Lingnan University

Digital Commons @ Lingnan University

Lingnan Theses and Dissertations | Theses and Dissertations

5-28-2019

Metaphysical reduction of necessity : a modified account

Pak Him LAI

Follow this and additional works at: https://commons.ln.edu.hk/otd

 Part of the Metaphysics Commons

Recommended Citation

Lai, P. K. (2019). Metaphysical reduction of necessity: A modified account (Master's thesis, Lingnan University, Hong Kong). Retrieved from https://commons.ln.edu.hk/otd/51/

This Thesis is brought to you for free and open access by the Theses and Dissertations at Digital Commons @ Lingnan University. It has been accepted for inclusion in Lingnan Theses and Dissertations by an authorized administrator of Digital Commons @ Lingnan University.

Terms of Use

The copyright of this thesis is owned by its author. Any reproduction, adaptation, distribution or dissemination of this thesis without express authorization is strictly prohibited.

All rights reserved.

METAPHYSICAL REDUCTION OF NECESSITY: A MODIFIED ACCOUNT

LAI PAK HIM

MPHI

LINGNAN UNIVERSITY

2019

METAPHYSICAL REDUCTION OF NECESSITY: A MODIFIED ACCOUNT

by
LAI Pak Him
黎柏謙

A thesis
submitted in partial fulfillment
of the requirements for the Degree of
Master of Philosophy in Philosophy

Lingnan University

2019

ABSTRACT

Metaphysical Reduction of Necessity:
A Modified Account

by

LAI Pak Him

Master of Philosophy

This thesis investigates the metaphysical nature of necessity. My study focuses primarily on the reduction of metaphysical necessity and the question of whether a necessary truth can be reductively defined. Theodore Sider (2011) develops a new reductive account of metaphysical necessity. Unfortunately, the multiple realizability problem posed by Jonathan Schaffer (2013) undermines the credibility of Sider's account. This underlies my motivation to search for a revised Siderian account of necessity. On this basis, I propose a modified version of Sider's account and argue that analytic, natural-kind and micro-reduction truths are necessary truths if and only if they express the same states of affairs as logical truths. Since logical truths are necessary truths, analytic, natural-kind and micro-reduction truths are necessary truths as well.

In this thesis, I will proceed as follows. Chapter 1 introduces Sider's account of necessity. In particular, I focus on his analysis of the necessity of a micro-reduction truth. His analysis is largely constitutive of the notion of a metaphysical semantics and some associated notions. I reconstruct his analysis and articulate his metaphysical semantics and the associated notions. Chapter 2 presents and clarifies the multiple realizability problem of Schaffer. His critique shows that Sider's metaphysical semantics cannot handle multiple realizability. Sider (2013e) refutes this claim by arguing that the multiple realizability problem is a problem for his analysis of the necessity of a micro-reduction truth, not for his metaphysical semantics itself. This is a good starting point to propound a modified account of necessity. Chapter 3 proposes and articulates my modified analysis of the necessity of a micro-reduction, where I show that the necessity of a micro-reduction truth can be reductively defined without appeal to Sider's metaphysical semantics. I argue that analytic, natural-kind and micro-reduction truths are necessary truths if and only if they express the same states of affairs as logical truths. Logical truths are necessary and so are they. After that, I consider a potential objection to my modified account and attempt to respond to it.

DECLARATION

I declare that this is an original work based primarily on my own research, and I warrant that all citations of previous research, published or unpublished, have been duly acknowledged.

_____

(LAI Pak Him)
Date:

CERTIFICATE OF APPROVAL OF THESIS

METAPHYSICAL REDUCTION OF NECESSITY:
A MODIFIED ACCOUNT
by
LAI Pak Him

Master of Philosophy

Panel of Examiners :

SIGNED
(Prof. HAMPTON Mark Andrew) (Chairman)

SIGNED
(Dr GOSWICK Dana Lynne) (External Member)

SIGNED
(Prof. MARSHALL Daniel Graham) (Internal Member)

SIGNED
(Prof. WAXMAN Daniel) (Internal Member)

Supervisor :

Prof. MARSHALL Daniel Graham

Approved for the Senate :

SIGNED

(Prof. MOK Ka Ho Joshua)
Chairman, Postgraduate Studies Committee

28 MAY 2019

Date

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Prof. Dan Marshall. Dan has been actively involved in every step of this thesis. He has always been generous in sharing his novel ideas with me and giving me excellent comments over the whole period of my writing this thesis. He has constantly reminded me of the demand of rigorous and clear argumentation. I was fortunate to have studied metaphysics and worked on this thesis under his supervision. That was the most valuable experience of my studies. I am grateful to him for his patience and care to me. I cannot imagine how much I owe intellectual debts to him. This thesis would have not been completed without his guidance and encouragement. For whatever there is of value in this thesis, he deserves all of the credit.

I am also grateful to many of my teachers, including Dr. Dana Goswick, Prof. Daniel Waxman, Prof. Jiji Zhang, Prof. Paisley Livingston, Prof. Derek Baker, Prof. Yujian Zheng, Prof. Jennifer Nado, Prof. Andrea Sauchelli, among others. Their comments, especially those from the members of the Panel of Examiners, make a substantial contribution to the completion of this thesis.

I also owe my thanks to fellow students and friends, Micheál Lacey, Christopher Atkinson, Alexander Ehmann, Jenny Hung, Haktan Akcin, Yeung Yu, Dongyu Xie, Ezechiel Thibaud, Shamik Chakravarty, David Villena Saldaña, Ka Yan Mok, Kei Yan Leung and Nok Nam Chan, for all valuable and enjoyable discussions over all these years. Special thanks are due to the staff members of my department, in particular Ms. Vinnie Cheung, for all their help.

I would like to thank my family and my aunt for their unconditional love and support, without which I would not be able to achieve my ambition. Finally, I owe many thanks than I can say to Ms. Ching Lam Ho whom I had the good fortune to meet. Her unswerving support and encouragement are of great importance for my entire studies at Lingnan. I sincerely dedicate this thesis to my family, my aunt and especially Ms. Ching Lam Ho in token of gratitude.

INTRODUCTION

I. An Overview of Modality

Consider the following sentences: 'All vixens are female' and 'All emeralds are green'. It is obvious that there are a number of similarities between both sentences. They both describe things in certain ways. They both are of a simple subject-predicate form. They both are true sentences. And so forth. However, they are different in terms of the *modes* or *ways* of truth: necessary and contingent truth. It is *necessarily* true that all vixens are female. It is necessarily true because it could not have been false that all vixens are female. It is *impossible* that there is a male vixen. To say that a vixen is male is like to say that a square is round. This is itself contradictory. By contrast, even though it is true that all emeralds are green, the truth is merely *contingent*. It is contingently true because it *could* have been false that all emeralds are green. It is *possible* that there is a red, purple or white emerald. It is necessarily true that all vixens are female whereas it is contingently true that all emeralds are green. The negation of 'All vixens are female' is impossible whereas the negation of 'All emeralds are green' is possible. This is how philosophers distinguish necessary truths from contingent truths in general.

Consider another pair of sentences: 'Socrates is a human' and 'Socrates is a philosopher'. Some philosophers claim that Socrates is *essentially* human. In other words, the property of being a human is an *essential* (or necessary) property of Socrates. In general, a property *P* of an object *O* is essential iff *O* could not have failed to have *P*. Socrates is essentially human because he could not have failed to have the property of being a human. But although Socrates has a property of being a philosopher, he has this property only *accidentally* (or contingently). For he could have been a poet and written lyric poetry rather than asking questions about moral knowledge. As opposed to an essential property, a property *P* of an object *O* is accidental iff *O* could have failed to have *P*. It is accidently true that Socrates is a philosopher because the property of being a philosopher is an accidental property of Socrates. Intuitively, we might be inclined to think that Socrates would still be Socrates if he could have failed to have the property of being a philosopher, but Socrates would *not* be Socrates if he

could have failed to have the property of being a human.[1] This is how philosophers distinguish essential properties from accidental properties in general.

The above considerations manifest the *de dicto* and *de re* distinction of modality. This is a generally accepted but still controversial distinction. Roughly speaking, a modal sentence is de dicto when it attributes its modal feature to a *whole sentence*, while a modal sentence is de re when it attributes its modal feature to a *particular object*. The contrast can be brought out by the following:

(De dicto) Necessarily, the number of the planets is greater than 7.

(De re) The number of the planets is such that it is necessarily greater than 7.

The first sentence is de dicto since it attributes necessity to the whole sentence—'the number of the planets is greater than seven'. Although the de dicto sentence is true, it is not necessarily true since clearly there might have been 6 planets or even less. The second sentence is de re because it attributes necessity to the object—the number that actually numbers the planets—8. The de re sentence is necessarily true since the number 8 is essentially greater then 7.

There are different strengths of necessity. In general, metaphysicians identify three kinds of necessity. Metaphysical necessity, the subject to explore in this thesis, is the most general kind of necessity. Logical necessity is a subset of metaphysical necessity. All logical necessities are metaphysical necessities but not all metaphysical necessities are logical necessities. For example, the sentence 'Mark Twain is Samuel Clemens' is considered as a metaphysically but not logically necessary truth. Physical necessity is a necessary truth given by the laws of nature. If the laws of nature are metaphysically necessary, then truths of physical necessity are true in all possible situations; but if they are contingent, truths of physical necessity are true in merely some possible situations where the laws of nature apply. In this thesis, my primary concern is the metaphysical kind of necessity.

There are also different strengths of possibility. There are at least two kinds of possibility in general: physical and metaphysical possibility. Physical possibility is a

[1] For an introduction of Essentialism, see Robertson & Atkins (2013), Cameron (2010), Roca-Royes (2011a), Roca-Royes (2011b).

narrow sort of possibility and the boundary of it is given by the laws of nature. The laws of nature tell us that it is *physically impossible* for one to travel faster than the speed of light. But if one believes that the laws of nature are merely contingent, then one would think possibility in a broader sense. According to metaphysical possibility, it is *metaphysically possible* for one to travel faster than the speed of light. It is also metaphysically possible that human beings are immortal. And so forth. The boundary of metaphysical possibility is very broad, though it is unclear and debatable how broad it is. For example, it is unclear to many philosophers whether metaphysical possibility allows the existence of impossibilities, such as a male vixen, a round square, etc.

Let us briefly talk about the philosophical importance of modality. Modality is indispensable in philosophical theorizing for many reasons. First, modal notions are everywhere in philosophy. The notions of disposition, counterfactuals and supervenience are often considered as modal notions. Philosophers provide analyses of these notions in terms of modality. Second, the development of modal logic in the middle of the 20th century has given philosophers reasons to scrutinize the metaphysical nature of necessity and possibility. At the core of the semantics of modal logic is a talk of possible worlds. The notion of possible worlds has been widely dominant in contemporary philosophy and has been continually calling out for explanation. Third, according to the traditional conception of philosophy, the goals of philosophical inquiries differ somewhat from those of empirical inquiries. Empirical science is a study of contingent reality of the world, whereas philosophy makes claims about what is necessarily true. For example, statements of fundamental metaphysics are claims of this sort. They are necessarily true if true and necessarily false if false. Last but not least, the radical revival of rethinking about modality derives largely from Saul Kripke's *Naming and Necessity* (1972). The revolutionary contributions of his lectures compel contemporary philosophers to reconsider the relation between epistemic and metaphysical modality. His and Hilary Putnam's (1975) convincing examples of the necessary *a posteriori* radically undermine the identification of a prioricity with necessity. Quine's (1951) "Two Dogmas of Empiricism" is another famous critique of analyticity that undermines the identification of analyticity with necessity. Putting all together, the reasons we considered suggest that the philosophical importance of modality is far-reaching and wide-ranging.

So far, I have introduced some basis ideas of the notion of modality. I hope that the preceding paragraphs have helped readers grasp the notion of modality.

II. My Plan of this Thesis

Moving on: This thesis explores the reduction of metaphysical necessity. The central questions of this thesis are constitutive of the following. What is a necessary truth? How a necessary truth can be reductively defined? To explore these two questions, here is my plan.

