see Table 1). However, this difference was

<sup>&</sup>lt;sup>98</sup> Note that for the Earth cases a Wilcoxon Signed-Ranks Test indicated a significant difference between responses to the Gold case and responses to the Tree case (Z = 2.14, p = .032, r = .1). However, because of the small effect size, and because no difference was found between the Twin Earth cases (Z = .245, p = .806) or for either the Earth or Twin Earth cases in the Chinese sample (Z = 1.28, p = .201 and Z = 1.01, p = .312), it seemed justified to combine the responses across the two cases to get the Total Rating Scores.

not borne out statistically, possibly because of the reduced sample size (U = 4483.00, p = .058).<sup>99</sup>

Differences between samples. Mann-Witney U tests indicate that the only difference between the NZ sample and the Chinese sample is on the Tree case, U = 2355.50, p = .003[U = 3556.00, p = .653, U = 4122.00, p = .310, U = 4145.50, p = .903, for the Gold case, Watercase and Metal case respectively]. This finding indicates that participants from the Chinesesample were more likely than participants from the NZ sample to be categorised as an'Internalist' on the Tree case. When collapsing across the two cases in each of the Earth andTwin Earth groups no significant differences were found between the two cultural samples<math>[U = 4066.50, p = .158, and U = 5154.00, p = .922 for the Earth and Twin Earth groups respectively] (See Figure 1). This lack of cross-cultural variation in intuitions contrasts Studies 5 and 6 about proper names from the previous chapter as well as other crosscultural experiments (e.g., in reference; Machery et al., 2004) where there have been significant differences uncovered between the different cultural samples.

<sup>&</sup>lt;sup>99</sup> Note that this is almost statistically significant at 5%.



*Figure 1.* Total categorisations of participants from -2 (Strong Internalist) to 2 (Strong Externalist) collapsed across both Twin Earth and Earth groups.

# 4.9 Study 8: US sample follow up

A follow up study with a US sample was also conducted. One hundred eighteen US participants were collected online (in the same manner as in Chapter 2). Sixty-eight participants were female (57.6%) and the mean age of participants was 38.64 (*SD* = 14.79). Half of the participants received the two Twin Earth cases and half of the participants received the two Earth cases. Note that all participants received the cases in the same order (Order-a).

	Internalist	Externalist	Mixed
Gold case	52.6	42.1	5.3
Tree case	47.5	47.5	5
Water case	45.5	43.2	11.4
Metal case	40	47.5	12.5

Table 3. shows the categorisation of participants (%) for the US sample.

The data from this sample indicate a lot of variation in responses (see Table 3). On two of the cases there is a greater proportion of Internalist responses (the Gold case and the Water case), on one case there is a greater proportion of Externalist responses (the Metal case), and on one case there is equal support for Internalism and Externalism (the Tree case). None of these differences turned out significant (e.g., there was not a significant difference between Externalist responses and other responses on the Metal case). This finding indicates that there were very similar amounts of support for Internalist and Externalist responses. The intuitions of participants who received the two Twin Earth cases was compared to participants who received the two Earth cases by comparing the Total Rating Scores. A Mann-Whitney test indicated no difference in responses, U = 1024.00, p = .541 (see Table 4).<sup>100</sup> Like in the NZ and the Chinese sample, it appears that there is a greater proportion of 'Mixed' responses to the Twin Earth cases than to the Earth (see Table 3). This was not borne out statistically, again possibly because of the smaller sample size (U = 3059.00, p = .239)

	Earth	Twin Earth
-2	17.8	20.4
-1	22.2	8.2
0	31.1	34.7
1	8.9	20.4
2	20	16.3

Table 4. shows the Total Rating Scores (%) for the US sample.

<sup>&</sup>lt;sup>100</sup> Note that there were no differences found in responses between each type of case; that is, between the two Twin Earth case and between the two Earth cases. This justifies combing the responses for each participant to get a Total Rating Score.

### 4.10 Discussion

The aim of Studies 7 and 8 was to uncover the semantic intuitions of ordinary language users about natural kind terms. To this end the following questions have been addressed by these experiments:

- 1. Is there evidence for The Ambiguity Thesis?
- 2. Are there cross-cultural differences?
- 3. Are there differences between considering Earth and Twin-Earth cases?
- 4. Does reflectiveness affect intuitions?
- 5. Is there evidence of a philosophical bullying phenomenon?

I will discuss the results of Studies 7 and 8 in terms of these questions; in this section I will discuss the first four questions, then, in Section 4.10, I will discuss the last question.

Firstly, the results of these studies indicate a large amount of variation in the intuitions of participants. This variation is good evidence for the Ambiguity Thesis. Although the results appear to indicate that there is slightly greater support for Externalism than Internalism, there were substantial disagreements found on all four of the cases among all of the cultural samples. Indeed, the greatest proportion of support for Externalism was only about 66% (by the NZ sample on the Tree case). Moreover, in the US sample there was actually slightly more support for Internalism. The Total Rating Scores (which combines a participants responses across the two cases they received) indicate that many participants did not have strict Externalist or Internalist intuitions, but rather that they had different intuitions across the two cases (i.e., gave an Internalist response to one case and an Externalist response to the other case) or had mixed intuitions to one or both cases (i.e., gave both an Internalist and an Externalist response to a case). This is easy to see from the Total Rating Scores by the proportions of participants who did not receive a score of -2 or 2, scores which would indicate that either only Internalist or only Externalist responses were made by that participant. This variation is accounted for by the Ambiguity Thesis. Recall the generalised version of the Ambiguity Thesis from Chapter 3, Section 3.9:

**The Ambiguity Thesis (general):** some terms and are ambiguous and have multiple possible meanings associated with them. A native speaker can use these terms and phrases with any of the possible contents depending on their communicative intention; no one disambiguation is correct outside of a context of utterance.

In this instance what the Ambiguity Thesis holds is that natural kind terms are ambiguous between an Externalist reading and an Internalist reading. A native speaker can disambiguate these terms with the content of either reading depending on the context of the utterance. The variation uncovered in Studies 7 and 8 is explained by participants disambiguating the natural kind terms in the given cases in different ways. The slight preference for Externalist disambiguations suggests that in these contexts the Externalist readings seemed like the more natural interpretation; however, that there is this preference is very unclear which indicates that Internalism also gives a very natural interpretation much of the time.

In terms of cross-cultural differences, unlike Studies 5 and 6 about proper names, there was only minimal cross-cultural variation uncovered in Studies 7 and 8. It was expected that there would be similar cross-cultural differences uncovered as in the proper names study; in this case as a preference for the Chinese sample to give Internalist responses and for the NZ sample and the US sample to give Externalist responses. However, while Externalist responses were preferred by the NZ sample to a slight degree, it is not clear that there is much of a preference between Externalism and Internalism for the other two samples. What this suggests is that, unlike in the case of proper names, the contextual cues for adopting each reading are similar across the cultural groups. That is, instances in which NZ and US native English speakers consider Externalism to give the most natural reading of a natural kind term also tend to be instances in which native Chinese speakers consider Externalism to give the most natural reading. This difference between natural kind terms and proper names might suggest that the two kinds of term are not as similar as some philosophers (e.g., Kripke, 1980) have believed; while Chinese native speakers have a preference to give a meaning consistent with the Frege-Russell Theory to proper names, they do not seem to also have a preference to give an Internalist meaning to natural kind terms. Alternatively it might simply be that the cases used in Studies 7 and 8 lent themselves particularly well to elicit Externalist intuitions from the Chinese sample. Indeed,

the Ambiguity Thesis explains variation in intuitions with context and surrounding content playing a large determining role. Had different scenarios been used it is possible that a cultural difference would have emerged.

Similarly to the lack of cultural variation, there also did not appear to be a difference between the Earth and Twin Earth cases. There was no difference found between participants who received the Earth cases compared to participants who received the Twin Earth cases in any of the cultural samples. It was expected that, if there was evidence of the Ambiguity Thesis, a relevant cue for determining whether Internalism or Externalism gave the best reading in a given instance would be whether a speaker was considering the actual world or an alternate world. This does not seem to be the case. However, one interesting feature worth noting is that in all of the cultural samples a larger proportion of participants in the Twin Earth group were categorised as 'Mixed' compared to participants in the Earth group (although this difference was only confirmed statistically for the NZ sample). This indicates that participants who received Twin Earth cases were more likely to endorse both an Internalist *and* an Externalist reason for why they considered an agent's judgement to be correct or incorrect. This might indicate that participant's intuitions were less clear when considering Twin Earth. I will talk more about this finding below in Section 4.12.

In terms of whether reflectiveness affects intuitions about natural kind terms, Study 7 did not find any evidence that they do. For the NZ sample, who also received a version of the CRT, there was no differences in the intuitions of participants based on either the version of the CRT they received or their CRT score. Moreover, although there was a significant gender difference in CRT score, this difference did not translate into gender differences on the natural kind cases. This finding is similar to Study 5 about proper names in which only a minimal difference was found between the 'reflective' version and the 'no CRT' version. The lack of connection between reflective thinking and philosophical intuitions in Study 7 challenges the claim that philosophers' intuitions are more reliable because they are more reflective (see the Expertise Objection; Chapter 1, Section 1.5). Indeed, it suggests that we should not dismiss the informativeness of ordinary language users as it does not appear that we can ignore the intuitions of ordinary language users for being less reflective than those of philosophers. Thus, any difference between the intuitions of philosophers and nonphilosophers, at least in this instance, should not be settled by appealing solely to the

philosophers (although as I will discuss in the next section, it does not look as though there is such a difference).