In chapter 1, I examine Sider's reductive account of metaphysical necessity. In particular, I focus on his analysis of the necessity of a micro-reduction truth. Let 'C_0' be a complex predicate which describes the actual microphysical states of New York City in every detail. Given this setting, the sentence 'Every C_0 is a city' is a necessary truth. Sider offers a reductive analysis to explain such a necessity. According to his analysis, the sentence 'Every C_0 is a city' is a necessary truth if and only if it is a logical consequence of a micro-reduction truth. What then is a micro-reduction truth? To answer, he then defines the notion of a micro-reduction truth in a complicated way. His definition is largely constitutive of the notion of a metaphysical semantics and some associated notions. I reconstruct his whole analysis and give a review of the notion of a metaphysical semantics, the notion of 'application axiom' and the notion of 'corresponding to'. I finish chapter 1 by showing how the sentence 'Every C_0 is a city' is logically derived from his definition of a micro-reduction truth.

In chapter 2, I present and clarify the multiple realizability problem of Schaffer. In brief, he argues that Sider's metaphysical semantics cannot handle multiply realizable truths. Since Sider's analysis of the necessity of a micro-reduction truth is mostly constitutive of his metaphysical semantics, something has gone wrong in either one (or both) if Schaffer's critique is successful. But Sider (2013e) argues that the multiple realizability problem is a problem for his analysis of the necessity of a micro-reduction truth, not for his metaphysical semantics itself. This unresolved problem leaves room for improvement in Sider's analysis.

In chapter 3, I develop a modified version of Sider's account of necessity. Given the multiple realizability problem, I show that the necessity of a micro-reduction truth can be reductively defined without appeal to Sider's metaphysical semantics. I argue

that analytic, natural-kind and micro-reduction truths are necessary if and only if they express the same states of affairs as logical truths. Logical truths are necessary and so are they. I understand the notion of 'express' as a semantic relation between sentences and states of affairs, and define 'express' in terms of the notion of a correct interpretation of David Lewis's (1984) reference magnetism. My use of the term 'state of affairs' is simple. A state of affairs is a way things are or a way things are not.[2] After that, I consider a potential objection: the notion of 'fit-with-use' in Lewis's reference magnetism is vague. To reply, I attempt to sketch a simple modification to make sense of the notion.

[2] This conception of states of affairs is suggested by Dan Marshall. I am indebted to him for this.

CHAPTER 1 SIDER'S REDUCTIVE ACCOUNT OF NECESSITY

This chapter introduces Sider's reductive account of necessity. In section 1.1, I outline Sider's reductive account. In section 1.2, I formulate his analysis of the necessity of a micro-reduction truth. In sections 1.3 & 1.4, in order to elucidate his analysis, I articulate the notion of a metaphysical semantics and some relevant notions. In section 1.5, I finish this chapter by reconstructing his analysis and show how it accounts for the necessity of a micro-reduction truth.

1. Sider's Reductive Account of Necessity[3]

In his book *Writing the Book of the World* (2011, chapter 12), Theodore Sider puts forth a new reductive account of metaphysical necessity.[4] Metaphysical necessity (necessity for short), in the sense of contemporary metaphysics, is generally but not universally understood as the most general kind of necessity: a truth is metaphysically necessary iff it is true in all possible situations. There are two main features of Sider's account. First, Sider argues that necessity is *reducible*. He claims that necessity is metaphysically non-fundamental since it "is unneeded for the most fundamental inquiries of metaphysics" (2011, 267). He makes a controversial analogy to support his claim: "just as scientists tend not to treat psychological, economic and political notions as being necessary for the most fundamental inquiries of mathematics and physics, so philosophers tend not to treat semantic, moral, epistemic, causal and modal notions as being necessary for the most fundamental inquiries of metaphysics" (2011, 267). The second feature is that necessity is identified with the "certain kinds" of truths. Roughly speaking, for a sentence *S* to be necessary is for *S* to be a mathematical truth, or a law of metaphysics, or an analytic truth, or a natural-kind truth, or a micro-

[3] Sider named his account (modal) Humeanism: "I prefer a third strategy for reducing modality, which I'll call Humean, for lack of a better word". (Sider 2011, 269) But just to avoid confusion with Humeanism in other philosophical areas, I will stick to Sider's reductive account of necessity or Sider's account for short.

[4] Sider briefly mentions his reductive account of modality in his "Reductive Theories of Modality" (2003 sec 4) and "Against Parthood" (2013a, sec 10) where he employs his account to argue against the argument from the possibility of gunk.

reduction truth, or a de re modal truth, or a logical consequence of such truths.

Before Sider, a number of metaphysicians had proposed different accounts to define necessity as *truths in all possible worlds*. Their views diverge only in terms of the nature of possible worlds. David Lewis (1986, sections 1.2-1.6) holds a concrete conception of possible world: possible worlds are identified as spatiotemporally and causally isolated entities. Unlike Lewis, abstractionists hold an abstract conception of possible world: possible worlds are identified as abstract entities—the maximal consistent sets of propositions, sentences, states of affairs, and so on.[5] David Armstrong (1989) holds a combinatorial conception of possible world: possible worlds are identified as combinations of the elements of the actual world.

Compared with the theories of possible worlds mentioned, there are a number of (alleged) theoretical merits in favor of Sider's account. Sider's account, if successful, gives a non-circular reduction of necessity. This alone is a significant achievement.[6] Sider's account, if successful, offers a parsimonious ontology and ideology since it avoids commitment of the existence of Lewisian possible worlds and taking modality as a primitive notion. Sider's account, if successful, enriches our understanding of the notion of metaphysical necessity. However, Sider's account will be flawed unless it defines each of the 'certain kinds' of truths *nonmodally*, and the definitions in question must be plausibly true. Otherwise, Sider's account loses its (alleged) theoretical merits. So far, it is not immediately obvious that Sider's account can actually accomplish this.

Let us present the reduction of necessity proposed by Sider. According to his account, necessity is reductively defined in terms of the following formulations (D1) and (D2):

(D1) A sentence *S* is a necessary truth iff *S* is a logical consequence of the modal axioms.

(D2) A sentence *S* is a modal axiom iff *S* is either

[5] See Plantiga (1974; 1976) and Adam (1974) for defenses of this view.

[6] David Lewis (1986, ch. 1) argues that only his theory of possible worlds can offer a non-circular reduction of modality and that it is one of the reasons to believe his modal realism. However, a number of philosophers reject this claim. For example, Marshall (2016) recently proposes a different criticism of Lewis's modal realism.

i) a mathematical truth,

ii) a law of metaphysics,

iii) an analytic truth,

iv) a natural-kind truth,

v) a micro-reduction truth, or

vi) a de re modal truth.[7]

Let us assume that *S* is written in a natural language, say English. On Sider's view, a sentence *S* in English is a necessary truth iff *S* in English is a mathematical truth, or a law of metaphysics, or an analytic truth, or a natural-kind truth, or a micro-reduction truth, or a de re modal truth, or a logical consequence of such truths. (D1) and (D2) are the identification of necessity. In what follows, let us briefly discuss each of these kinds of truths.

Logical Consequences

Sider claims that the notion of a logical consequence can be explained in terms of a model-theoretic account (2011, 273). According to this account, a sentence *S* is a logical consequence of a set of sentences $K_1,K_2,\ldots K_n$ iff there is no model in which $K_1,K_2,\ldots K_n$ are true and S is false. Alternatively, he proposes an attempted analysis of the notion of a logical consequence in terms of Lewis's account of lawhood.[8] According to Sider's attempted analysis, a sentence *S* is a logical consequence of $K_1,K_2,\ldots K_n$ iff *S* is entailed from $K_1,K_2,\ldots K_n$ on the basis of the laws of logic. For the laws of logic, it is defined in terms of a regularity in the best deductive system that achieves the best combination of simplicity and strength. Either way, he does not take the notion of a logical consequence to be metaphysically basic.

Mathematical Truths

Sider's definition of a mathematical truth is pretty simple. A sentence *S* is a mathematical truth iff *S* concerns only mathematics and *S* is true (Sider 2011, 273).

---

[7] A trivial point: Sider's (2011) original text uses "proposition" as the bearer of modal properties. But without further elaborations of my metaphysical assumption of proposition, let us stick to "sentence".

[8] For more, see Sider (2011, sections 10.3; 3.1), Lewis (1973; 1983; 1986; 1994).

Analytic Truths

On Sider's view, analytic truths are defined as the following: a sentence *S* is an analytic truth iff *S* is true and definitional (2011, 274; 193).[9] Sider provides no further definition of the notion of definitionality. He only describes that such a notion can be roughly understood as a constraint on the meaning of an expression. For example, the sentence 'Every vixen is a female fox' is an analytic truth since it is true and has a definitional constraint on 'vixen'. When a competent speaker of English interprets 'vixen', the interpretation made by the speaker has certain "definitional constraint" on the meaning of 'vixen'. It is worth noting that Sider's notion of analyticity carries no commitment to "truth by convention" in the traditional sense of analyticity.[10] On his view, truth and definitionality come apart. Analytic sentences are just the sentences that both are true and definitional. At any rate, we will consider a modified analysis to explain the necessity of an analytic truth in chapter 3.

Laws of Metaphysics

Sider's account of metaphysical laws is similar to Lewis's account of lawhood (2011, 275).[11] His definition of a metaphysical law is presumably of the form:

> (ML) A sentence *S* is a metaphysical law iff *S* is a regularity in the best system, where a best system is a deductive system that achieves the best combination of simplicity and strength.

How to balance simplicity and strength is an important matter. If complexity is cheap, then special-science regularities counts as laws. If complexity is expensive, then only laws of logic count as laws. (Recall his proposed definition of the notion of a logical consequence.) If complexity is made somewhere in between special-science regularities and laws of logic, then laws of physics count as laws. The central claim of his account of a metaphysical law: if a cost of complexity is made intermediate

[9] For more, see Sider (2011, sections 12.4; 9.8).
[10] A popular version of truth by convention is defended by Ayer's (1936, 31-87).
[11] See Lewis (1973; 1983; 1986; 1994), and Sider (2011 sections 3.1; 12.5).

between laws of logic and laws of physics, then laws of metaphysics count as laws and laws of physics drop out.

Sider also claims that sentences of fundamental metaphysics are to be necessarily true if true or necessarily false if false. Possible examples of metaphysical laws are these:

Parthood is transitive.
There are no past or future objects.
Objects have temporal parts.
Any objects have a mereological sum.[12]

Sider's account of a metaphysical law is neutral as to whether 'parthood is transitive' is true or its negation is true. On his view, it does not matter whether parthood is transitive or parthood is not transitive. Whichever is true should be counted as a necessary truth.

Natural-kind Truths

A typical example of natural-kind truths is this: Every water molecule is a H_2O molecule. Sider's definition is as follows: A sentence *S* is a natural-kind truth iff *S* is a sentence of the form ⌜$\forall x(Fx \rightarrow Gx)$⌝, where *F* is a natural-kind term and *G* expresses the deep explanatory feature of the property expressed by *F* (Sider 2011, 283). One might find "the deep explanatory feature" obscure and even perhaps redundant. For example, it might be simpler to say instead: a sentence *S* is a natural-kind truth iff *S* is a sentence of the form ⌜$\forall x(Fx \rightarrow Gx)$⌝, where *F* is a natural-kind term and *G* expresses the same property as *F*. For present purposes, however, we will not proceed any further for this. As we will see, chapter 3 proposes a modified analysis to explain the necessity of a natural-kind truth.

De Re Modal Truths

This is a standard way to define de re modality: a sentence *S* is a de re modal sentence

[12] The examples come from Sider (2011, 274).

if *S* is a sentence of the form ⌜□*Fa*⌝ or ⌜∃*x*□(*Fx*)⌝. Sider basically adopts this standard and attempts to specify it. According to Sider, a sentence *S* is a de re modal truth iff i) *S* is true and ii) *S* is a sentence of the form ⌜□*Fa*⌝ or ⌜∃*x*□(*Fx*)⌝.

The first kind of de re modal truths is of the form: ⌜□*Fa*⌝. Following Kripke's (1972, 106-143) famous examples, Sider further specifies different types of '□*Fa*': i) 'Necessarily, *a* is *F*' where *a* is a proper name and *F* is a sortal predicate, ii) 'Necessarily, *a* is *F*' where *a* is a proper name and *F* specifies material origins, or iii) 'Necessarily, *a* is *b*' where *a* and *b* are proper names (2011, 287). 'Donald Trump is a human' is an example of the form i). Let '*A*' be a name of the table in front of me and let '*B*' abbreviate the piece of wood. Suppose that '*A*' is actually made from '*B*'. Then '*A* is *B*' is a de re modal truth of the form ii). Finally, 'Hesperus is Phosphorus' is of the form iii).

The second kind of de re modal truths is of the form: ⌜∃*x*□(*Fx*)⌝. This is a sentence that contains a variable inside the scope of a modal operator which is bound to a quantifier outside. On Sider's view, this kind of de re modality can accommodate with the above categories i)-iii) by some modification.[13]

It might be argued that Sider's definition of de re necessity is too narrow. For the sake of argument, let us suppose that being a cartoon mouse is an essential property of Mickey Mouse. On this basis, the sentence 'Necessarily, Mickey Mouse is a cartoon mouse' is a de re modal truth. But according to the definition suggested by Sider, it is not obvious that the sentence is of a de re modal truth, depending on what he means by a "sortal" predicate.

However, Sider claims that the above syntactic definition is just one approach he could take to define de re necessity. An alternative approach to define de re modal truths is to appeal to Lewis's counterpart theory.[14]

Micro-reduction Truths

We will give an in-depth study of the notion of a micro-reduction truth in the rest of

[13] Sider claims that the modification can be made by adopting Tarskian truth-theory to assign truth-conditions to sentences with free variables. But for present purposes, we will not discuss this complication.

[14] See Sider (2002; MS)

this chapter.

Before moving on to Sider's definition of a micro-reduction truth, it is worth mentioning a novel feature of Sider's reductive account. On this account, there is no underlying *unification* of the 'certain kinds' of truths. The following quotations of Sider manifest the non-unifying feature of his account of necessity.

> "What determines the "certain sort" of propositions? Nothing "metaphysically deep". For the Humean, necessity does not carve at the joints. There are many candidate meanings for 'necessary', corresponding to different "certain sorts" our linguistic community might choose. Since none of these candidates carves at the joints, our linguistic community is free to choose whichever of these it likes. Perhaps the choice is arbitrary[.]" (2011, 269)

> "The Humean shares the conventionalist's goal of accounting for modality in a fundamentally amodal world. To that end, it's important that the "certain sorts" of propositions invoked by the Humean are not objectively distinguished, that no joint in reality encircles the class of logical and mathematical truths. So what does select this class? Something about us, says the Humean. Perhaps the choice of the "certain sorts" is conventional." (2011, 270)

> "The core idea of the Humean account, then, is that necessary truths are truths of certain more or less arbitrarily selected kinds." (2011, 271)

As we have seen from the quotations, Sider's account does not intend to explain necessity by formulating a single necessary and sufficient condition for being the 'certain kinds' of truths. Sider is perfectly willing to admit that the identification of necessity with the 'certain kinds' of truths is more or less heterogeneous. The import of the identification is to reflect *ipso facto* our linguistic behaviors and conventions of what to mean by the word 'necessity'.

Having concisely described Sider's reductive account and each of the 'certain kinds' of truths, we will proceed to articulate Sider's definition of a micro-reduction truth in the following sections.

2. Sider's Analysis of the Necessity of a Micro-reduction Truth

Let 'C_0' be a complex predicate which describes the actual microphysical states of New York City in every detail. On any reasonable account, (C) is surely necessary.

(C) Every C_0 is a city.
A logical form of (C): $\forall x(C_0 x \rightarrow \text{city } x)$.

According to Sider's account, (C) is a necessary truth since (C) is a logical consequence of a micro-reduction truth. Let us recall the reduction of necessity (D1) and (D2) provided above.

(D1) A sentence *S* is a necessary truth iff *S* is a logical consequence of the modal axioms.

(D2) A sentence *S* is a modal axiom iff *S* is either i) a mathematical truth, ii) a law of metaphysics, iii) an analytic truth, iv) a natural-kind truth, v) **a micro-reduction truth**, or vi) a de re modal truth.

Putting all together, Sider argues that since (C) is a logical consequence of a micro-reduction truth, (C) is a necessary truth. From there, Sider's account is said to give a reductive analysis of the necessity of (C).

However, it is not immediately clear exactly what a micro-reduction truth is like. Therefore, Sider attempts to define the notion of a micro-reduction truth in terms of the following.

(D3) A sentence *S* is a micro-reduction truth iff for a true **metaphysical semantics** *M*, and for an **application axiom** *A* in *M*, *S* **corresponds to** *A*.

At a first glance, it is difficult to fully understand (D3) without further elaborations. In order to properly understand (D3), I begin in exegetical mode by clarifying each of the technical terms such as "metaphysical semantics", "application axiom" and "correspond to" respectively. Let us firstly explain the notion of a metaphysical

semantics.

3. Defining Metaphysical Semantics

In his (2011, section 7.4), Sider develops his metaphysical semantics to shed light on the problem of fundamentality: in what way do non-fundamental facts hold *in virtue of* fundamental facts? [15] Before proceeding, I shall outline Sider's theory of fundamentality and how he thinks of the relation between the fundamental and the non-fundamental.

Sider's Theory of Fundamentality[16]

When metaphysicians talk about fundamentality, they talk about the idea that there is something metaphysically basic or rock-bottom in the world. In general, they all agree that fundamental facts are metaphysically basic or rock-bottom facts, but they hold different views about what fundamental facts are like. On Sider's view, facts about subatomic particles are fundamental facts, whereas facts about someone's walking in a park are non-fundamental facts. But this is merely an intuitive distinction between the fundamental and the non-fundamental. To make the distinction more precisely, Sider makes two claims about the fundamental and non-fundamental truths: *Completeness* and *Purity*.

> **Completeness**
>
> "Every non-fundamental truth holds in virtue of some fundamental truths." (2011, 105)

> **Purity**
>
> "Fundamental truths involve only fundamental notions." (2011, 106)

[15] For more, see Sider's (2011, chapter 7)

[16] My presentation of Sider's theory of fundamentality is sketchy and incomplete. For example, I have passed over other alternative approaches to the nature of 'hold in virtue of', such as grounding approach. For more on grounding, see Fine (2001; 2010; 2012); Rosen (2010); Schaffer (2009); anthology by Correia and Schnieder (2012). For more on Sider's theory of fundamentality, see his (2009; 2011 chapters 7-8).

Consider the truth that there is a city. Given Purity, the truth that there is a city is a non-fundamental truth since it contains a non-fundamental term 'city', and it expresses a non-fundamental notion. Given Completeness, the truth that there is a city must hold in virtue of some fundamental truths. However, there seems to be a problem. How do non-fundamental truths "hold in virtue of" fundamental truths? In other words, what is the connection between the fundamental and the non-fundamental? To explain the nature of the connection in question, Sider proposes a metaphysical type of semantic theory—"Metaphysical Semantics". He defines 'hold in virtue of' in terms of his metaphysical semantics as follows.

> **Completeness (new version)**
>
> "Every sentence that contains expressions that do not carve at the joints has a metaphysical semantics" (2011, 116).

This new version of Completeness makes a new requirement of being fundamental. The basic idea of Sider's metaphysical semantics is that for any non-fundamental term *T*, there is a metaphysical semantics that gives a meaning to *T* in fundamental language. Given the new version of Completeness, Sider's metaphysical semantics, if successful, in principle can describe the non-fundamental story of the world by using only fundamental terms. However, it is natural to wonder how exactly Sider's metaphysical semantics works. Hence, we move on to explain Sider's metaphysical semantics in detail.

Metaphysical Semantics

So much for the theory of fundamentality. I hope that the brief presentation outlined above has helped to clarify Sider's underlying motivation for developing an account of how the fundamental the non-fundamental are connected to each other. As we have just seen, his metaphysical semantics is a metaphysical type of semantic theory (that takes the form of a truth theory) which explains how the fundamental and the non-fundamental are connected to each other. The definition of Sider's metaphysical semantics is as follows.

(D4) *M* is a metaphysical semantics iff

(1) *M* takes the form of a truth theory,

(2) *M* is true,

(3) Every sentence ϕ in the meta-language of *M* must be stated in purely fundamental expressions, and

(4) All the T-theorems of *M* must be explanatory.

The First Characterization

Sider defines the notion of a metaphysical semantics as a certain type of semantic theory. But what type of theories of meaning does his metaphysical semantics actually take, given that there are a considerable number of such theories? Here is a suggestion from Sider.

> "A metaphysical semantics is a semantic theory with two distinctive features. First, meanings are to be given in purely joint-carving terms. For example, if the semantic theory takes the form of *a truth-theory*, then the truth-conditions must be stated in perfectly joint-carving terms" (2011, 112 my emphasis).

The quotation suggests that Sider's metaphysical semantics takes the form of a truth theory. Here, I assume that the type of semantic theory in question is Donald Davidson's (1967) theory of meaning.[17]

In order to understand Sider's metaphysical semantics, let us briefly outline Davidson's theory of meaning. In his (1967) "Truth and Meaning", Davidson claims that the notion of meaning is best understood in terms of the notion of truth. He also claims that a theory of truth which specifies the truth conditions for all sentences of a language would provide sufficient knowledge to specify the meanings for all sentences of the language in question. In this sense, to give the truth conditions for a sentence is to give the meanings for the sentence under consideration. To develop a theory of meaning which enables us to generate theorems that specify the truth conditions for

[17] A remark. In this thesis, I will assume that Sider's metaphysical semantics take the form of Davidson's truth theory, unless there is strong evidence to show otherwise. Although Sider (2011) has never explicitly claimed whether his metaphysical semantics takes the form of *Davidson's version* or *other versions* of theories of meaning, it seems to me that Sider's metaphysical semantics as presented in (2011) goes hand in hand with Davidson's theory of meaning.

all sentences of a language, Davidson adopts Tarski's (1936) theory of truth and defines the notion of meaning in terms of Tarski's definition of truth—known as Convention T. Given Tarski's Convention T, Davidson's theory of meaning yields a theorem of the form:

S is true iff *P*.

Theorems of this form are generally called T-theorems. The function of a T-theorem is to give meanings to each sentence of a natural language. In a T-theorem, '*S*' is a name of any sentence of the object-language and '*P*' is a sentence in the meta-language. The object-language is a language for which we try to give meanings and the meta-language is a language which translates the object-language. There is no constraint whether the object- and meta-language have to be identical or different. It is permissible to give the meanings for English sentences (object-language) in English sentences (meta-language). In this case, Davidson's theory of meaning would yield a T-theorem like this:

'Snow is white' is true iff snow is white.

It is also permissible to give the meanings for German sentences (object-language) in English sentences (meta-language). In this case, Davidson's theory of meaning would yield a T-theorem like this:

'Schnee ist weiss' is true iff snow is white.[18]

So far, we have presented some basic ideas of Davidson's theory of meaning. Now let us return to Sider's metaphysical semantics.

Sider attempts to explain the connection between the fundamental and non-fundamental in terms of the notion of a metaphysical semantics. Like Davidson who claims Tarski's theory of truth which specifies the truth conditions for all sentences of

[18] This example is adopted from the entry in "Donald Davidson" in Stanford Encyclopedia of Philosophy.

a language would provide sufficient knowledge to give the meanings for all sentences of the language under study, Sider makes a similar claim that Davidson's theory of meaning which gives the meanings of all sentences of a natural language would provide sufficient knowledge to give the meanings of all sentences of a non-fundamental language. Unlike Davidson's theory of meaning, the object-language for which Sider tries to give the meaning is a non-fundamental language, and the meta-language is a fundamental language.[19] For Sider's metaphysical semantics, to give a "metaphysical" truth-condition for a sentence of non-fundamental language is to give a meaning to the sentence in question.