### 4.11 Philosophical bullying/filtering

A new question that I wanted to investigate in this study was whether there was any indication of philosophical bullying. The reason for this addition was due to the previous literature suggesting that there is a difference between the intuitions of philosophers and non-philosophers about the semantic content of natural kind terms. This is seen by the tendency for philosophers to espouse Externalism (while this claim seems to be assumed by many philosophers it is also suggested by Bourget & Chalmer's study; see Section 4.2) along with the verbal reports from philosophers that their undergraduate students have Internalist intuitions (e.g., Jackson; see Section 4.2). I suggested that if these verbal reports were borne out in the present studies across all samples, that is, if it was found that participants did tend to hold Internalist intuitions, then a good explanation would be that a philosophical bullying phenomenon was occurring. Indeed, I suggested that this explanation would be motivated over the explanation offered by the Expertise Objection, namely, that the difference occurs because philosophers have superior intuitions, if there was not global divergences in the intuitions of non-philosophers as there is in philosophers (found in Bourget & Chalmers study). Moreover, I suggested that if reflective thinking did not affect philosophical intuitions (which it did not) then this would be evidence against the Expertise Objection. Any evidence of this phenomenon would warrant investigation into the practices of teaching philosophy as well as further experimental philosophy to determine how widespread this phenomenon is occurring within philosophy. However, the present results do not support this phenomenon, at least with regard to natural kind terms. Indeed, the intuitions of participants in Studies 7 and 8 seem to mirror the intuitions of philosophers, at least in regard to Bourget and Chalmers (2013) study; that is, some variation between Externalism and Internalism with a stronger tendency towards Externalism. This finding is also evidence against the Expertise Objection as the advocate of the Expertise Objection should also contend that the greater philosophical expertise of philosophers leads to a greater consensus among philosophers than among in philosophers' non-philosophers. This

was obviously not the case in this chapter.

The data from the US participants is not as clear on this point. In this sample there was not greater support for either Externalism or Internalism. This contrasts Bourget and Chalmers' study and thus indicates some difference between philosophers and non-philosophers. However, note here that the US sample was not explicitly comprised of undergraduate students, the sample instead is best categorised as the general public. What this might suggest is that the general public do not have a tendency for Externalist intuitions even if philosophers and undergraduate philosophy students do.

Although there is a lack of evidence of philosophical bullying was uncovered, I think it is a viable explanation for any case in which there is a discrepancy between the intuitions of philosophers and non-philosophers. Philosophers have in the past assumed that a discrepancy between the intuitions of philosophers and non-philosophers should be taken as evidence in favour of the Expertise Objection. The possibility of a philosophical bullying phenomenon provides an alternative explanation that would warrant more rather than less experimental work.

# 4.12 Alternative explanations

In this section I will introduce and reject a number of possible alternative explanations of the present data.

- I have already discussed and rejected an interpretation of the data according to which intuitions are unreliable and uninformative (see Chapter 2, Section 2.5.3), and so any distribution of intuitions amongst non-philosophers would be compatible with the correct semantic theory about natural kind terms. Because Meta-Internalism is well motivated, this hypothesis should be ruled out.
- 2) Another explanation is that Externalism is supported and the non-Externalist responses can be explained as noise. This is the approach that Jyllkä et al. (2009) take in their study. While their study uncovered a lot of variation in the intuitions of participants about natural kind terms, they concluded that the data provides support

for Externalism because there was a greater proportion of Externalist responses compared to Internalist or hybrid responses. Because participants could explicitly make a response consistent with a hybrid theory, that is, 'on the one hand yes, on the other no', Jyllkä et al. did not think a hybrid account could be used to account for the variation. Instead they explained the variation as participants having unclear intuitions. Similarly, because the present results indicate more support of Externalism than Internalism, the data could be interpreted as evidence only for Externalism. The benefit of this conclusion is that it is consistent with the greater endorsement of Externalism by philosophers and thus suggests that both philosophers and non-philosophers tend to have Externalist intuitions, but that there is some variation due to some non-philosophers and, to a lesser extent, some philosophers having unclear intuitions on the matter. However, I do not think that this interpretation accurately captures the variation uncovered in either Jyllkä et al.'s study or Studies 7 and 8. In Section 4.4 I argued that Jyllkä et al.'s data actually look like good evidence of the Ambiguity Thesis. The reason that their hybrid theory is not supported is because they do not test for an ambiguity view, but for a different kind of hybrid account, namely what looks like a pluralist view. Indeed, according to the Ambiguity Thesis we should not expect participants to give the hybrid response because the surrounding context in the vignettes forces us to give just one disambiguation of the natural kind term. Participants will naturally give one disambiguation to a term but the particular disambiguation will differ participant to participant. Thus, we should expect to see a mix of Internalist and Externalist responses. This is exactly what we do witness in Jyllkä et al.' study and the present studies. Moreover, there is an extremely high proportion of participants that give different responses to different cases in both Jyllkä et al.'s study (81%) and the present studies (approximately 60% of the NZ sample were not categorised as either a strict Externalist or a strict Internalist). This variation in responses is accounted for on the Ambiguity Thesis. It does not sit so well with an explanation that any variation from Externalist responses is due to unclear intuitions because it suggests that a majority of participants then have unclear intuitions some of the time. A theory that predicts widespread confusion in this way is not the most charitable interpretation of the data if an alternative explanation is consistent with the view that participants

are usually rational and have clear intuitions.

3) A third explanation is that the results could be explained by a component view. A component view is a type of hybrid theory according to which there are at least two different and independent factors - e.g., perhaps a term's referent and its descriptive content— that jointly determines or constitutes its semantic value. On one version of this kind of view, a term like 'Elizabeth II' picks out the Queen of England (descriptive content) in all possible worlds where it picks out anything at all, but it never picks out anyone other than Elizabeth Windsor in any possible world, even if that person happens to be the Queen of England (referential content). That is, the meaning of 'Elizabeth II' fits both the descriptive content and the referential content of the term. Genone and Lombrozo (2012), discussed in Section 4.4, present a hybrid theory in which they argue both descriptive and causal factors influence people's intuitions about the reference of various terms such as natural kind terms. While they do not specify how their hybrid theory should be characterised, it looks like a good candidate for a component theory. We also see Jyllkä et al. (2009), following Braisby et al. (1993), present a hybrid theory according to which a natural kind term can be used with an 'externalistic sense' or an 'epistemic sense' (see Section 4.4).<sup>101</sup> Earlier I argued that the target of their study was best understood as a pluralist theory. However, here we see Jyllkä et al. suggest that it might not be enough that what is picked out by a natural kind term fit the descriptive content associated to the term, it might also need to have the right essential properties:

[E]ven the causal descriptivists are externalists in the referential sense: for instance, the description *actual watery substance* does not alone suffice to determine the reference of 'water', but only together with the actual, possibly unknown, nature of water. [Jyllkä et al. (2009) fn. 1]

This comment suggests that Jyllkä et al. actually had a kind of component theory in mind. A mixed theory differs to an ambiguity theory because, according to a component theory there is only one meaning of a term, but multiple factor determine that meaning. An ambiguity theory posits multiple independent semantic

<sup>&</sup>lt;sup>101</sup> Note that as we saw in the last paragraph, Jyllkä et al. (2009) do not think that a hybrid theory can account for their results but instead argue that their results support Externalism.

contents of a term; however, only one is ever expressed in a given occasion of utterance. Therefore, for something to be picked out by a term it only needs to meet the conditions of one content; that is, to be picked out by 'water' a substance only needs to fit the descriptive content 'watery substance...' *or* it only needs to have the chemical composition  $H_2O$ , depending on the context.

The present results do not support a component theory because most participants only gave an Internalist *or* an Externalist response. A component theory predicts that both sets of conditions need to be fulfilled. The vignettes describe a scenario in which an agent makes a statement about a substance that only fits the descriptive content of a term, that is, only fulfils the Internalist conditions. In order to be consistent with a component theory participants' would need to have given an Externalist response; the Internalist responses alone do not account for the Externalist conditions. Thus, the following responses are at odds with a component theory:<sup>102</sup>

-Yes- because Oscar means the same thing as Twin Earthians do when they use the word 'water'; they all mean something like 'the colourless, odourless liquid that falls from the sky and fills the oceans, lakes and seas.'

**-No**- because when Oscar uses word 'water' he means something like "whatever scientists on Earth refer to when they use the word 'water'".

We also see this in Genone and Lombrozo's (2012) study. Although they argue that both causal and descriptive factors are important for determining reference, as it was their participants only relied on *one* of these factors at a time. That is, in the cases where responses were mixed at roughly 50% between 'same referent' and 'different referent', individual participants actually had very strong intuitions one way or another. In one of their experiments participants made their responses on a 7-point scale. This allows us to see how important participants considered each of the two factors to be. As it was, Genone and Lombrozo found that most participants gave a very high or very low rating. See the following:

<sup>&</sup>lt;sup>102</sup> Note these responses are only at odds with a component theory when they are given in isolation. Participants were able to give multiple responses so could have given one of these Internalist responses AND one of the Externalist responses which would have accorded with a component theory.

While mean ratings for shared reference were again near the middle of the scale (4.44 on a 1–7 scale), this was not because a majority of participants had weak or middling intuitions. Figure 2 reports the distribution of ratings on the 1–7 scale, and reveals what appears to be a bimodal distribution: the mean of 4.44 resulted not from a majority of participants clustered around 4 and 5, but rather from a combination of ratings near the bottom and top of the scale.12 This suggests that participants had strong intuitions in response to the questions about shared reference. [p. 730-731]

This indicates that participants did not think that both causal and descriptive factors were important at the same time, but instead that only one factor was important for determining reference. This finding is evidence against a component theory, and looks like support for an ambiguity (cf. Brock, 2004, p. 283).