A metaphysical semantics of a language begins by stipulating a number of axioms that assign meanings to each of the lexical items in that language, yielding T-theorems of the form:

T-theorem: *S* is true iff ϕ.

In T-theorems of a metaphysical semantics, '*S*' is a name of a sentence in non-fundamental language for which we are trying to give meanings, and 'ϕ' is a sentence in fundamental language which translates *S* in non-fundamental language. By yielding T-theorems, Sider claims that a metaphysical semantics can give meanings to each sentence of non-fundamental language in fundamental language. Given Sider's metaphysical semantics, we get a better grip on how the non-fundamental connects to the fundamental: they are semantically connected to each other by a metaphysical semantics.[20]

[19] Another difference between Davidson's theory of meaning and Sider's metaphysical semantics concerns with the conception of "meaning". The former holds a Fregean traditional conception of sentence-meaning where the meaning of a sentence is an object of thought or a propositional attitude or whatever is like, that concerns with the objective, strict and literal meaning of the sentence. However, this is an open question whether or not Sider would take the traditional one. I think he probably does not endorse the traditional conception of meaning. Also, I am unsure which conception of meaning Sider would take, and what Sider means by 'meaning' in his metaphysical semantics. In section 2.5, we will see that this unclarity leads to a problem of explanatory power in Sider's metaphysical semantics

[20] In fact, it is still unclear to me how 'ϕ' gives meanings for '*S*' and I have a hard time understanding how the non-fundamental and the fundamental are "semantically" connected to one another by a metaphysical semantics. Do they have same meaning by expressing the same proposition or state of affairs? Or do they have same meaning by having the same mental state of the speaker? These are pressing questions for Sider's metaphysical semantics. We will consider more on this in section 2.5.

The Second Characterization

As we have mentioned, Sider's metaphysical semantics takes the form of a truth theory. As a result, the 'iff' in the T-theorems is best plausibly understood as the material biconditionals.

The Third Characterization

Sider requires that every sentence ϕ in the meta-language of a metaphysical semantics must be stated in purely fundamental expressions. It is an essential element of his metaphysical semantics. As just mentioned, a metaphysical semantics gives meanings to each sentence of non-fundamental language in purely *fundamental* language. Hence, ϕ must be written in fundamental language. As Sider puts it explicitly,

> "Metaphysical semantics is more ambitious in that by giving meanings in fundamental terms, it seeks to achieve something not sought by linguistic semantics: to show how what we say fits into fundamental reality." (2011, 112)

> "The requirement that meanings are 'given' in purely joint-carving terms amounts to the requirement that ϕ be phrased in purely joint-carving terms" (2011, 113).

For Sider, predicates like "is fundamental" and "is joint-carving" are synonymous. But what are fundamental expressions? Based on Purity (fundamental truths involve only fundamental notions), here is my best guess:

> An expression is fundamental iff it expresses a fundamental notion.

For instance, the expression 'city' is counted as a non-fundamental expression because it expresses a non-fundamental notion, whereas the expression 'charge' is counted as a fundamental term because it expresses a fundamental notion. But what are fundamental notions? For Sider, 'is fundamental' is a primitive notion. In his (2011, chapter 13), Sider illustrates rather than defining all his primitive ideologies. Let us illustrate them as follows:

1. First-order logical constants (‘∀’, ‘¬’, ‘∨’, ‘=’, etc.),

2. Predicates of fundamental physics (‘is a unit negative charge’, ‘is quark flavors’, ‘is a mass’, etc.)

3. Predicate ‘∈’ for set-membership, and

4. The notion of structure (‘is fundamental’)

Sider’s illustration of his primitive ideologies helps to explain which notion counts as fundamental or non-fundamental. From there, we can roughly distinguish the fundamental and the non-fundamental notions by this:

A notion is fundamental iff it is expressed by either (1), (2), (3), or (4);

In any case, this is Sider’s illustration of a fundamental notion.[21] Though it may be debatable, it helps to illustrate which notion counts as fundamental or non-fundamental in terms of his theory of fundamentality. For example, the notion ‘city’ is non-fundamental because it is expressed the expression ‘city’ and such an expression does not belong to either (1), (2), (3) or (4). By contrast, the notion ‘mass’ is fundamental because it is expressed by the expression ‘mass’ and such an expression belongs to (2).

Since Sider requires that every sentence ϕ in a metaphysical semantics must only be written in fundamental expressions, ϕ can only be written in first-order logical constants, predicates of fundamental physics, set-membership predicate, the notion of structure, and nothing else.

The Fourth Characterization

Sider’s metaphysical semantics requires that all the T-theorems must be explanatory.

[21] There are good reasons to think that Sider’s illustration of a fundamental notion is disputable. For there could have been other instantiated fundamental properties. For example, being a color might be an instantiated fundamental property, according to some theory of color. Similarly, consciousness might be an instantiated fundamental property, according to dualism. And so on.

At bottom, a metaphysical semantics is an *explanation* of how the fundamental and the non-fundamental are connected to each other. He defines the notion of a metaphysical semantics to be a semantic theory whose job is to yield its T-theorems that give meanings to each sentence of non-fundamental language in fundamental language. Just as a theory of meaning seeks to explain our linguistic phenomena, a metaphysical semantics must *in some sense* explain our linguistic phenomena.[22] In this regard, Sider writes:

> "A metaphysical semantics must successfully explain the linguistic behavior of the population in question [.]" (2011, 114)

Let us set this aside and assume for the sake of argument that Sider's metaphysical semantics is an (alleged) explanatory theory. As we will see in chapter 2, it is probably not the case.

Instead of trying to validate that Sider's metaphysical semantics is an explanatory theory, let us procced to how a metaphysical semantics generates its T-theorems and gives meanings to each sentence of non-fundamental language in fundamental language.

To start off, we need to set up a set of axioms that assign meanings to each of the lexical items in a given language. Let us stipulate our language to be English. For the sake of simplicity, we will simplify our language where there are two lexical items: one name 'New York City' and one predicate 'is a city'. For any name, there is a reference axiom ⌜'New York City' refers to New York City⌝. For any predicate, there is an application axiom ⌜$\forall x$('is a city' applies to $x \leftrightarrow \phi x$)⌝. To build up an atomic sentence, we need another axiom to combine references and predicates. Hence, there is a compositionality axiom: ⌜For any 'a', and for any 'F', 'aF" is true iff 'F" applies to the reference of 'a'⌝. From there, we are able to derive the T-theorems from these

[22] Sider points out that the explanatory goals between his metaphysical semantics and a theory of meaning are different. Unlike a theory of meaning in general, what a metaphysical semantics tries to explain is more modest. For example, it does not aspire to integrate with theories of cognitive science, psychology, syntactic theory, theories of action and rationality, and so on (Sider 2011, 112-113). How exactly does a metaphysical semantics explain linguistic phenomena? It is an important matter, but we will not proceed further until chapter 2.

axioms. An illustration of a metaphysical semantics for our simplified English looks like this:

Reference Axiom: 'New York City' refers to New York City.

Application Axiom: $\forall x$('is a city' applies to $x \leftrightarrow \phi x$).

Compositionality Axiom: For any name '*a*', and for any predicate '*F*', '*aF*' is true in English iff '*F*' applies to the reference of '*a*'.

T-theorem NY: 'New York City is a city' is true in English iff ϕ.

The T-theorem NY shows that a metaphysical semantics for our simplified English gives a meaning to the English sentence 'New York City is a city' in perfectly fundamental expressions ϕ. Thus understood, the fundamental and the non-fundamental are semantically connected to each other by a metaphysical semantics. This is the goal of Sider's metaphysical semantics.

After that, Sider quickly recognizes a (potentially fatal) problem for his metaphysical semantics, if it takes the form of Davidson's theory of meaning. Consider the following quotation of Sider.

> "Returning to metaphysical semantics of the truth-theoretic form: what exactly is required of a fundamental truth-condition ϕ for a sentence S in [a metaphysical semantics]? We know from the literature on Donald Davidson's approach to semantics that sentences ϕ and ϕ' can have the same truth-value, even necessarily so, despite the fact that ϕ is an appropriate truth-condition for S while ϕ' is not. 'Snow is white' is an appropriate truth-condition in a (nonmetaphysical) semantics for 'Snow is white'; neither 'Grass is green' nor 'Snow is white and 2+2=4' is appropriate. A metaphysical semantics must successfully explain the linguistic behavior of the population in question, and a truth-theory with necessarily true conjuncts tacked onto each of its truth-conditions is presumably not explanatory, though it's a hard question why not. I have no particular answer to the question, though I suspect that the approach of section 3.2 is applicable. And if the question proves intractable, the metaphysical semanticist could abandon the Davidsonian approach. I have chosen that approach largely because it's simple, not because metaphysical semantics is wedded to it." (Sider 2011, 114)

As Sider said, there is a widely recognized problem in Davidson's theory of meaning, as first pointed out by Foster (1976). The problem is that the theory inevitably entails problematic T-theorems of the sort: ⌜'Snow is white' is true iff snow is white and grass is green⌝. It is a bizarre consequence of Davidson's theory of meaning. It is bizarre because even though the problematic T-theorem is *true*, we are hard to see why the conjunct 'grass is green' contributes anything to the meaning of 'snow is white'. There must be something wrong with Davidson's claim:

> It is this: the definition [of Tarski's] works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence—any sentence—to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language. (1984, 24)

If Sider's metaphysical semantics takes the form of Davidson's theory of meaning, then it faces the same problem too. As we can see from the quotation, Sider thinks that his metaphysical semantics can avoid the problem by adopting another theory of meaning. He claims that Davidson's theory of meaning is just one form that a metaphysical semantics might take. An immediate question: what a non-Davidsonian theory of meaning might a metaphysical semantics adopt? Proponents of Sider's metaphysical semantics are invited to explore more on this.

We have already gone through the details of Sider's metaphysical semantics and clarified each of the essential features of it. To recapitulate, there are four essential features consisting of Sider's metaphysical semantics: i) it takes the form of a truth theory; ii) it must be true in a sense of material biconditionals; iii) it requires that every sentence ϕ must be stated in purely fundamental terms; iv) all the T-theorems of it must be explanatory. Next, we move on to clarify the notions of "application axiom" and "correspond to".

4. Defining 'Application Axiom' and 'Correspond to'

Let us proceed to the other technical terms: "application axiom" and "correspond to".

To refresh our memory, let us recall Sider's definition of a micro-reduction truth again:

> (D3) A sentence *S* is a micro-reduction truth iff for a true **metaphysical semantics** *M*, and for an **application axiom** *A* in *M*, *S* **corresponds to** *A*.

Since we have explained the notion of a metaphysical semantics previously, we will go on to articulate 'application axiom' and 'correspond to' respectively.

Application Axiom

Actually, we have already seen how an application axiom plays its role in a metaphysical semantics in the last section. According to Sider's official formulation, application axioms of a metaphysical semantics have the form:

> (a) Predicate '*F*' applies to an object *x* iff ϕx. (2011, 285)

To grasp (a), it would be better to see what an application axiom of Davidsonian theory of meaning looks like:

> (b) Predicate '*F*' applies to an object *x* iff *Fx*.

In a Davidsonian theory of meaning, (b) specifies the condition under which, for any predicate '*F*', '*F*' applies to an object *x*. To see this, let us try to give a theory of meaning for 'Donald Trump is a human' in a simplified English. Suppose that this simplified English has only two lexical items: one name 'Donald Trump' and one predicate 'is a human'. For any name, there is a reference axiom: 「Donald Trump' refers to Donald Trump」. For any predicate, there is an application axiom: 「'is a human' applies to $x \leftrightarrow x$ is a human」. Furthermore, there is a compositionality axiom to combine references and predicates: 「For any '*a*', and for any '*F*', '*aF*' is true iff '*F*' applies to the reference of '*a*'」. This can be illustrated as follows:

> Reference Axiom: 'Donald Trump' refers to Donald Trump.

Application Axiom: ‘is a human’ applies to $x \leftrightarrow x$ is a human.

Compositionality Axiom: For any name ‘*a*’, and for any predicate ‘*F*’, ‘*aF*’ is true iff ‘*F*’ applies to the reference of ‘*a*’.

T-theorem DT: ‘Donald Trump is a human’ is true in English iff Donald Trump is a human.

T-theorem DT shows that a Davidsonian theory of meaning for English gives a truth condition for the English sentence ‘Donald Trump is a human’ in English. Given that the notion of truth is defined in terms of the notion of meaning, it follows that a Davidsonian theory of meaning for English gives a meaning to the English sentence ‘Donald Trump is a human’ in English. This is exactly how an application axiom plays its role in a Davidsonian theory of meaning. As we have illustrated in section 1.3, the application axiom of a metaphysical semantics does almost exactly the same thing. There is a crucial difference between Sider’s metaphysical semantics and a Davidsonian theory of meaning. Sider’s metaphysical semantics requires that the right bijunct of a metaphysical semantics must be phrased in purely fundamental expressions mentioned in section 1.3. Unlike (b), (a) is required to specify the condition under which for any predicate ‘*F*’, ‘*F*’ applies to an object *x* iff *x* is ϕ, where ϕ is phrased in purely fundamental expressions. So, let us formulate (D5) as a proposed definition of an application axiom of Sider’s metaphysical semantics.

(D5) A sentence *S* is an application axiom of a true metaphysical semantics *M* iff *S* is a sentence of the form ⌜$\forall x$(‘*F*’ applies to $x \leftrightarrow \phi x$)⌝ and *S* is an axiom of *M*.

Corresponding to

Having explained ‘application axiom’, we proceed with explaining ‘correspond to’. Here is a suggested definition:

(D6) A sentence S corresponds to an ⌜application axiom⌝ of a true metaphysical semantics M iff S is a sentence of the form ⌜$\forall x(Fx\leftrightarrow\phi x)$⌝ and ⌜$\forall x$(‘$F$’ applies to $x\leftrightarrow\phi x$)⌝ is an application axiom in M.

As we have seen, (C) is a necessary truth.

(C) Every C_0 is a city.

But (C) does not seem to be a mathematical truth, or a law of metaphysics, or a natural kind truth, or an analytic truth, or a de re modal truth or a logical consequence of such truths. How does Sider account for such a necessity? His answer is that (C) is a logical consequence of a micro-reduction truth. After that, he defines a micro-reduction truth to be a sentence of the form ⌜$\forall x(Fx\leftrightarrow\phi x)$⌝, and explains that it corresponds to the application axiom of a true metaphysical semantics. This is the idea of (D6). As Sider puts it,

> “This problem [the problem of how the necessity of a micro-reduction truth should be explained] threatens the Hmuean as well, since sentences like (C) do not seem to be mathematical truths, analytic truths, laws of metaphysics, natural kind axioms or logical consequences of such sentences.
>
> The Humean should introduce another group of modal axioms, derived from a theory of metaphysical truth-conditions for the language in question. There are, I assume, metaphysical truth-conditions for statements about cities, smiles, and candy. The metaphysical semantics generating these metaphysical truth-conditions, let’s assume, contains axioms of the form “Predicate F applies to object x iff $\phi(x)$”, where F may be ‘city’, ‘smile’, ‘candy’, and so on, and where $\phi(x)$ is phrased in purely fundamental terms. For each such axiom of the metaphysical semantics, we should add a corresponding modal axiom: $\forall x(F\text{x}\leftrightarrow\phi x)$.” (Sider 2011, 285)

According to the quotation of Sider, he suggests that there is an application axiom in a metaphysical semantics: ⌜$\forall x$(‘F’ applies to $x\leftrightarrow\phi x$)⌝. This has already manifested in (D5). He then defines a micro-reduction truth as a sentence of the form: ⌜$\forall x(Fx\leftrightarrow\phi x)$⌝, which corresponds to the application axiom ⌜$\forall x$(‘F’ applies to $x\leftrightarrow\phi x$)⌝ of a true

metaphysical semantics. This is the idea of saying that ⌜$\forall x(F\text{x}\leftrightarrow\phi x)$⌝ corresponds to the application axiom of a true metaphysical semantics and is the sentential form of Sider's definition of a micro-reduction truth.

5. Revisiting Sider's Analysis of the Necessity of a Micro-reduction Truth

In the previous sections, we have clarified the technical terms 'metaphysical semantics', 'application axiom', and 'correspond to'. Now, we are in the position to understand Sider's analysis of the necessity of a micro-reduction truth. Sider's analysis runs as follows:

> (D1) A sentence *S* is a necessary truth iff *S* is a logical consequence of the modal axioms.
>
> (D2) A sentence *S* is a modal axiom iff *S* is either i) a mathematical truth, ii) a law of metaphysics, iii) an analytic truth, iv) a natural-kind truth, v) **a micro-reduction truth**, or vi) a de re modal truth.

(D1) and (D2) are of Sider's account of necessity. He then defines a micro-reduction truth as (D3).

> (D3) A sentence *S* is a micro-reduction truth iff for a true metaphysical semantics *M*, and for an application axiom *A* in *M*, *S* corresponds to *A*.

To understand (D3), we have to understand what it is meant by 'metaphysical semantics', 'application axiom' and 'correspond to'. In section 1.3, we have therefore provided a suggested definition of Sider's metaphysical semantics.

> (D4) *M* is a metaphysical semantics iff (1) *M* takes the form of a truth theory, (2) *M* is true, (3) every sentence ϕ in the meta-language of *M* must be stated in purely

fundamental expressions, and (4) all the T-theorems of *M* must be explanatory.

In section 1.4, we have also defined and interpreted 'application axiom' and 'correspond to' as (D5) and (D6).

(D5) A sentence *S* is an application axiom of a true metaphysical semantics *M* iff *S* is a sentence of the form ⌜∀*x*('*F*' applies to *x*↔ϕ*x*)⌝ and *S* is an axiom of *M*.

(D6) A sentence *S* corresponds to an ⌜application axiom⌝ of a true metaphysical semantics *M* iff *S* is a sentence of the form ⌜∀*x*(*Fx*↔ϕ*x*)⌝ and ⌜∀*x*('*F*' applies to *x*↔ϕ*x*)⌝ is an application axiom in *M*.

Putting (D4)-(D6) together, we are in the position to understand (D3). Let us rephrase (D3) more succinctly into (D7).

(D7) A sentence *S* is a micro-reduction truth iff for a true metaphysical semantics *M*, and for an application axiom ⌜∀*x*('*F*' applies to *x*↔ϕ*x*)⌝ *A* in *M*, *S* is a sentence of the form ⌜∀*x*(*Fx*↔ϕ*x*)⌝ that corresponds to *A* in *M*.

However, it is obvious that ⌜∀*x*(*Fx*↔ϕ*x*)⌝ is the sentential form of Sider's definition of a micro-reduction truth and itself has no semantic value because '*F*' and 'ϕ' just predicate letters without any particular meaning. They will have some particular meanings when they get assigned meanings in an interpretation. Similarly, just by (D7) alone, Sider still cannot account for the necessity of (C). For it tells us nothing as to whether or not (C) count as being necessary. But he claims that this problem can be solved. To accomplish it, we substitute the predicate 'is a city' for '*F*' and some complex and fundamental predicates of the actual microphysical states of being a city for 'ϕ':

(T) ∀*x*(city(*x*)↔ϕ*x*).

If a metaphysical semantics for English can give a meaning to an English sentence 'x is a city' in perfectly fundamental expressions ϕ, then ϕ must cover every conceivable circumstance where a city is like. Since clearly there are infinitely many ways a city could be, ϕ must fit with intuitively correct usage of 'is a city' by holding in every conceivable circumstance. Since Sider allows that the language of ϕ could be infinite, let us assume that ⌜$\forall x(x$ is a city$\leftrightarrow\phi x)$⌝ should be properly written as (T'):

(T') $\forall x[\text{city}(x)\leftrightarrow\phi_1(x) \vee \phi_2(x) \vee \phi_3(x) \vee \phi_4(x) \vee \phi_5(x) \vee\ldots \phi_n(x)]$[23]

where '$\phi_1(x)$' is a (complex) formula which describes the actual microphysical states of Los Angeles, '$\phi_2(x)$' is a (complex) formula which describes the actual microphysical states of Chicago, '$\phi_3(x)$' is a (complex) formula which describes the actual microphysical states of Houston, and so forth. Thus, Sider claims that in this case, the actual microphysical states of New York City—C_0—will be one of the 'ϕs(x)'. It is this.

(T'') $\forall x[\text{city}(x)\leftrightarrow\boldsymbol{C_0(x)} \vee \phi_1(x) \vee \phi_2(x) \vee \phi_3(x) \vee \phi_4(x) \vee \phi_5(x) \vee\ldots \phi_n(x)]$.

Of course, Sider cannot fully specify the details of each of the disjuncts including 'C_0' in the right-hand side. Presumably, the specification will be infinitely open-ended. Even so, Sider contends that given (T''), this toy-model of a metaphysical semantics has already provided sufficient resource to account for the necessitation of (C). The idea is that (C) is necessary because (C) is a logical consequence of (T'').

To see this, let us demonstrate how (C) is logically derived from (T''). To improve readability, let us simplify (T'') into ⌜$\forall x(\text{city}(x)\leftrightarrow\phi^*(x)$⌝, and let '$\phi^*(x)$' be an abbreviation of '$C_0(x) \vee \phi_1(x) \vee \phi_2(x) \vee \phi_3(x) \vee \phi_4(x) \vee \phi_5(x) \vee\ldots \phi_n(x)$'.[24] Here is a proof:

[23] See Sider (2011, 285 footnote 36)

[24] '$\phi^*(x)$' should be read as a shorthand for the infinite disjunction of all possible situations in which a city is like. The actual microstate of New York City 'C_0' is just one of the infinite disjuncts.

1. $\forall x$[city(x)$\leftrightarrow\phi^*(x)$]	Modal axiom
2. City(x)$\leftrightarrow\phi^*(x)$	Universal Quantifier Elimination 1
3. $\phi^*(x)\rightarrow$City(x)	Biconditional Elimination 2
4. $C_0(x)$	Hypothesis
5. $\phi^*(x)$	Disjunction Introduction 4
6. City(x)	Implication Elimination 3, 5
7. $C_0(x)\rightarrow$City(x)	Implication Introduction 4, 6
8. $\forall x[C_0(x)\rightarrow$city($x$)]	Universal Quantifier Introduction 7

This natural deduction validates that ⌜$\forall x(C_0(x)\rightarrow$city($x$))⌝ is a logical consequence of ⌜$\forall x$(city(x)$\leftrightarrow\phi^*(x)$)⌝. In other words, it proves that (C) is a logical consequence of (T"). Given (D7), (T") is a micro-reduction truth under a particular interpretation.[25] Since (C) is a logical consequence of a micro-reduction truth, (C) is a necessary truth.

(C) Every C_0 is a city.

A logical form of (C): ⌜$\forall x(C_0(x)\rightarrow$city($x$))⌝.

(T) $\forall x$(city(x)$\leftrightarrow\phi x$).

(T") ⌜$\forall x$[city(x)$\leftrightarrow C_0(x)\vee\phi_1(x)\vee\phi_2(x)\vee\phi_3(x)\vee\phi_4(x)\vee\phi_5(x)\vee\ldots\phi_n(x)$]⌝

This is Sider's analysis of the necessity of a micro-reduction truth. Now we finish this chapter by recapitulating what we have done so far. In section 1.1, I have outlined Sider's account of necessity in general. In section 1.2, I have explicated his analysis of the necessity of a micro-reduction truth and identified three technical terms for further explanations. In section 1.3-4, I have clarified each of these technical terms. In section 1.5, I have reconstructed Sider's analysis and showed how it accounts for the necessity of a micro-reduction truth.