4) We might also consider pluralist theories, the other view consistent with a hybrid theory, to explain the present data. As noted, pluralist theories hold that individual terms have multiple semantic values, and multiple distinct contents are associated with the term on any occasion of utterance. One such pluralist theory that might be considered here is neo-descriptivism. Nimtz (2004) describes neo-descriptivism as a descriptivist (i.e., Internalist) theory that is combined with a two-dimensional semantics which helps it overcome the problems descriptivism faces.<sup>103</sup> According to neo-descriptivists (e.g., Jackson and Chalmers) a natural kind term has two intensions, each with a different semantic content. Chalmers (2006) puts this point as follows:

Two-dimensionalism is naturally combined with a *semantic pluralism*, according to which expressions and utterances can be associated with many different semantic (or quasi-semantic) values, by many different semantic (or quasi-semantic) relations. On this view, there should be no question about whether the primary intension or the secondary intension is *the* content of an utterance. Both can be systematically associated with utterances, and both can play some of the roles that we want contents to play. [p. 601]

The primary intension captures the descriptive content associated with a natural kind term, which fixes the extension of that term for every world considered as

<sup>&</sup>lt;sup>103</sup> Note that Nimtz (2004) thinks that neo-descriptivist's still face the same problems that classical descriptivism does.

actual. Thus, it accounts for epistemic possibilities about how the actual world could have been. Accordingly the primary intension is understood *a priori* by a speaker. The secondary intension of a natural kind term fixes the extension of that term in all worlds considered as *counterfactual*. That is, the extension of a natural kind term is determined by the actual state of the world; thus, 'water' rigidly designates  $H_2O$ . According to neo-descriptivism, Internalism gets the primary intension right, while Externalism gets the secondary intension right. Thus, neo-descriptivism accommodates both the Kripkean intuition that 'water' picks out the same stuff in all possible worlds (considered as counterfactual), but also the intuition that 'water' could have had a different extension than ' $H_20$ ' (i.e., in different worlds, considered as actual, the term 'water' picks out different stuff). Importantly, according to neodescriptivism, both of these intensions are expressed in every instance that a natural kind term is used. Thus, neo-descriptivism, as well as any pluralist theory, differs to an ambiguity theory because an ambiguity thesis contends that only one of multiple possible meanings is expressed by a token use of an ambiguous term. Chalmers makes this distinction between pluralist and ambiguity theories clear when he defends his two-dimensional theory against the objection that ordinary expressions are not ambiguous (cf. Bealer, 2002, and Marconi, 2005). This objection states that two-dimensionalism cannot account for the statement 'it is metaphysically necessary but not epistemically necessary that water is H<sub>2</sub>O' because it accounts for the two claims, 'it is metaphysically necessary that water is H<sub>2</sub>O' and 'it is epistemically necessary that water is  $H_2O'$ , by holding that water expresses the primary intension in the first case, and the secondary intension in the second case. Chalmers (2006) makes the following response:

Two-dimensionalism does not hold that ordinary expressions are ambiguous. 'Water' has exactly the same content in both (3) [it is metaphysically necessary that water is  $H_2O$ ] and (4) [it is epistemically necessary that water is  $H_2O$ ] above: in both contexts (and in all contexts) it has both a primary intension and a secondary intension (or equivalently, it has a complex semantic value involving both a primary and a secondary intension). This does not entail that 'water' is ambiguous, any more than the distinction between character and content entails that indexicals are ambiguous. [p. 610]
Chalmers makes it clear that on his view a term has both the primary intension and the secondary intension in every instance it is used. Thus, 'water' always has the same contents.

We might think that a pluralist theory can account for the variation in the present results because if both an Externalist and an Internalist meaning are expressed by a natural kind term, it looks as though neo-descriptivism, at least, could account for the responses made by participants. Thus it looks like we can explain the variation in responses based on whether participants were considering Twin Earth as actual or as counterfactual. However, I think there a few reasons to resist a pluralist explanation of the present data. Firstly, pluralist theories hold that natural kind terms always have the same content. This claim is at odds with the number of participants who changed their response across different cases. That is, participants who gave an Internalist response to one case and an Externalist response to the other case they received. Because the content is considered to always be the same, pluralism predicts that participants should always give the same response. Secondly, pluralism predicts that this response should be indicative of both Internalism and Externalism simultaneously. However, only a very small proportion of participants gave both an Internalist and an Externalist response to a case (i.e., were labelled 'Mixed' for a given case). On the Earth cases the proportion of participants were labelled 'Mixed' was negligible (between 2.2% and 5.3%). These responses should, I think, be dismissed as noise. On the Twin Earth cases, this is less clear as there was a greater proportion of 'Mixed' responses (up to 14.9%). Furthermore, this proportion was found to be significantly greater than the proportion of 'Mixed' responses for the Earth cases for the NZ sample. This might suggest that the 'Mixed' responses to the Twin Earth cases should not be dismissed as noise. Nonetheless, if pluralism was correct we should expect there to be a much greater proportion of participants labelled 'Mixed'. Indeed, given that participants were explicitly instructed that they could choose more than one response, if the participants considered that more than one reading was expressed then we should expect that they would have made use of this option.

The proportion of 'Mixed' responses on the Twin Earth cases makes an interesting

finding as it does not clearly support either the Ambiguity Thesis or a pluralist theory. There are two ways I think it could be explained, one which takes the neodescriptivist view. Firstly, it might be that the Ambiguity Thesis is correct, however, participants sometimes think a term is captured by both Internalism and Externalism in cases set on Twin Earth because they have difficulty considering the counterfactual case. That is, some participants might have had trouble committing to a single reading because they were not really imaging that the counterfactual scenario had or could occur. Without really imaging the scenario, context would not be able to determine which reading is expressed. Because, according to the Ambiguity Thesis, it is context and surrounding content that determines which reading is expressed in a given instance, if participants did not consider the counterfactual cases to give a context that could rule out a reading, they might have given both responses simply because it would avoid giving the 'wrong' answer. A second explanation is that neo-descriptivism is correct, but that many participants opted to answer from the perspective of only one of the intensions. That is, all participants should have given 'Mixed' responses on the Twin Earth cases, but many did not and decided instead to choose whether the world should be considered as actual or counterfactual. This would explain the greater proportion of 'Mixed' responses to the Twin Earth cases because in the Earth cases a natural kind, such as 'gold', has the same extension on both the primary and secondary intension, namely 'entities with the atomic number 79'. In the Twin Earth cases, on the other hand, they have different extensions. If it is the case that this neo-descriptivist explanation can account for the data in Studies 7 and 8, a further study could aim to rule this explanation out. This avenue for future research is one I discuss in Chapter 5.

5) Lastly, we might consider a pragmatic explanation. An important project across this thesis has been to rule out any possible pragmatic interpretation of the data to show that any ambiguity uncovered is a genuine semantic ambiguity. While I do not know of anyone who has objected to the postulation of an ambiguity in natural kind terms by appealing to pragmatic considerations, I do think that an appeal to the semantic-pragmatic distinction would make a good response to some of the previous work in this area. To illustrate this I will discuss three studies which have uncovered variation

in the responses made to statements containing natural kind terms. In each of these three studies the authors offer some explanation of the variation. I will argue that a pragmatic explanation might better account for the experimental results of all of these studies. I will then argue that the same appeal to pragmatic considerations is not applicable to Studies 7 and 8.

One study I discussed above, in Section 4.4, was conducted by Braisby et al. (1996). In this study they presented participants with scenarios about a natural kind for which something is discovered about an essential or a non-essential property of either a single member of the natural kind or all members of that natural kind. They then asked participants whether they agreed or disagreed with a number of statements about the natural kind. What they found was lot of variation in the responses of participants so there was not clear support for either Internalism or Externalism. Moreover, from the same scenario some participants were willing to endorse a number of seemingly contradictory statements like 'Tibby is a cat' and 'Tibby is not a cat'. They argue that these results do not support (something like) Externalism. Instead, they explain the variation by appealing to a 'representational change' theory. According to this theory, ordinary language users can make use of different concepts associated to a natural kind term. They might associate a natural kind term to a set of non-essential properties which determine the reference. Or, after a discovery is made, they might associate a set of essential properties to the natural kind term. Although they do not state it explicitly, their theory looks like an ambiguity thesis; however, each disambiguation is determined by the mental states of the users, so it appears to be an Internalist theory. While Jyllkä et al. (2009) give some grounds to criticise Braisby et al.'s study, I will contend that a pragmatic theory could account for the results after discussing the next two studies.

The second relevant study was conducted by Machery and Seppälä (2010) (mentioned in Chapter 1, Section 1.2) in which they found that participants were willing to endorse two seemingly contradictory sentences, such as, 'in a sense tomatoes are vegetables' and 'in a sense tomatoes are not vegetables', about half the time. What they argue is that these results support the 'heterogeneity hypothesis', later amended to the 'polysemy hypothesis' by Machery (2014).

According to the polysemy hypothesis, ordinary speakers represent some terms with several different co-referential concepts. So a term like 'dog' has several meanings because it can lexicalise any of the concepts associated to the term by the speaker. Specifically Machery claims that these co-referential concepts might be a prototype of dogs, a set of exemplars of dogs, or a causal theory of dogs. Notably, the results of Machery and Seppälä's study might also be explained on a pluralist account. I will argue shortly, (and as I mentioned in Chapter 1, Section 1.2), that a pragmatic explanation could also account for these results.

In another study, Nichols, Pinillos and Mallon (Forthcoming) looked at how participants thought the name of a legendary creature, namely the 'Catoblepas', referred. They presented participants with a vignette about this creature; one version of the vignette went as follows:

In the Middle Ages, animal researchers described a distinctive kind of mammal. They called it catoblepas. The catoblepas was said to be like a bull but with a head so heavy that the animal has to keep its head down at all times. It was also thought that the catoblepas had scales on its back. In addition, the researchers said that looking into the animal's eyes causes immediate death. Of course there is nothing that meets this description, but researchers know that it was based on reports of encounters with wildebeest. Many scientists in the middle ages, when faced with wildebeest, would often call them catoblepas.

Participants were then asked to rate how strongly they agreed with the following statements (on a 6-point scale):

- (1) Catoblepas exist.
- (2) Catoblepas are wildebeest.