[25] More specifically, (C) is a logical consequence of (T) since (T) is a shorthand for (T"). As just mentioned, (T) is of a sentence of the form ⌜$\forall x(Fx\leftrightarrow\phi x)$⌝ under a particular interpretation of 'F' and 'ϕ'. Hence, (T) is a micro-reduction truth under a particular interpretation of 'F' and 'ϕ'.

CHAPTER 2 SCHAFFER'S MULTIPLE REALIZABILITY PROBLEM

This chapter introduces the multiple realizability problem. In section 2.1, I present a schematic argument to illustrate the multiple realizability problem of Schaffer. In section 2.2-2.5, I articulate each of the premises in the argument and how they reach to the conclusion that Sider's metaphysical semantics cannot handle multiple realizability. In section 2.6, I consider an impact on Sider's analysis of the necessity of a micro-reduction truth.

1. A Schematic Argument

In his "Metaphysical Semantics Meets Multiple Realizability" (2013), Jonathan Schaffer poses a deeply challenging objection to Sider's metaphysical semantics. In his article, Schaffer argues that Sider's metaphysical semantics cannot handle multiply realizable non-fundamental truths. To put it more carefully, let us standardize Schaffer's objection into the following schematic argument:

1. There are multiply realizable non-fundamental truths.

2. There are three conditions of adequacy for Sider's metaphysical semantics.

3. If there are multiple realizable non-fundamental truths, Sider's metaphysical semantics cannot satisfy the three conditions of adequacy.

---

4. *Therefore*, Sider's metaphysical semantics cannot handle multiply realizable non-fundamental truths.

This schematic argument illustrates the multiple realizability problem of Schaffer. In what follows, I will articulate each of the premises in this argument. Let us begin with premise 1.

2. Multiple Realizability

There are multiply realizable non-fundamental truths. For example, the truth that Moore has hands is obviously one of them. Such a non-fundamental truth is multiply realizable since its truth can hold in virtue of many different fundamental bases.[26] Let us define a multiply realizable non-fundamental truth as follows:

> (R) *T* is a multiply realizable non-fundamental truth iff i) *T* is non-fundamental, ii) *T* is true, and iii) *T* can hold in virtue of many different actual and non-actual fundamental bases.

Here, we will follow Sider's proposed standard of fundamentality stated in section 1.3. Consider the truth that Moore has hands. Such a truth is a non-fundamental truth since it contains a non-fundamental term 'hand'. The fundamental base of our world could be any sort of distributions of fields, permutations and combinations of particles, variations of physical principles, or whatever. Suppose that the actual fundamental base of our world is field-theoretic physics. Thus, it is actually the case that the truth that Moore has hands holds in virtue of field-theoretic physics. Presumably, it could have been the case that such a truth can hold in virtue of particle-theoretic physics, other than field-theoretic physics. Even more improbably, it could have been the case that such a truth can hold in virtue of an alien science where the fundamental physical properties are alien properties other than being mass, charge, spin, and so forth. Thus understood, non-fundamental truths 'Moore has hands', 'Vienna is a city', etc., are all multiply realizable because its truths can hold in virtue of many different actual and non-actual fundamental bases. Thus, premise 1 is validated. There are multiply realizable non-fundamental truths.

Schaffer offers an analogy. Given that multiple realizability poses a serious problem for reductive physicalism in philosophy of mind, so it also poses a similar problem for Sider's metaphysical semantics. Just as reductive physicalism cannot

[26] In this thesis, I am neutral on the disputes over the connection between the fundamental and the non-fundamental. So, it would be better to use 'hold in virtue of', rather than the notion of grounding, the notion of a metaphysical semantics, and so on.

identify a multiple realizable state "pain" with a specific physical state, Sider's metaphysical semantics cannot identify a multiple realizable non-fundamental truth with a specific fundamental truth. Of course, this analogy is just for a heuristic purpose. The basic idea of it is to show that there might be something wrong with this one-one identity view of the connection between the fundamental and the non-fundamental.

3. The Three Conditions of Adequacy

Let us examine premise 2. There are three conditions of adequacy for Sider's metaphysical semantics. As mentioned earlier, a metaphysical semantics of a language begins by stipulating a number of axioms that assign meanings to each of the lexical items in that language, yielding T-theorems of the form:

> A sentence *S* of *L* is true in *L* iff ϕ.

Schaffer (2013, 738) claims that Sider's metaphysical semantics must satisfy at least three conditions of adequacy. Here, let us understand 'conditions of adequacy' as constraints on Sider's metaphysical semantics. (Recall our suggested definition of Sider's metaphysical semantics in section 1.3. This makes a similar point.) Let us examine the three conditions of adequacy suggested by Schaffer.

The first condition of adequacy is Fitting. It says that ϕ in Sider's metaphysical semantics must fit with intuitively correct usage in every conceivable circumstance.

> Fitting: The biconditionals must fit with intuitive usage by holding in [all] conceivable circumstances. (Schaffer 2013, 739 my amendment)

Sider's metaphysical semantics must obey Fitting. After all, to specify the semantic connection between *S* and ϕ in the T-theorems is the ambition of Sider's metaphysical semantics. As Sider claims, "metaphysical semantics is more ambitious in that by giving meanings in fundamental terms, it seeks to achieve something not sought by linguistic semantics: *to show how what we say* [in non-fundamental language] *fit into fundamental reality*" (Sider 2011, 112 my emphasis). Sider's metaphysical semantics

must be able to give meanings to each sentence of non-fundamental language in fundamental language and it must match with intuitively correct usage of what we say in non-fundamental language in every conceivable circumstance.

(Two complications here. *First*, there is an uncertainty as to how much fit with intuitive usage Sider's metaphysical semantics requires. Sider contends that since his metaphysical semantics is just a toy model, it allows some imperfect matches with intuitive usage. We should not expect a toy metaphysical semantics to match with intuitively correct usage in every possible world or conceivable circumstance. He argues:

> "A more reasonable goal is the construction of "toy" metaphysical truth-conditions. These will be toy in at least two ways, First, they needn't match with intuitively correct usage in absolutely all possible worlds or conceivable circumstances. The mesh need only be approximate (the more mesh, the better)" (Sider 2011, 118).

But how much mismatch can be tolerated by a toy metaphysical semantics? Unfortunately, Sider made no mention of this. *Second*, there is an unresolved debate over whether Sider's metaphysical semantics requires to obey Fitting. He himself denies it.

> "The multiple realizability problem [of Schaffer] is generated by the assumption that metaphysical truth-conditions must obey a principle of 'Fitting', as Schaffer calls it—that they 'must fit with intuitive usage by holding in most conceivable circumstances'. […] But I don't think that Fitting is required by my theory of metaphysical truth-conditions itself." (Sider 2013e, 768)

But one thing for sure is that Sider contends that there is a *real* and *non-toy* metaphysical semantics. "What is the point of toy metaphysical truth-conditions? One point is to convince us that real, non-toy metaphysical truth-conditions exist" (Sider 2011, 117). Given this, we are thereby reasonable to expect that a real and non-toy metaphysical semantics must perfectly fit with intuitively correct usage in every conceivable circumstance. Therefore, a real and non-toy metaphysical semantics must obey Fitting. Imperfect match is only for a toy metaphysical semantics. Even so, we should also expect that a toy metaphysical semantics must fit with intuitively correct

usage as much as possible. But Sider owes us an explanation of how much mismatch can be tolerated by a toy metaphysical semantics.)

The second condition of adequacy is Sparse. It says that ϕ in Sider's metaphysical semantics must be phrased in purely fundamental terms.

> Sparse: The right bijunct must connect to the fundamental by being cast in purely fundamental terms. (Schaffer 2013, 739)

This constraint is plainly obvious. As Sider puts it, "[t]he requirement that meanings are 'given' in purely joint-carving terms amounts to the requirement that ϕ be phrased in purely joint-carving terms" (2011, 113). Although Sider's metaphysical semantics take a linguistic approach, it *per se* is a metaphysical account of the connection between the fundamental and the non-fundamental. After all, the goal of Sider's metaphysical semantics is to give meanings to each sentence of non-fundamental language in fundamental language.

The third condition of adequacy is Explanatory. It says that Sider's metaphysical semantics must explain our linguistic phenomena.

> Explanatory: The biconditionals must explain our linguistic behaviour. (Schaffer 2013, 739)

This constraint is also indisputable. To restate, Sider's metaphysical semantics is an explanatory theory of the connection between the fundamental and the non-fundamental. Sider (2011, 114) says that "a metaphysical semantics must successfully explain the linguistic behavior of the population in question". Thus, Sider's metaphysical semantics must obey Explanatory.

Putting all together: there are three conditions of adequacy for Sider's metaphysical semantics. They are Fitting, Sparse and Explanatory. Although there will and must be a controversy over the three conditions of adequacy, let us assume that the triple constraints are plausibly correct, and that they are essential ingredients of Sider's metaphysical semantics.[27] Hence, premise 2 is justified. There are three

[27] As we can see, Sider (2011, 116-117; 2013e, 768-769) rejects the principle of Fitting.

conditions of adequacy for Sider's metaphysical semantics.

4. Sider's Metaphysical Semantics Meets Multiple Realizability

Having articulated premise 1 and 2, let us examine premise 3. Here we shall consider whether Sider's metaphysical semantics can satisfy the three conditions of adequacy Fitting, Sparse and Explanatory, when it meets multiply realizable non-fundamental truths. So far, we agree that there are multiply realizable non-fundamental truths, like the truth that Moore has hands. We also agree that there are three conditions of adequacy for Sider's metaphysical semantics. So far so good. Now, we will see whether or not Sider's metaphysical semantics is compatible with multiple realizability.

For the sake of argument, let us assume that there is a metaphysical semantics for the multiple realizable non-fundamental truth that Moore has hands. A true metaphysical semantics of English will yield a theorem of the form:

> T-theorem M: 'Moore has hands' is true in English iff ϕ.

Fitting requires that T-theorem M must fit with intuitively correct usage of what we say by 'Moore has hands' in English in every conceivable circumstance. Sparse requires that ϕ in T-theorem M must be written in purely fundamental terms. Exploratory requires that T-theorem M must explain our linguistic behavior. Does T-theorem M satisfy the triple constraints? Schaffer argues that T-theorem M fails to do so. To show this, he asks the following question:

> Question: "What goes in for ϕ in the right bijunct of [T-theorem M] when there is multiple realizability" (Schaffer 2013, 741)?

What and how exactly can Sider's metaphysical semantics answer to Question? Sider (2011, 118-121) gives a few examples to show how his metaphysical semantics can give a meaning to the sentence 'There exists an atom of hydrogen'. However, Schaffer argues that it is unclear how Sider's metaphysical semantics can give a meaning for the multiply realizable non-fundamental truth that Moore has hands. Natural-kind terms such as hydrogen can only be realized by the actual fundamental base of the

world. Clearly, Sider's metaphysical semantics has no problem with this. But the truth that Moore has hands is another story. Although it is perfectly fine to say that the meaning of 'There exists an atom of hydrogen' is 'There exist an electron and a proton bonded to each other', it is highly problematic that the meaning of 'Moore has hands' is 'The subatomic particles of Moore's hands have such-and-such masses and charges and spins in such-and-such spatial arrangements'. Moore could have hands if our world is fundamentally field-theoretic physics. Or Moore could have hands if our world is fundamentally particle-theoretic physics but the subatomic particles has some other fundamental properties rather than mass, charge and spin. Or Moore could have hands if our world is fundamentally … And so forth. In fact, the truth that Moore has hands can be multiply realized via any conceivable circumstance where things might fundamentally be. This is the multiple realizability problem and Sider's metaphysical semantics is supposed to accommodate this kind of multiple realizability.

For Schaffer, he (2013, 741) claims that there is no answer to Question that can satisfy the triple conditions of adequacy: Fitting, Sparse and Explanatory. To defend this claim, he considers three possible answers to Question that Sider's metaphysical semanticists might take to handle multiple realizability, and then shows that none of them can satisfy more than one of the triple adequacy conditions. Let us see how it works.

5. Three Possible Answers

Now, we will consider the three possible answer to Question suggested by Schaffer. For the sake of argument, let us assume that the actual fundamental base of our world is field-theoretic physics. The first suggested answer is that ϕ in T-theorem M is phrased by the actual fundamental base, namely field-theoretic physics. The second suggested answer is that ϕ in T-theorem M is phrased by a disjunction of physically possible bases of our actual field-theoretic physics, where a physically possible base is one that is compatible with the true physical laws. In this case, ϕ in T-theorem M would be a relatively long disjunction of various physically possible distributions of field. The third suggested answer is that ϕ is T-theorem M is phrased by a disjunction

of all possible fundamental bases. In this case, ϕ in T-theorem M would be an open-ended infinitary disjunction of not only the actual fundamental base, but also the physically possible bases of our actual field-theoretic physics, plus anything that could be the fundamental base of our world, no matter how alien and inconceivable it sounds like. Things like God's arrangement, Cartesian evil demon, Nozick's experience machine, the mind of Berkeley's deity, etc., could be the fundamental base of our world and should be included in the infinitely open-ended disjunction. Schaffer suggests that these are the three possible answers to Question that Sider's metaphysical semantics might take to handle multiple realizability.[28] Next, let us closely examine each of them and see whether they can satisfy the triple conditions of adequacy.

The First Answer: Actual Realizer

Schaffer suggests that the first answer that Sider's metaphysical semantics may attempt to respond to the multiple realizability problem is to insert the actual fundamental base into ϕ in T-theorem M as follows.

> T-theorem M1: 'Moore has hands' is true in English iff the actual field-theoretic physics is such and such.

Suppose again that our world is fundamentally field-theoretic. The right bijunct of T-theorem M1 is filled in the highly complex details of actual field-theoretic physics that actually realize the truth that Moore has hands.

---

[28] In addition to the three possible answers, Schaffer (2013, 742-743) considers a fourth answer that Sider might take to be his favored answer. This answer is divided into two steps. The first step is to quantify over role properties of being a hand. In this case, T-theorem M would be written as follows:

> T-theorem MR: 'Moore has hands' is true in English iff Moore has properties that play the hands role. (Schaffer 2013, 742)

Next, the fourth answer proceeds to the second step. Since 'properties that play the hands role' in the right bijunct of T-theorem MR is plainly a non-fundamental term by the proposed standard of Sider's fundamentality. So the second step is to give a metaphysical semantics to such a non-fundamental term.

> T-theorem MR2: 'Moore has properties that play the hands role' is true in English iff ϕ.

Schaffer argues that the multiple realizability problem reappears. The truth that Moore has properties that play the hands role is also multiply realizable, since its truth can still hold in virtue of many different fundamental bases.

Let us evaluate the first answer. It satisfies Sparse but does not satisfy Fitting and Explanatory. The first answer satisfies Sparse since the right bijunct of T-theorem M1 is written in purely fundamental field-theoretic expressions. Recall that we suppose that our world is fundamentally field-theoretic physics. Unfortunately, the first answer fails to satisfy Fitting. It requires that T-theorem M1 must fit with intuitive usage by holding in *every* conceivable circumstance. Clearly, T-theorem M1 makes 'Moore has hands' in English come out false in a circumstance where the fundamental base of our world is particle-theoretic physics. In fact, any conceivable circumstance where a slightly different distribution of fields in physics other than the actual one would suffice to prove that T-theorem M1 fails to obey Fitting. Moreover, there are countlessly many and various conceivable ways things might fundamentally be. The actual realizer of the non-fundamental truth that Moore has hands is just one of the many. This is exactly the multiple realizability problem.

The first answer does not satisfy Explanatory either. Schaffer (2013, 744) writes that "[T-theorem M1] does not explain what we say since it is overly specific, includes irrelevant details, and misses the generalizations." He then goes on to explain what an apt explanation for non-fundamental truths is like. For him, "[a]n apt explanation for why English speakers say 'Moore has hands' involves something like the presence of *hands*. This is an apt level of explanation, with the appropriate specificity, relevance of information and amenability to generalization" (2013, 744). Here, I agree with Schaffer. For the reasons he gives are cogent and legitimate.

In addition, there is another reason to think T-theorem M1 is not an apt explanation for our linguistic behavior. For it is unclear to me exactly what Sider's metaphysical semantics tries to explain. I am not sure that I fully understand what Sider takes to be his conception of "meaning" in his metaphysical semantics. After all, his metaphysical semantics is a metaphysical kind of semantic theory that gives 'meanings' to every sentence of non-fundamental language in fundamental language. But he claims that the explanatory goal of his metaphysical semantics differs from those of theories of meaning. So he writes:

> "[T]he explanatory goals differ from those of linguistic semantics. In one way they are more ambitious, and in another, more modest. Metaphysical semantics

> is more ambitious in that by giving meanings in fundamental terms, it seeks to achieve something not sought by linguistic semantics: to show how what we say fts into fundamental reality. Metaphysical semantics is more modest in that it tries to explain a narrower range of phenomena. […] According to a traditional conception (largely associated with Frege), meaning plays a broad theoretical role: the meaning of a sentence is conventionally encoded by that sentence, grasped by anyone who understands the sentence, is communicated when the sentence is used; sentence-meanings are the objects of thought and other propositional attitudes, and so on." (2011, 112)

From Sider's quotation, we can see explicitly that his metaphysical semantics does not associate with the traditional conception of meaning and so does not take it to be propositional attitudes, objects of thought, states of affairs, etc. This suggests that there is an obscurity of what he takes to be his conception of 'meaning' in his metaphysical semantics. Since Sider's metaphysical semantics is not integrated with the traditional conception of meaning, it is permissible to go beyond the understanding of competent speakers when it assigns a certain 'meaning' to a given sentence of non-fundamental language in purely fundamental language. He claims:

> But she [a metaphysical semanticist] is not concerned to integrate her semantics with other linguistic or psychological theories. Thus she is not trying to integrate her semantics with syntactic theory, for example. And *she is free to assign semantic values that competent speakers would be incapable of recognizing as such, for she is not trying to explain what a competent speaker knows when she understands her language.* She might, for example, assign to an ordinary sentence about ordinary macroscopic objects a meaning that makes reference to the fundamental physical states of subatomic particles. And she might simply ignore Frege's (1952/1892) puzzle of the cognitive nonequivalence of co-referring proper names, since she is not trying to integrate her semantics with theories of action and rationality. (2011, 113 my emphasis)

Given the special feature of Sider's conception of meaning, I find this difficult to understand exactly what Sider's metaphysical semantics tries to explain. It seems to me that there is an inevitable tension between its 'metaphysical character' and 'semantic character' of his metaphysical semantics. In order to do the *metaphysical* works, his metaphysical semantics must hold an innovative and unfamiliar conception of meaning differing from a theory of meaning. If it makes no difference from the traditional one, what makes it so special and metaphysical? After all, one of the

metaphysical ambitions of his metaphysical semantics is to account for the connection between the fundamental and the non-fundamental. But in order to do the *semantic* works, his metaphysical semantics, like any theory of meaning, must in some sense follow the traditional and orthodox conception of meaning to explain what we say in non-fundamental language. Otherwise, it cannot be a declared semantic and explanatory theory. As I take it, this underlying tension is the main obstacle to apt explanations for our linguistic phenomena.

To conclude, even though the first answer satisfies Sparse, it does not satisfy Fitting and Explanatory.

The Second Answer: Potential Actual Realizers

Next, Schaffer considers the second answer that Sider's metaphysical semantics might take to handle the multiple realizability problem. This answer inserts the actual fundamental base—field-theoretic physics, plus all physically possible bases of our actual field-theoretic physics into ϕ in T-theorem M as follows.

> T-theorem M2: 'Moore has hands' is true in English iff the actual field-theoretic physics is such and such, or a physically possible distribution of field is such and such, or…

Suppose again that the actual fundamental story of our world is field-theoretic. The right bijunct of T-theorem M2 is filled in a relatively long disjunction of various physically possible distributions of field that could realize the truth that Moore has hands.

Let us evaluate the second answer. Schaffer claims that it is indeterminate as to whether T-theorem M2 satisfies Sparse. It is indeterminate precisely because T-theorem M2 contains a disjunction in the right bijunct. Sider (2011, 216-219) argues for the claim that disjunction is a fundamental term. But this claim is a controversial claim and not every philosopher endorses this. Since we need not go into detail about his argument for fundamental logical connectives, let us assume that the second answer satisfies Sparse. Even so, it satisfies neither Fitting nor Explanatory for the same reasons as the first answer. In terms of Fitting, T-theorem M2 makes a little progress

in capturing with intuitively correct usage in every conceivable circumstance. But that is clearly not enough, since T-theorem M2 makes 'Moore has hands' in English come out false in any circumstance where the fundamental base of our world is outside of the disjunction containing in T-theorem M2. For example, the fundamental base could be the mind of Berkeley's deity, but clearly the disjunction under consideration does not cover this. For Explanatory, again, T-theorem M2 is still not an apt explanation for 'Moore has hands' because of its inappropriate specificity and irrelevant information and failure in generalization. Just as the first answer, it is still unclear to me what T-theorem M2 tries to explain.

To conclude, the second answer satisfies Sparse, but it fails to obey Fitting and Explanatory.

The Third Answer: Possible Realizers

Finally, Schaffer considers the third answer that Sider's metaphysical semantics might take to escape from the multiple realizability problem. For this answer, it inserts all possible fundamental bases into ϕ in T-theorem M as follows.

> T-theorem M3: 'Moore has hands' is true in English iff the actual field-theoretic physics is such and such, or a physically possible distribution of field is such and such, or a particle-theoretic physics is such and such, or God's arrangement is such and such, or Cartesian evil demon is such and such, or Nozick's experience machine is such and such, or the mind of Berkeley's deity is such and such, or …

Continue to suppose that the fundamental story of our world is field-theoretic. The right bijunct of T-theorem M3 is filled in an infinitely open-ended disjunction of the actual field-theoretic physics, all physically possible bases of our actual field-theoretic physics, particle-theoretic physics, other merely possible fundamental bases, etc., that could possibly realize the truth that Moore has hands.

Let us evaluate the third answer. Finally, T-theorem M3 satisfies Fitting since it captures intuitively correct usage in every conceivable circumstance of both actual and possible fundamental bases. However, T-theorem M3 fails to obey Sparse and Explanatory. With respect to Sparse, the infinitely open-ended disjunction contained

in T-theorem M3 violates the use of fundamental terms. Recall that we assume that our world is fundamentally physical (actual physics), whether physics is field-theoretic or particle-theoretic. But when T-theorem M3 contains such a disjunction, the terms of God's arrangement, Cartesian evil demon, Berkeley's deity or whatever are clearly non-fundamental terms. This plainly violates the requirement of Sparse. As to Explanatory, T-theorem M3 is still far away from being an apt explanation for 'Moore has hands'. For the infinitely open-ended disjunction of T-theorem M3 remains full of overly specific and irrelevant details about what fundamentally realizes the truth that Moore has hands. Indeed, it is highly questionable whether such a disjunction has anything to do in *explaining* what we say by 'Moore has hands' in English.

To conclude, the third answer satisfies Fitting, but it fails to obey Sparse and Explanatory.

Putting all together: the above considerations show that none of the three suggested answers can satisfy more than one of the three conditions of adequacy. For the first and second answers, they satisfy Sparse but fails to obey Fitting and Explanatory. For the third answer, it satisfies Fitting but fails to obey Sparse and Explanatory. Thus, premise 3 is proved. If there are multiple realizable non-fundamental truths, Sider's metaphysical semantics cannot satisfy the three conditions of adequacy

6. The Impact on Sider's Analysis of the Necessity of a Micro-reduction Truth

Let us revisit the schematic argument outlined in the beginning of this chapter. The argument is as follows:

1. There are multiply realizable non-fundamental truths.

2. There are three conditions of adequacy for Sider's metaphysical semantics.

3. If there are multiple realizable non-fundamental truths, Sider's metaphysical

semantics cannot satisfy the three conditions of adequacy

---

4. *Therefore*, Sider's metaphysical semantics cannot handle multiply realizable non-fundamental truths.