While there was some variation in responses, what Nichols et al. found was that participants tended to disagree with (1) and agree with (2). This, Nichols et al. argue is evidence for both a descriptivist theory of reference [in the case of (1)] and a causal-historical theory of reference [in the case of (2)]. They use these results as evidence for an ambiguity theory of reference. That is, that the reference of a natural kind term can be determined by either the causal-historical theory or a

#### descriptivist theory.

In all three of these studies, by Braisby et al., Machery and Seppälä, and Nichols et al., there is evidence that natural kind terms are ambiguous in some sense because of the variation uncovered in intuitions. In each case we see the authors offer some kind of explanation that appeal to an ambiguity in the terms, or at least in the reference of the terms (note that these ambiguity theses are not the same as the Ambiguity Thesis defended in this thesis).<sup>104</sup> However, I think that the results of all three of these studies could be easily, if not better, explained by appealing to pragmatic consideration. In the case of participants endorsing seemingly contradictory sentences, such as 'Tibby is a cat' and 'Tibby is not a cat' (Braisby et al., 1996), and 'in a sense tomatoes are vegetables' and 'in a sense tomatoes are not vegetables' (Machery, 2014), it might be that participants were willing to find an interpretation of the sentences that made them non-contradictory, even if, interpreted literally, they would take the sentences to be contradictions. For example, participants might have endorsed 'in a sense tomatoes are not vegetables' because tomatoes are actually classified as a fruit (semantic meaning) but also endorsed 'in a sense tomatoes are a vegetable' because we treat tomatoes like vegetables (speaker's meaning).<sup>105</sup> If participants were given the opportunity to give reasons for their responses in these studies any instances of speaker's meaning might be revealed. Alternatively, these studies might have been able to rule out the possibility of participants appealing to pragmatic interpretations by changing the sentences to reflect that it is the semantic meaning that is being sought; for instances it might thus be that if participants were given the statements 'tomatoes are actually vegetables' and 'tomatoes are actually not vegetables' that participants would only endorse one of the two sentences. In the case of Nichols et al.'s study, it might be that participants disagreed with 'Catoblepas exist' because there are no actual Catoblepas, and that participants agreed with 'Catoblepas are wildebeest'

<sup>&</sup>lt;sup>104</sup> This is because these authors are either proposing that natural kind terms are ambiguous between different concepts that we associate with the terms or are ambiguous in the way that they refer rather than in the terms themselves.

<sup>&</sup>lt;sup>105</sup> It's worth noting that Machery & Seppälä (2010) in a follow-up experiment show that participants treated sentences like 'in a sense tomatoes are vegetables' as literal interpretations rather than metaphorical interpretations. However, it is not clear that this has a bearing on the semantic meaning/speaker's meaning distinction.

because what was thought to be Catoblepas were actually wildebeest. That is, participant might have been interpreting 'Catoblepas are wildebeest' as 'what was mistaken as Catoblepas are actually wildebeest'. If this is the case then the agreement with this sentence would be due to participants giving a charitable reading to the sentence. Notably, Nichols et al. mention this possibility, that there results could be confounding semantic reference with speaker's reference. However, they do not address this concern in any way.

Given that appeals to pragmatic considerations seem to account well for these previous experiment results, it might be that the same argument is made to Studies 7 and 8. That is, it might be argued that a unitary thesis is correct but that participants take into account what a speaker intends to communicate (because of a lack of knowledge). According to this argument, either all of the Internalist responses or all of the Externalist responses can be given a speaker's meaning interpretation; for instance, it might be that Externalism is correct but that participants sometimes give Internalist responses because from the agents point of view Internalism better captures what they were attempting to communicate. However, I attempted to thwart this kind of explanation in my experiments. This was attempted in two ways. Firstly, the questions explicitly asked whether an agent was *correct* in what they uttered. Obviously the agent would take themself to be correct, but by asking this question is was reasoned that participants would consider whether the agent was correct regardless of whether they considered themself to be. Secondly, this was reflected in the answer choices in which it was stated that what an agent says might mean something different to what the agent thinks they mean (e.g., 'unbeknownst to X'). It was thought that by providing the multiple different answer choices that a participant would be able to explain exactly what they meant by their answer. Moreover, participants were able to give their own response [i.e., respond (e)]. There were only forty instances in which participants gave their own response to one the four cases. In these responses we can see both instances of Externalist and Internalist responses that rule out a speakers meaning interpretation. Here are some examples of Externalist responses:

She is not right about the tree because she is using the wrong name even though she

doesn't know the difference.

Oscar believes it to be true in similar comparison but literally it is not true as XYZ has a different chemical compound so cannot be noted as water (or H2O).

No because what he thinks is gold and what is actually gold is not the same thing.

### And an Internalist response:

Yes- the object he has found is gold coloured therefore literally he has found gold.

These responses seem to indicate that participants were distinguishing semantic meaning from speaker's meaning in these experiments. Thus, I think there is good reason to consider the variation uncovered as indicative of a genuine semantic ambiguity in natural kind terms.

# 4.13 Conclusion

Studies 7 and 8 are further evidence for the Ambiguity Thesis, this time in the domain of natural kind terms. Ordinary language users have intuitions consistent with both Internalism and Externalism. The variation in responses is best captured by appealing to an ambiguity in the terms.

# **Chapter 5: Concluding Remarks**

### 5.1 Concluding remarks

The aim of this thesis has been to inform on three traditional debates about the semantics of definite descriptions, proper names, and natural kind terms using the methods of experimental philosophy. The eight studies presented in Chapters 2-4 provide new evidence regarding these debates and extend the findings of previous experimental work in these areas. Specifically, they provide evidence for the Ambiguity Thesis. The Ambiguity Thesis holds that some terms and phrases are ambiguous and have multiple possible meanings associated with them. These terms and phrases can be used by a native speaker with any of the possible disambiguations depending on the communicative intent. By endorsing the Ambiguity Thesis we can overcome many challenges that face traditional views about the semantics of referential and denoting terms, which presuppose that such terms are all univocal. Moreover, it accounts for the striking flexibility with which we use various terms, and does not have the consequence that we are using these terms incorrectly or metaphorically or non-standardly in many instances. Indeed, we know that people use various terms with different intentions, for example, we know that speakers often use definite descriptions referentially, and also to mean whatever fits the descriptive content of 'the F'. Moreover, we know that each use is widespread amongst native speakers of English. The Ambiguity Thesis maintains that neither use is non-standard. This fits well with the claim that the meaning of our terms is determined by how the terms are actually used. Many philosophers have made this kind of point; for example, Seegar's (2010) Constitutivism which holds that the meaning of some terms is constituted simply by how most people use them, and Devitt (2004) and Reimer (1998) who claim that if there is a common way a term is used then the meaning associated to that usage is a genuine meaning of the term. Indeed, it is this feature of language that I think makes it particularly amendable to the methodology of experimental philosophy. Moreover, if we take Meta-Internalism seriously, that the correct theory of meaning for some terms is determined by what the speaker takes to be the relevant determining factors, then it is entirely relevant what ordinary language users take the meaning of their terms to be (see Section 4.2).

I think that the results uncovered in this thesis also justify future experimental work in this area. Indeed, I think there are at least six future avenues that could be investigated:

- Whether there is evidence for the Ambiguity Thesis for any other kinds of predicates, denoting phrases (such as indefinite descriptions) or referring terms (such as demonstratives or indexicals).
- 2) Whether divergences between philosophical orthodoxy and common sense intuitions about the semantics of other linguistic terms is best explained by philosophical expertise—'philosophers have *better* intuitions than non-philosophers because they have the professional skills required to determine the semantic value of terms in language'—or by a philosophical bullying phenomenon—'philosophers have *worse* intuitions than non-philosophers because they are unduly influenced and biased to favour the intuitions, whatever they happen to be, of those few dominant authorities in the field'. Indeed, while no evidence was found to support a philosophical bullying hypothesis with regard to natural kind terms in this thesis, it might be the case that intuitions are biased in this way with respect to other linguistic terms. Good areas to investigate would be ones where there seems to be a discrepancy between the intuitions of philosophers and the intuitions of non-philosophers.
- 3) Relatedly, whether there are any relevant differences between philosophers and non-philosophers prior to formally learning philosophy (i.e., at university). Livengood et al. (2010) suggested that rather than looking at what philosophers do that distinguishes them from non-philosophers, we should look at how they differ in temperament. In particular, they argue that, based on the results of the Cognitive Reflectivity Test, philosophers are particularly reflective in their thinking. This introduces the further question of whether philosophers' begin more reflective, and are therefore drawn towards the study of philosophy, or whether philosophers do not start out more reflective but instead learn to be reflective through the course of their study. The answer to this question will tell us something about the kind of person who becomes a philosopher. The present investigation went some way into answering this question as in Chapter 3 and Chapter 4 both first year philosophy

student (i.e., people drawn to philosophy) and first year non-philosophy students were given the CRT as well as philosophical questions. No difference was uncovered between the CRT scores of the two groups, or between their philosophical intuitions about either proper names or natural kind terms. This suggests that philosophers become more reflective through the course of their study. However, in both studies the philosophy students did score higher than the non-philosophy students on the CRT on average, this difference just did not reach the level of significance. This might suggest some degree of greater reflective thinking. Future research could focus more closely on this question and possibly provide more cognitive and personality tests to investigate any differences between the two groups.

4) A greater emphasis could be made to experimentally investigate our intuitions about actual and counterfactual cases. This would assist us in interpreting the evidence outlined in this thesis, and in particular, would help determine whether a pluralist or an ambiguity thesis better explains the data collected. Previous research suggests that people have different intuitions depending on whether they are considering the actual world or an alternative world, for example, Roskies and Nichols (2008) look at intuitions about moral responsibility in possible situations that vary with respect to how deterministic the universe is. Chapter 4 of this thesis investigated intuitions about the semantics of natural kind terms, and did not find any significant difference between participants considering cases set on Earth and participants considering cases set on Twin Earth. The vignettes, however, did not clearly ask participants to consider worlds and universes counterfactually. The 'Actual Earth' cases required participants to consider an epistemic possibility that the actual earth is different. While, the 'Twin Earth' cases *could be* interpreted as thought experiments asking participants to consider alternative possible worlds, they could just as easily be interpreted as thought experiments asking participants to consider an epistemic possibility about the actual universe (an epistemic possibility in which there is a planet interestingly related to Earth). I think that further research should investigate whether intuitions vary as the grammatical emphasis of the 'Twin Earth' vignette varies. In particular, this should be done while keeping a two-dimensional semantics in mind. For instance, we could look at

whether participants tend to evaluate cases differently depending on whether the Twin Earth cases are presented in the indicative or the subjunctive mood. One way this might be done is to present participants with something like one of the following vignettes (along with similar vignettes presented in the same ways):

#### Universe-Considered-As-Actual Case (Epistemic Possibility)

Imagine that somewhere in <u>our</u> universe there <u>is</u> in fact a planet almost exactly like Earth called Twin Earth, but with one tiny difference: the colourless, odourless liquid that falls from the sky and fills the oceans, lakes and seas on Twin Earth is not H2O. Instead, imagine that this stuff <u>in fact</u> <u>has</u> a different chemical compound, XYZ. Indeed, suppose XYZ <u>is</u> indistinguishable from H2O at normal temperatures and pressures on Twin Earth. Imagine next that there <u>are in fact</u> lots of English speaking visitors to Twin Earth from Earth. Suppose <u>they all in fact</u> enjoy quenching their thirst by drinking XYZ.

If Oscar, one of the English speaking visitors to Twin Earth, <u>makes</u> the statement "there is water on Twin Earth," <u>is</u> his statement literally true?

#### Universe-Considered-As-Counterfactual Case (Metaphysical Possibility)

<u>Although our universe is not in fact like this</u>, imagine what it <u>would be like if</u> <u>there were</u> a planet almost exactly like Earth called Twin Earth, but with one tiny difference: the colourless, odourless liquid that falls from the sky and fills the oceans, lakes and seas on Twin Earth is not H2O. Instead, imagine that this stuff <u>had</u> a different chemical compound, XYZ. Indeed, suppose XYZ <u>were</u> indistinguishable from H2O at normal temperatures and pressures on Twin Earth. Suppose that Oscar, an inhabitant of Twin was to enjoy quenching his thirst by drinking XYZ. If Oscar <u>was to make</u> the statement "there is water on Twin Earth," <u>would</u> his statement be literally true? Two-dimensional theories, such as neo-descriptivism, predict that native speakers of English should answer 'yes' to the first case, but 'no' to the second. The ambiguity hypothesis would be undermined to the extent that experimental results confirmed this prediction, and supported to the extent that experimental results deviated from this prediction.

5) I think the most important future avenue is one that moves away from collecting data about our intuitions from surveys and instead employs other methods such as reaction times and brain imaging. I have argued that the Expertise Objection does not challenge the work of experimental philosophy. However, it is the case that surveying intuitions can introduce a lot of noisy data that needs to be explained as participant error. Moreover, formulating surveys is hard and can introduce a lot of biases. By employing these other methods we may be able to supplement the findings of survey studies and, moreover, further quell the Expertise Objection. Notably, Greene et al. (2001), Greene et al. (2004), and other studies have already employed brain imaging methods to investigate moral judgements. In the domain of the philosophy of language, as we saw in Chapter 2, Schwarz (2015) has used reaction times in his study of definite descriptions. Other such studies could be conducted about the meanings of our terms and phrases. In particular I think that a study that tested how fast participants make 'false' responses compared to 'neither true nor false' responses to sentences containing empty definite descriptions would be able to inform on the debate about definite descriptions. This is because if participants made significantly faster 'neither true nor false' responses we might take this as evidence that the Frege-Strawson Theory gives us the most common reading of definite descriptions. According to Devitt (2004) and Reimer (1998) this would give us good reason to consider the Frege-Strawson Theory to give a genuine semantic reading. With regard to proper names and natural kind terms, it would be interesting to investigate whether one reading was consistently made faster than another reading; for example, whether Externalist readings were made significantly faster than Internalist readings. This could tell us whether one reading, like Externalism, was a more natural or more commonly used reading because it is able to be made more easily than another reading, like Internalism. Moreover, it might

be that a cross-cultural difference could be uncovered with this kind of test. Indeed, in the natural kind terms study no cross-cultural differences were uncovered. However, it might be that while there is no difference in final intuitions, Chinese participants make Internalist responses faster than Externalist responses and New Zealand participants make Externalist responses faster than Internalist responses. Thus, a time trial might be able to uncover new information that is not available from the present studies.

6) Lastly, a related method would be to move away from collecting data about independent naive intuitions. Two ways that this could be done is, firstly, we could look at linguistic behaviour rather than intuitions from linguistic corpus data. To the extent that there is evidence of ordinary language users disambiguating these terms and phrases of interest in different ways there would be evidence in favour of the Ambiguity Thesis. For example, if we found evidence that ordinary language users regularly use the term 'gold' to mean 'yellow metal' and, in other instances, 'metal with the atomic number 79', this would be supporting evidence for the Ambiguity Thesis.<sup>106</sup> Secondly, we could focus groups to see what ordinary language users would say after discussing the questions among themselves. In this scenario we would not get individual intuitions. Instead we would get a judgement based on the amalgamated opinions of a group who can challenge each other which might lead to more reflected on intuitions and possibly a convergence of opinion. There is also a strong possibility that the responses in these groups would be based on a large amount of philosophical bullying. While both of these suggestions are problematic, I think they could yield interesting subsidiary evidence to surveys.

Indeed, there are many new opportunities for philosophers to explore that will be able to inform on traditional questions from the philosophy of language domain. Certainly, I take the Ambiguity Thesis to give an answer to some of these important questions.

<sup>&</sup>lt;sup>106</sup> I say regularly because otherwise we would not consider it to be a default meaning and could therefore be an instance of speaker's meaning.

# References

- Abrusán, M., & Szendröi, K. (2013). Experimenting with the King of France: Topics, verifiability and definite descriptions. *Semantics and Pragmatics*, *6*, 1-43.
- Ackerman, B. P. (1979). Children's understanding of definite descriptions: pragmatic inferences to the speaker's intent. *Journal of Experimental Child Psychology*, *28*(1), 1–15.
- Adams, F., Fuller, G., & Stecker, R. (1997). The Semantics of Fictional Names. *Pacific Philosophical Quarterly*, *78*(2), 128–148.

Alexander, J. (2012). Experimental philosophy: An introduction. Cambridge: Polity Press.

- Alexander, J., Mallon, R., & Weinberg, J. M. (2010). Accentuate the negative. *Review of Philosophy and Psychology, 1* (2), 297-314.
- Alexander, J., & Weinberg, J. M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass, 2*, 56–80.

Amaral, F. S. (2008). Definite descriptions are ambiguous. Analysis, 68 (300), 288–297.

- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.) *Groups, leadership and men*. Pittsburgh, PA: Carnegie Press.
- Atlas, J. (2004). Descriptions, linguistic topic/comment, and negative existentials: A case study in the application of linguistic theory to problems in the philosophy of. In
   Marga Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 342—360).
   Oxford: Oxford University Press.

Bach, K. (1981). Referential/attributive. Synthese, 49 (2), 219-244.

Bach, K. (1987). Thought and reference. Oxford: Oxford University Press.

- Baz, A. (2012). When words are called for: a defense of ordinary language philosophy. Havard: Harvard University Press.
- Bealer, G. (2002). Modal epistemology and the rationalist renaissance. InT. Szabo-Gendler &
  J. Hawthorne (Eds.), *Conceivability and possibility* )pp. 71-125). Oxford: Oxford
  University Press.
- Besson, C. (2012). Empty Natural Kind Terms and Dry-Earth. Erkenntnis, 76(3), 403–425.
- Bird, A., & Tobin, E. "Natural Kinds", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <a href="http://plato.stanford.edu/archives/spr2015/entries/natural-kinds/">http://plato.stanford.edu/archives/spr2015/entries/natural-kinds/</a>.
- Blackburn, W. (1988). Wettstein on definite descriptions. *Philosophical Studies*, 53<u>(2)</u>, 263-278.
- Boghossian, P. A. (1998). What the Externalist Can Know A Priori. *Philosophical Issues, 9*, 197-211.
- Bourget, D., & Chalmers, J. (2014). What do philosophers believe? *Philosophical Studies*, 170 (3), 465-500.
- Braisby, N., Franks, B., & Hampton, J. (1996). Essentialism, word use, and concepts. *Cognition, 59*, 247-274.
- Braun, D. (1993). Empty names. Noûs, 27 (4), 449-469.
- Brock, S. (2004). The ubiquitous problem of empty names. *The Journal of Philosophy, 101* (6), 277-298.
- Brock, S. (2014). The phenomenological objection to fictionalism. *Philosophy and Phenomenological Research, 88* (3), 574–592.
- Burgess, J. (2013). Kripke. Cambridge: Polity Press.
- Burgess, J. (1996). Marcus, kripke, and names. Philosophical Studies, 84, 1-47.

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

- Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*, 284-299.
- Chalmers, D. J. (2006). Two-Dimensional Semantics. In *Oxford Handbook of the Philosophy of Language*. Oxford: Oxford University Press.
- Chemla, E., & Bott, L. (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes, 28* (3), 241-260.
- Cohen, J. (1980). Notes on Kripke and the Description Theory Kripk's Arguments, 1–7.
- Cohnitz, D., & Haukioja, J. (2012). Meta-Externalism vs Meta-Internalism in The Study of Reference. *Australasian Journal of Philosophy*, *91* (3), 475-500..
- Copestake, A., & Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal* of Semantics, 12(1), 15–67.
- Crane, T. (1991). All the Difference in the World. *The Philosophical Quarterly, 41,* 1–25.
- Culbertson, J., & Gross, S. (2009). Are linguists better subjects? *The British Journal for the Philosophy of Science, 60* (4), 721-736.
- Cullen, S. (2010). Survey-Driven Romanticism. *Review of Philosophy and Psychology*, 1 (2), 275–296.
- Deutsch, M. (2009). Experimental philosophy and the theory of reference. *Mind & Language, 24* (4), 445-466.
- Devitt, M. (1974). Singular terms. The Journal of Philosophy, 71 (7), 183-205.
- Devitt, M. (1981). Designation. Columbia: Columbia University Press.
- Devitt, M. (1996). *Coming to our senses: A naturalistic program for semantic localism*. Cambridge: Cambridge University Press.

- Devitt, M. (2004). The Case for Referential Descriptions. In Marga Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 280–305). Oxford: Oxford University Press.
- Devitt, M. (2007). Referential descriptions and conversational implicatures. *European Journal* of Analytic Philosophy, 3 (2), 7-32.
- Devitt, M. (2011). Experimental semantics. *Philosophy and Phenomenological Research*, *82*(2), 418-435.
- Devitt, M. (2012). Whither experimental semantics? *Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia*, 27 (1), 5–36.
- Devitt, M., & Sterelny, K. (1999). *Language and Reality* (2nd Ed). Oxford: Blackwell Publishers.
- Donnellan, K. S. (1966). Reference and Definite Descriptions. *Philosophical Review*, 75 (3), 281-304.
- Dummett, M. (1981). Frege: Philosophy of language. Cambridge: Cambridge University Press.

Elbourne, P. (2005). Situations and individuals. Cambridge, MA: MIT Press.

Elbourne, P. (2013). Definite descriptions. Oxford: Oxford University Press.

- Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine : Journal of the Association of American Medical Colleges, 79* (10), 70–81.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100* (3), 363–406.

Feltz, A. (2009). Experimental Philosophy. Analyse & Kritik, 2, 201-219.

- Feltz, A., & Cokely, E. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*, 18, 342–350.
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19* (4), 25-42.
- Frege, G. (1892). Über Sinn und Bedeutung. Zeitschrift Für Philosophie Und Philosophische Kritik, 100, 25–50.
- Frege, G. (1893). Grundgesetze der Arithmetik. Jena: Hermann Pohle.
- Gelman, S. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65-95.
- Gelman, S. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8* (9), 404-409.
- Gelman, S. A., & Markman, E. M. (1986) Categories and induction in young children. *Cognition, 23*, 183-209.
- Gelman, S. A., & Markman, E. M. (1987) Young children's inductions from natural kinds: the role of categories and appearances. *Child Devolpment, 58*, 1532–1541.
- Genone, J. (2012). Theories of Reference and Experimental Philosophy. *Philosophy Compass,* 7 (2), 152-163.
- Genone, J., & Lombrozo, T. (2012). Concept possession, experimental semantics, and hybrid theories of reference. *Philosophical Psychology*, *25* (5), 717-742.
- Gray, R. (2006). Natural phenomenon terms. Analysis, 66 (2), 141-148.
- Greene, J., Nystrom, L., & Engell, A. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44* (2), 389-400.

- Greene, J., & Paxton, J. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America, 106* (30), 12506–12511.
- Gundel, J. (1977). *The Role of Topic and Comment in Linguistic Theory*. Bloomington, Indiana: Indiana University Linguistics Club.
- Haduong, P., Horn, L., & Knobe, J. (2013). *Mind the (Truth-Value ) Gap* (Unpublished dissertation). Haven, Connecticut: Yale University.
- Hales, S. (2006). *Relativism and the foundations of philosophy*. Cambridge, MA: MIT Press.
- Hanfling, O. (2000). *Philosophy and ordinary language: the bent and genius of our tongue*. New York: Routledge.
- Hansen, N. (2014). Contemporary Ordinary Language Philosophy, *Philosophy Compass, 9* (8), 556–569.
- Haukioja, J. (2015). Advances in Experimental Philosophy of Language. Bloomsbury Publishing.
- Heim, I. (1990). E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, *13*, 137–77.

Heim, I., & Kratzer, A. (1998). Semantics in generative grammar. Oxford: Blackwell.

Hofmann, F. (2010). Intuitions, concepts, and imagination. *Philosophical Psychology*, 23 (4), 529-546.

Horn, L. (1989). A natural history of negation. Chicago: University of Chicago Press.

- Horvath, J. (2010). How (not) to react to experimental philosophy. *Philosophical Psychology*, 23 (4), 447–480.
- Ichikawa, J., Maitra, I., & Weatherson, B. (2011). In defense of a Kripkean dogma. *Philosophy and Phenomenological Research, 82*, 418–435.

- Jackman, H. (2009). Semantic intuitions, conceptual analysis, and cross-cultural variation. *Philosophical Studies*, *146* (2), 159–177.
- Jackson, F. (1998a). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, F. (1998b). Reference and description revisited. *Philosophical Perspectives, 12,* 201-218.

Jackson, F. (2011). Language, names, and information. Oxford: Blackwell Publishing.

Jaszczolt, K. M., "Defaults in Semantics and Pragmatics", The Stanford Encyclopedia of Philosophy (Winter 2014 Edition), Edward N. Zalta (ed.), URL = <a href="http://plato.stanford.edu/archives/win2014/entries/defaults-semantics-pragmatics/">http://plato.stanford.edu/archives/win2014/entries/defaults-semantics-pragmatics/</a>>.

- Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. British Journal of Psychology, 63 (3), 395-400.
- Jylkkä, J., Railo, H., & Haukioja, J. (2009). Psychological essentialism and semantic externalism: Evidence for externalism in lay speakers' language use. *Philosophical Psychology*, 22 (1), 37–60.
- Kaiser, M., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory and Cognition, 14* (4), 308-312.
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), Themes from Kaplan (pp. 481-563). Oxford: Oxford University Press.
- Katz, J. J. (1990). Has the description theory of names been refuted? In G. Boolos (Ed.)
   *Meaning and Method: Essays in Honour of Hilary Putnam* (pp. 31–61). Cambridge:
   Cambridge University Press.
- Katz, J. J. (1994). Names without bearers. Philosophical Review, 103, 1-39.

- Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action, 10* (2), 95-118.
- Keil, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A.
   Marras (Eds.), Language Learning and Concept Acquisition (pp. 133-153). Norwood,
   New Jersey: Ablex.
- Keller, S. (2011). Social Psychology and Philosophy: Problems in Translation. *Nous, 45* (4), 776–791.
- Knobe, J. (2007). Experimental philosophy and philosophical significance. *Philosophical Explorations*, *10* (2), 119-121.
- Knobe, J., & Nichols, S. (2007). An experimental philosophy manifesto. In J. Knobe & S.Nichols (Eds.), *Experimental Philosophy* (pp. 3-14). Oxford: Oxford University Press.
- Knobe, J., & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, *126*(1), 72–86.
- Kornblith, H. (2010). What reflective endorsement cannot do. *Philosophy and Phenomenological Research, 80*, 1-19.

Kripke, S. (1972/1980). Naming and Necessity. Cambridge MA: Harvard University Press.

- Kripke, S. (1977). Speaker's reference and semantic reference. *Midwest Studies In Philosophy*, *2*, 255–276.
- Kripke, S. (2013). *Reference and Existence: The John Locke Lectures*. New York: Oxford University Press.

Kroon, F. (1987). Causal-descriptivism. Australasian Journal of Philosophy, 65 (1), 1-17.

Lam, B. (2010). Are Cantonese-speakers really descriptivists? Revisiting cross-cultural semantics. *Cognition*, *115* (2), 320–329.

- Lasersohn, P. (1993). Existence presuppositions and background knowledge. *Journal of Semantics*, *10* (2), 113–122.
- Lau, J., & Deutsch, M. "Externalism About Mental Content", The Stanford Encyclopedia of Philosophy (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <a href="http://plato.stanford.edu/archives/sum2014/entries/content-externalism/">http://plato.stanford.edu/archives/sum2014/entries/content-externalism/</a>>.
- Livengood, J., Sytsma, J., Feltz, A., Scheines, R., & Machery, E. (2010). Philosophical temperament. *Philosophical Psychology*, *23* (3), 313–330.
- Locke, J. (1894). *An Essay Concerning Human Understanding*. Oxford, Oxfordshire : Clarendon Press.
- Loftus, E. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7 (4), 560–572.

Ludlow, P. "Descriptions", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/descriptions/>.Ludlow, P., & Neale, S. (1991). Indefinite descriptions: In defense of Russell. *Linguistics and Philosophy*.

- Ludlow, P., & Segal, G. (2004). On a unitary semantical analysis for definite and indefinite descriptions. In Marga Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 420-436). Oxford: Oxford University Press.
- Ludwig, K. (2007). The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy, Philosophy and the Empirical, 31* (1), 128–159.

Ludwig, K. (2010). Intuitions and relativity. *Philosophical Psychology*, 23 (4), 427-445.

Lutz, S. (2009). Ideal Language Philosophy and Experiments on Intuitions. *Studia Philosophica Estonica*, 2 (2), 117–139.

Lycan, W. (Unpublished manuscript). Metaphysics and the paronymy of names.

- Machery, E. (2012). Expertise and intuitions about reference. *Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia*, 27 (1), 37–54.
- Machery, E. (2014). Concepts: Investigating the heterogeneity hypothesis. *Advances in Experimental Philosophy of Mind* (203-232). London: Bloomsbury Academic.
- Machery, E., Deutsch, M., Mallon, R., Nichols, S., Sytsma, J., & Stich, S. P. (2010). Semantic intuitions: Reply to Lam. *Cognition*, *117* (3), 361–366.
- Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, *92* (3), B1-B12.
- Machery, E., Olivola, C., & Blanc, M. De. (2009). Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis*, *69* (4), 689–694.
- Machery, E., & Seppälä, S. (2010). Against hybrid theories of concepts. *Anthropology & Philosophy*, *10*, 97-125.
- Machery, E., Sytsma, J., & Deutsch, M. (2015). Speaker's Reference and Cross-cultural Semantics. In A. Bianchi (Ed.), *On Reference* (pp. 62-76). Oxford: Oxford University Press.
- Marconi, D. (2005). Two-dimensional semantics and the articulation problem. *Synthese*, *143* (3), 321-349.
- Martí, G. (2009). Against semantic multi-culturalism. Analysis, 69, 42-8.
- McCloskey, M. (1983). Intuitive physics. Scientific American, 248 (4), 122-130.
- McKay, T., & Nelson, M. "Propositional Attitude Reports", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <a href="http://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/">http://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/</a>>.
- Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44* (12), 1469-1481.

- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). New York: Cambridge University Press.
- Mill, J. S. A system of logic. New York: Longmans, Green, 1947. (Originally published, 1843.)
- Mizrahi, M. (2014). Philosophical Expertise? *Experimental Philosophy*. Retrieved June 23, 2015, from http://philosophycommons.typepad.com/xphi/2014/10/philosophical-expertise.html
- Mizrahi, M. (2015). Three arguments against the expertise defense. *Metaphilosophy*, 46, 52-64.
- Moore, J. S. (1924). The meaning of meaning. Mind.
- Murday, B. (2014). Definite Descriptions and Semantic Pluralism. *Philosophical Papers*, 43 (2), 255–284.
- Nadelhoffer, T., & Feltz, A. (2008). The actor–observer bias and moral intuitions: adding fuel to Sinnott-Armstrong's fire. *Neuroethics*, *1* (2), 133-144.
- Nadelhoffer, T., Kvaran, T., & Nahmias, E. (2009). Temperament and intuition: A commentary on Feltz and Cokely. *Consciousness and Cognition, 18*, 351-355.
- Nadelhoffer, T., & Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations, 10* (2), 123-149.
- Nado, J. (2014a). Philosophical Expertise. Philosophy Compass, 9, 631–641.
- Nado, J. (2014b). Philosophical expertise and scientific expertise. *Philosophical Psychology*, 1–19.
- Naess, A. (1953). Interpretation and preciseness: A contribution to the theory of communication. Dordrecht, The Netherlands: Springer.
- Naess, A. (1966). Communication and argument. London: Allen & Unwin.

Neale, S. (1990). Descriptions. Cambridge, MA: MIT Press.

Neale, S. (2004). *This, that, and the other*. In Marga Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 68-182). Oxford: Oxford University Press.

Neale, S. (2005). A century later. Mind, 114 (456), 809-871.

Nelson, M. (2002). Descriptivism defended. *Noûs, 36* (3), 408-436.

Nichols, S., Pinillos, A., & Mallon, R. (Forthcoming). Ambiguous reference. *Mind*.

Nimtz, C. (2004). Two-dimensional and natural kind terms. Synthese, 138, 125-148.

- Nisbett, R., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, *108*, 291-310.
- Olivola, C. & Machery, E. (2014). Is psychological essentialism an inherent feature of human cognition? *Behavioral and Brain Sciences*, *37*, 499.
- Oppy, G. (1997). Discussion. The philosophical insignificance of Gödel's slingshot. *Mind, 106,* 121-142.
- Paxton, J., Ungar, L., & Greene, J. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*, 163–177.
- Pethő, G. (2005). On intuitions about proper names. *Acta Linguistica Hungarica, 52*, 281-301.
- Phillips, J. K., Klein, G., & Sieck, W. R. (2008). Expertise in Judgment and Decision Making: A
   Case for Training Intuitive Decision Skills. *Blackwell Handbook of Judgment and Decision Making*. Malden, MA: Blackwell Publishing.

Pigliucci, M. (2011). One more take on the role of intuition in philosophy. *Rationally Speaking*. Retrieved June 23, 2015, from http://rationallyspeaking.blogspot.co.nz/2011/11/one-more-take-on-role-of-intuition-in.html. Pinillos, N., Smith, N., & Nair, G. (2011). Philosophy's new challenge: experiments and intentional action. *Mind & Language, 26*, 115-139.

Putnam, H. (1973). Meaning and reference. The Journal of Philosophy, 70 (19), 699-711

Putnam, H. (1975). The meaning of 'meaning'. In *Philosophical Papers, Vol. 2: Mind, Language and Reality*. Cambridge University Press.

Quine, W. (1960). Word and object. Cambridge, MA: MIT Press.

- Raatikainen, P. (2006). Against causal descriptivism. Mind and Society, 5 (1), 78-84.
- Recanati, F. (2003). Literalism and contextualism: Some varieties. In G. Preyer (Ed.) *Contextualism*. Oxford: Oxford University Press

Recanati, F. (2004). Literal meaning. Cambridge: Cambridge University Press.

Reimer, M. (1998). Donnellan's distinction/Kripke's test. Analysis, 58 (2), 89-100.

- Reimer, M., & Bezuidenhout, A. (2004). Presuppositions and Truth-value gaps. In Marga
   Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 307-314). Oxford:
   Oxford University Press.
- Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics in Pragmatics and Philosophy I. *Philosophica Anc Studia Philosophica Gandensia Gent, 27*, 53-94.
- Richard, M. (1987). Attitude ascriptions, semantic theory, and pragmatic evidence. *Proceedings of the Aristotelian Society*, *87*, 243-62.
- Roberts, L. (1984). Ambiguity vs. generality: removal of a logical confusion. *Canadian Journal of Philosophy, 14* (2), 295-313.
- Roskies, A., & Nichols, S. (2008). Bringing moral responsibility down to earth. *Journal of Philosophy, 105* (7), 371-388.

Russell, B. (1905). On Denoting. Mind, 14 (56), 479-493.

Russell, B. (1957). Mr. Strawson on referring. Mind, 66 (263), 385-389.

Russell, B. (1996). My philosophical development. New York: Routledge.

Salmon, N. (1986). Frege's puzzle. Atascadero, CA: Ridgeview.

Salmon, N. (1988). Reflexivity. In N. Salmon & S. Soames (Eds.), *Propositions and Attitudes* (pp. 240-274). Oxford: Oxford University Press.

Salmon, N. (1991). The pragmatic fallacy. *Philosophical Studies, 63*, 83-97.

- Schoubye, A. (2009). Descriptions, truth value intuitions, and questions. *Linguistics and Philosophy, 32* (6), 583-617
- Schulz, E., Cokely, E., & Feltz, A. (2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition, 20* (4), 1722–1731.
- Schwarz, F. (2015). False but slow: Evaluating statements with non-referring definites. Journal of Semantics.
- Schwitzgebel, E. (2009). Do ethicists steal more books?. *Philosophical Psychology*, 22(6), 711-725.
- Schwitzgebel, E., & Cushman, F. (Unpublished manuscript). Professional Philosophers' Susceptibility to Order Effects and Framing Effects in Evaluating Moral Dilemmas.
- Schwitzgebel, E., & Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind and Language*, *27* (2), 135–153.
- Searle, J. R. (1978). Literal meaning. Erkenntnis, 13(1), 207-224.
- Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. R. (1985). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Seegar, M. (2010). Experimental philosophy and the twin earth intuition. *Grazer Philosophische Studien, 80,* 237-244.

Segal, G. (2000). A slim book about narrow content. Cambridge, Massachusetts: MIT Press.

- Sennet, A. (2002). An ambiguity test for definite descriptions. *Philosophical studies*, 111(1), 81-95.
- Sennet, A. "Ambiguity", The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <a href="http://plato.stanford.edu/archives/spr2015/entries/ambiguity/">http://plato.stanford.edu/archives/spr2015/entries/ambiguity/</a>.
- Sinnott-Armstrong, W. (2007). Abstract + concrete = paradox. In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 209-230). Oxford: Oxford University Press.
- Sinnott-Armstrong, W. (2008). Framing moral intuitions. In S. Walter (Ed) *Moral psychology, Vol 2: The cognitive science of morality: Intuition and diversity* (pp. 47-76). Cambridge, MA: MIT Press.
- Smith, A. (2005). Natural Kind Terms: A Neo-Lockean Theory. *European Journal of Philosophy, 13,* 70–88.
- Soames, S. (1989). Direct reference, propositional attitudes and semantic content. *Philosophical Topics*, *15*, 44-87.
- Soames, S. (1998). The modal argument: Wide scope and rigidified descriptions. *Noûs, 32*, 1–22.
- Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.
- Soames, S. (2009). *Philosophical analysis in the twentieth century, volume 2: The age of meaning* (Vol. 2). NJ: Princeton University Press.
- Sosa, D. (2001). Rigidity in the Scope of Russell's Theory. Noûs, 35 (1), 1-38.

- Sosa, E. (2007). Experimental philosophy and philosophical intuition. *Philosophical Studies*, *132* (1), 99–107.
- Stanley, J., & Szabó, Z. G. (2000). On quantifier domain restriction. *Mind and language*, 15 (2 & 3), 219-261.
- Stich, S., & Tobia, K. P. (2015). Experimental Philosophy and the Philosophical Tradition.
- Strawson, P. F. (1950). On referring. Mind, 320-344.
- Strawson, P. F. (1971). Logico-linguistic papers (Vol. 516). London: Methuen.
- Sytsma, J., & Livengood, J. (2011). A new perspective concerning experiments on semantic intuitions. *Australasian Journal of Philosophy*, *89* (2), 315-332.
- Sytsma, J., & Livengood, J. (2012). Experimental Philosophy and philosophical disputes. Essays in Philosophy, 13, 145-160.
- Sytsma, J., Livengood, J., Sato, R., & Oguchi, M. (2015). Reference in the Land of the Rising Sun: A Cross-cultural Study on the Reference of Proper Names. *Review of Philosophy and Psychology*, 6 (2), 213-230.
- Tennessen, H. (1960). Vindication of the Humpty Dumpty attitude towards language: A rejoinder to Anfinn Stigen and other critics. *Inquiry: An Interdisciplinary Journal of Philosophy, 3*, 185-198.
- Tennessen, H. (1959). What should we say? Inquiry, 2 (1-4), 265-290.
- Tetlock, P. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.
- Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts?. *Philosophical Psychology*, *26*(5), 629-638.
- Tobia, K. P., Chapman, G. B., & Stich, S. (2013). Cleanliness is next to morality, even for philosophers. *Journal of Consciousness Studies*, *20*(11-12).

Turri, J. (2015). An Open and Shut Case: Epistemic Closure in the Manifest Image. Ann Arbor,MI: Michigan Publishing, University of Michigan Library.

Villanueva, E., & Boghossian, P. A. (2013). Externalist Priori \* Can, 9(1998), 197–211.

- Von Fintel, K. (2004). Would you believe it? The King of France is back! In Marga Reimer & Anne Bezuidenhout (Eds.), *Descriptions and Beyond* (pp.315-341). Oxford: Oxford University Press.
- Weatherson, B. (2007). Doing philosophy with words. *Philosophical Studies*, *135* (3), 429–437.
- Weinberg, J. M., Gonnerman, C., Buckner, C., & Alexander, J. (2010). Are philosophers expert intuiters? *Philosophical Psychology*, *23* (3), 331–355.
- Weinberg, J. M., Weinberg, J. M., Nichols, S., Nichols, S., Stich, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, *29* (1 & 2), 429–460.
- Wettstein, H. K. (1981). Demonstrative reference and definite descriptions. *Philosophical Studies*, *40*(2), 241-257.
- Wikforss, Å. (2008). Semantic externalism and psychological externalism. *Philosophy Compass, 3*, 158-181.

Williamson, T. (2008). The philosophy of philosophy. Hoboken, NJ: John Wiley & Sons.

- Williamson, T. (2011). Philosophical expertise and the burden of proof. *Metaphilosophy*, 42 (3), 215-229.
- Zacharska, B. (n.d.). Definite descriptions: Semantic or pragmatic polysemy?. Citeseer.
- Zwicky, A., & Sadock, J. (1975). Ambiguity tests and how to fail them. Syntax and semantics, 4 (1), 1-36.
- Zwicky, A. M., & Sadock, J. M. (1987). A Non-test for Ambiguity. *Canadian Journal of Philosophy*, *17* (1), 185-187.

# **Appendix A**

#### Chinese translations of Frege's case and the Cat case

### Frege's case:

古时候,希腊人为晚上肉眼可见的第一颗明亮的天体取了个名字,叫做"黄昏星",又完全另 外地为清晨肉眼可见的最后一颗明亮的天体取了个名字,叫做"晨星"。事实上,古希腊人认 为黄昏星跟晨星是不同的星星。直到很多年以后,天文学家才发现黄昏星跟晨星根本不是星星 ,而是同一个行星,即金星。假设一个古希腊人阿戈斯说"黄昏星永远明亮",而另一个古希 腊人阿多尼亚说"晨星永远明亮"。

你认为阿戈斯跟阿多尼亚说的话指的是否是一回事?请从以下选项中选择你认为最有可能是正确答案的选项。

- (A) 是,他们指的是一回事。
- (B) 不是,他们指的不是一回事。

### The Cat case:

假设有一天,约翰发现了一只流浪猫,并决定收养它。因为这只猫长得很可爱,所以约翰给它 起名为"基蒂"。约翰告诉他的朋友说他收养了一只叫做基蒂的宠物,并且说"基蒂是一只猫 "。现在假设,当约翰不在家的时候,约翰的猫往往会四处闲逛。约翰的猫常常进入简的房子 ,而简很快就喜欢上了这只猫。她不知道这只猫是约翰的。有一个冬天,简自行决定给约翰的 猫取名为"基蒂",单纯只是因为她喜欢这个名字。之后她跟家人说起基蒂,并且说"基蒂是 一只猫"。

你认为约翰与简说的"基蒂是一只猫"指的是否是一回事?请从以下选项中选择你认为最有可 能是正确答案的选项。

(A) 是,他们指的是一回事。

(B) 不是,他们指的不是一回事。
# **Appendix B**

## **Chinese Translations**

### Gold case:

假设现在是1860年,远在亨利·莫塞莱发明原子序数的概念之前。威廉是一名美国勘探者,在 加里福利亚南部的圣贝纳迪诺山脉寻找贵重金属。当他在一个山谷里勘探的时候,发现了一大 块闪烁的沉重的黄色金属。威廉欢呼起来: "我发现了金子"。威廉发现的矿块的原子序数恰 好为34,而我们今天称之为"金子"的金属的原子序数为79。

当威廉声称"我发现了金子"时,他这句话的字面意思对吗?请选择看上去最正确的选项:

- (1) 对
- (2) 不对

0

(3) 不确定

- (A) -对- 因为威廉用的"金子"这个词跟我们用的"金子"这个词指的是同一个意思;我 们指的都是"闪烁的沉重的黄色金属"。
- (B) -不对- 因为威廉用的"金子"这个词跟我们用的"金子"这个词指的是同一个意思, 我们指的是"原子序数为79"的金属(但威廉对此并不知情)。
- (C) -对- 因为威廉用的"金子"这个词跟我们用的"金子"这个词指的不是同一个意思; 威廉指的是"原子序数为34的物体",而我们指的是"原子序数为79的物体"。
- (D) -不对- 因为威廉用的"金子"这个词跟我们用的"金子"这个词指的是同一个意思, 我们指的都是"科学家 - 化学家及地质学家 - 使用'金子'这个词时所表达的意思"
- (E) 出于其他原因。(请把你的答案及理由写在下面。)

### Tree case:

假设只有专家才能对山毛榉树和榆树加以区别。山毛榉树和榆树有一些共同的重要特征:它们 的常见地在世界的相同区域(即北美)、都能长到大致相同的高度、长着样子相似的叶子、冬 天落叶。但是,山毛榉树属于山毛榉科,而榆树则属于榆科。假设简(像大多数人一样)无法 分辨山毛榉科植物与榆科植物之间的不同。在一个大热天,简进入了一片榆科植物林,她指着 这些树兴奋地说:"它们是山毛榉树。"

当简声称"它们是山毛榉树"时,她的判断对吗?请选择看上去最正确的选项:

- (1) 对
- (2) 不对
- (3) 不确定

- (A) -对- 因为简用的"山毛榉树"这个词跟大多数人用的"山毛榉树"这个词指的是同一 个意思;他们指的都是"北美常见的平均树高为90英尺(27米)、长着椭圆形树叶的 落叶乔木"。
- (B) -不对- 因为简用的"山毛榉树"这个词跟大多数人用的"山毛榉树"这个词指的是同 一个意思;他们指的都是"山毛榉科植物"(而简对此并不知情)。
- (C) -对- 因为简用的"山毛榉树"这个词跟专家们用的"山毛榉树"这个词指的不是同一 个意思;简指的是"榆科植物",而专家们指的则是"山毛榉科植物"。
- (D) -不对- 因为简用的"山毛榉树"这个词跟大多数人用的"山毛榉树"这个词指的是同一个意思;他们指的都是"专家们使用'山毛榉树'这个词时所表达的意思"。
- (E) 出于其他原因。(请把你的答案及理由写在下面。)

#### Water case:

假设有一个叫做孪生地球的星球。孪生地球跟地球几乎一模一样,但是这两个星球之间有一个 细微的差别:在孪生地球,从天空落下来的、充满湖泊海洋的无色无味的液体不是H2O,而是 另一种化合物XYZ。事实上,在常温常压下,XYZ与H2O难以区分。假设现在有一名地球上的游 客奥斯卡到了孪生地球。他一到那里就欢呼起来:"孪生地球上有水。"随后便喝了一杯XYZ 来解渴。

当奥斯卡声称"孪生地球上有水"时,他这句话的字面意思对吗?请选择看上去最正确的选项:

- (1) 对
- (2) 不对
- (3) 不确定

- (A) -对- 因为奥斯卡用的"水"这个词跟孪生地球人用的"水"这个词指的是同一个意思;他们指的都是"从天空落下来的、充满湖泊海洋的无色无味的液体"。
- (B) -不对- 因为奥斯卡使用"水"这个词时指的是"地球上的科学家使用'水'这个词时 所表达的意思"。
- (C) -对- 因为当奥斯卡使用"水"这个词的时候,他讲的是孪生地球人的语言,孪生地球人用的"水"这个词指的是"XYZ"(而奥斯卡对此并不知情)。
- (D) -不对- 因为奥斯卡用的"水"这个词跟孪生地球人用的"水"这个词指的不是同一个 意思;奥斯卡指的是"H2O"而不是"XYZ"。
- (E) 出于其他原因。(请把你的答案及理由写在下面。)

Metal case:

原子序数为13的金属被地球上的专家称为"铝"。原子序数为42的金属被地球上的专家称为 "钼"。这两种金属都很硬,颜色介于灰色与银色之间,但是,地球上很少能找到原子序数为 42的金属。假设除了科学专家以外,无人能区分这两种金属。现在假设有一个叫做孪生地球的 星球。孪生地球跟地球几乎一模一样,但是这两个星球之间有一个细微的差别:在孪生地球 上,原子序数为42的金属是最常见的化学元素之一,而原子序数为13的金属却很少被发现。 所以,孪生地球上大多数的锅具都是由原子序数为42的金属制成。另外,在孪生地球上,原子 序数为13的金属被称为"钼",而原子序数为42的金属被称为"铝"。假设现在有一名地球 上的游客卡尔到了孪生地球,他一到那里便欢呼起来:"孪生地球上有铝制的锅具。"接着便 拿起一个用原子序数为42的金属制成的锅。

当卡尔声称"孪生地球上有铝制的炊事用具"时,他这句话的字面意思对吗?请选择看上去最 正确的选项:

- (1) 对
- (2) 不对
- (3) 不确定

- (A) -对- 因为卡尔用的"铝"这个词跟孪生地球人用的"铝"这个词指的是同一个意思;他们指的都是"制造锅具使用的银色硬金属"。
- (B) -不对- 因为卡尔使用"铝"这个词时指的是"地球上的专家使用'铝'这个词时所表达的意思"。
- (C) -对- 因为当卡尔使用"铝"这个词的时候,他讲的是孪生地球人的语言,孪生地球人用的"铝"这个词指的是"原子序数为42的金属"(而奥斯卡对此并不知情)。
- (D) -不对- 因为卡尔用的"铝"这个词跟孪生地球人用的"铝"这个词指的不是同一个意思;卡尔指的是"原子序数为13的金属"而不是"原子序数为42的金属"。
- (E) 出于其他原因。(请把你的答案及理由写在下面。)