Now, we are in the position to fully understand the argument. Let me summarize how Schaffer proves premises 1-3. He proves premise 1 by showing that there are multiply realizable non-fundamental truths. For instance, the truth that Moore has hands is of such a truth. To prove premise 2, he demonstrates that there are three conditions of adequacy for Sider's metaphysical semantics: Fitting, Sparse and Explanatory. He proves premise 3 by considering three possible answers that Sider's metaphysical semantics might take to respond to the multiple realizability problem and showing that none of them can satisfy more than one of the triple conditions of adequacy. Therefore, he concludes that Sider's metaphysical semantics cannot handle multiply realizable non-fundamental truths. This is the overall picture of Schaffer's multiple realizability problem.

In his "Reply to Jonathan Schaffer" (2013), Sider attempts to defend a fourth answer to Question in response to the multiple realizability problem. His attempted defence is constitutive of several tentative proposals, presuppositions and highly sophisticated arguments. For present purposes, we will not examine it and leave aside the question as to whether it succeeds in refuting the problem. At any rate, we need not make any conclusion for this. This is an important issue for proponents of Sider's metaphysical semantics. What is important to us instead concerns the fact that Sider admits that there is something wrong with his analysis of the necessity of a micro-reduction truth and the matter demands a modification of it. Sider writes:

> "Though each of these retreats is possible, I myself prefer a different one. I want to argue that the multiple realizability problem [of Schaffer] is primarily a problem for my account of modality, and not for my account of metaphysical truth-conditions per se, and thus that it can be solved by modifying my theory of modality (though I will not here attempt this modification)."

> "The multiple realizability problem [of Schaffer] is generated by the assumption that metaphysical truth-conditions must obey a principle of 'Fitting', as Schaffer

> calls it—that they 'must fit with intuitive usage by holding in most conceivable circumstances'. Now, something like Fitting is most certainly implied by my Humean theory of necessity: I say that the 'axioms' of a theory of metaphysical truth-conditions are ipso facto necessarily true (section 12.9) (and it's likely that the metaphysical truth-conditions we have been discussing would count as axioms). But I don't think that Fitting is required by my theory of metaphysical truth-conditions itself." (Sider 2013e, 768)

I will take the quotation of Sider to be a friendly invitation or an important motivational factor to propose a modified analysis of the necessity of a micro-reduction. For this reason, I will in the next chapter propose a modified version of Sider's reductive account of necessity to explain the necessity of a micro-reduction truth.

CHAPTER 3 A MODIFIED ACCOUNT OF NECESSITY

This chapter proposes and defends a modified account of necessity. In section 3.1, I develop a modified version of Sider's account of necessity. In section 3.2 and 3.3, I define the notion of 'express' in terms of Grice's theory of meaning and Lewis's reference magnetism. In section 3.4 and 3.5, I revisit my modified account and present a new reductive analysis of the necessity of a micro-reduction truth as well as an analytic truth and a natural-kind truth. In section 3.6, I consider a potential objection to my modified account and attempt to reply it.

1. A Modified Account

As we have seen in chapter 2, the multiple realizability problem of Schaffer seriously threatens the credibility of Sider's analysis of the necessity of a micro-reduction truth. More importantly, Sider admits that a revision of his account of necessity is needed to be done. In this chapter, I propose and defend a modified analysis of the necessity of a micro-reduction truth. I show that my proposed analysis can define the notion of a micro-reduction truth without appeal to Sider's metaphysical semantics. I then argue that analytic, natural-kind and micro-reduction truths are necessary iff they express the same states of affairs as logical truths. Since logical truths are necessary truths, analytic, natural-kind and micro-reduction truths are necessary truths as well. To begin with, let me present my proposed account as follows:

(*) A non-modal qualitative sentence S in a language L is a de dicto necessary truth iff S is a non-modal qualitative, de dicto and true sentence such that

(1) There is a state of affairs ϕ such that **i)** S in L expresses ϕ and **ii)** for some sentence S^* in some language of predicate logic L^* and for some interpretation I^*, S^* in L^* expresses ϕ under I^*, and

(2) S^* in L^* is a logical consequence of the set of mathematical truths in L^* under

I^* and metaphysical laws in L^* under I^*.

Frist, let us restrict my modified account to an account of *de dicto* necessity and leave de re necessity aside.[29] Let us also restrict my modified account to an account that applies only to non-modal qualitative sentences. A qualitative sentence is a sentence that expresses a qualitative state of affairs. A qualitative state of affairs is a state of affairs that does *not* concern any particular entity. Possible examples of a qualitative state of affairs are: the state of affairs that there is a cube, and the state of affairs that every emerald is green. Possible examples of a non-qualitative state of affairs are: the state of affairs that Trump is an American, and the state of affairs that Plato is the teacher of Aristotle. Roughly speaking, a qualitative sentence is a sentence that does not contain any name or concern any particular entity.

Second, let us suppose that there is a language of predicate logic L^* and an interpretation I^* that maps each predicate to a perfectly natural property or relation. Suppose also that L^* contains enough predicates and names such that i) for every object O, I^* maps a name in L^* to O, and ii) for every property or relation P, I^* maps a predicate in L^* to P. Thus understood, L^* under the interpretation I^* is a language that is closely skin to Sider's fundamental language.

To explain (2), I shall define mathematical truths in L^* under I^* and metaphysical laws in L^* under I^*. A mathematical truth ψ^* in L^* under I^* expresses a mathematical truth ψ in a language L iff i) ψ is a mathematical truth in L, and ii) ψ^* under I^* expresses the same state of affairs as ψ in L. Here, let us take L to be a language whose vocabularies consist of mathematical, English and set-theoretic expressions. A metaphysical law χ^* in L^* under I^* expresses a metaphysical law χ in English iff i) χ is a metaphysical law in English, and ii) χ^* under I^* expresses the same state of affairs as χ in English. The notion of 'expressing' here is understood as an interpretation function in a logician sense. But let us postpone our explanation of this idea until section 3.4. I will assume that mathematical truths are necessary on their own and there is no further feature to explain their necessity. We can understand a mathematical truth in L as a sentence in L that is true and concerns only mathematics.

[29] Unfortunately, this thesis cannot offer a non-circular account of de re necessity.

In regard to the notion of a metaphysical law, I will appeal to Sider's approach to metaphysical laws. His approach to laws of metaphysics is similar to Lewis's approach to laws of nature.[30] Sider's definition of a metaphysical law is presumably the following:

> (ML) A sentence *S* is a metaphysical law iff *S* is a regularity in the best system, where a best system is a deductive system that achieves the best combination of simplicity and strength.

The best combination of simplicity and strength is an important matter. If complexity is cheap, then special-science regularities counts as laws. If complexity is expensive, then only laws of logic count as laws. (Sider also defines the notion of a logical truth by appeal to Lewis's account of lawhood.[31] Recall section 1.1.) If complexity is made somewhere in between special-science regularities and laws of logic, then laws of physics count as laws. The central claim of his approach to metaphysical laws: if a cost of complexity is made intermediate between laws of logic and laws of physics, then laws of metaphysics count as laws and laws of physics drop out. This is Sider's approach to metaphysical laws. Here, I will follow Sider and borrow his approach to be my definition of a metaphysical law in *L*.

For the notion of a logical consequence, let us assume a model-theoretical account of a logical consequence, according to which a sentence *S* is a logical consequence of a set of sentences $K_1,K_2,\ldots K_n$ iff there is no model in which $K_1,K_2,\ldots K_n$ are true and *S* is false.

Claim (*) is the core claim of this thesis. To articulate (1), I shall begin by articulating what I mean by "state of affairs" and "express" in i). First, let me sketch my use of the term 'state of affairs'. A state of affairs is a way things are or a way things are not.[32] They are not literally true or false, but obtaining or not obtaining. Following Pollock (1984, 54), I will interpret obtaining and not obtaining as "truth-like properties". Prima facie, the state of affairs that Mark Twain is the author of

---

[30] See Lewis (1973; 1983; 1986; 1994), and Sider (2011 sections 3.1; 12.5).
[31] See Sider (2011, section 10.3).
[32] Again, I am indebted to Dan Marshall for this suggestion.

Adventures of Huckleberry Finn obtains iff it is true that Mark Twain is the author of Adventures of Huckleberry Finn. Likewise, the state of affairs that some koalas are poached eggs does not obtain iff it is false that some koalas are poached eggs. Let the term 'states of affairs' be understood as stated.

Second, let us understand 'express' as a semantic notion. Here, I assume that the 'expressing' relation between sentences and states of affairs is a semantic relation. To avoid circularity, I reject the 'expressing' relation to be modally understood. If one does not share the assumption of mine, then there is an alternative approach to take. For example, one could take 'express' as a primitive notion and gives examples to explain (*). Presumably, this approach causes a problem of ideological extravagance. In addition, it seems to me that meaning is not fundamental. Given these reasons, I shall explore two theories of meaning in the following two sections.

2. Paul Grice's Theory of Meaning

We are now dealing with this question: in virtue of what do sentences come to have the meanings that they have. There are two major theories of meaning in respect to this: Paul Grice's (1957) theory of meaning and David Lewis's (1984) reference magnetism. Let us examine the former first.

Grice makes a significant distinction between what sentences mean (sentence-meaning) and what speakers mean by utterances (speaker-meaning). The idea is that sentence-meaning concerns with the standard, literal, and ordinary meaning of a given sentence, while speaker-meaning concerns with what speakers intend to convey to the hearers by uttering a given sentence. Grice's distinction is a nice explanation for the cases of sarcasm and metaphor. Consider this situation. A philosophy student *x* invites his/her supervisor to write him/her a recommendation letter for the application of the graduate school. In the letter, the supervisor comments on his/her philosophical talent by writing this sentence:

(A) Student *x* is an excellent basketball player.

The sentence-meaning of (A) is the literal meaning that student *x* is an excellent

basketball player, while the speaker-meaning is what the supervisor wants to convey (B) to the readers.

(B) Student *x* is not good in doing philosophy.

Grice's distinction shows that sentence-meaning and speaker-meaning are greatly different from each other. It is plain to see that the sentence literally means (A) while the speaker by uttering (A) actually means (B).

Furthermore, Grice's account of meaning is a reductive account. On this view, the reduction contains two steps. First, sentence-meaning is explained in terms of speaker-meaning. As we have just seen in our example, the sentence (A) is used by the supervisor to mean something else, namely (B). Second, speaker-meaning is further explained in terms of the intentions of the speakers. In the above case, what the supervisor means by writing down (A) is to *intend* i) that the readers come to believe in (B), ii) that the readers recognize the supervisor's intention, and iii) that the readers come to believe in (B) on the basis of the recognition of the supervisor's intention. Putting all together, sentence-meaning is ultimately defined in terms of the intentions of the speakers.

Let us say a bit more on the notion of speakers' intention in Grice's theory of meaning. Obviously, speakers' intentions play an important role in explaining speaker-meaning, but other mental states—beliefs, desires, recognitions, etc.—also do. How does Grice himself think of the nature of mental states? Unfortunately, there is no explicit statement in his (1957) article to show which views he really holds. Devitt and Sterelny (1999, 149) suggest that Grice tacitly endorses a behavioristic view of mental states, according to which mental states are identified with behavioral dispositions.[33] But this issue goes beyond the scope of this thesis.

[33] See Grice (1957, 386-388). Grice remarks that there are linguistic and behavioral (non-linguistic in Grice own word) conventions for our intentions and actions. The linguistic convention nicely explains why an individual who asks for a pump at a fire does not mean a bicycle pump. The behavioral convention nicely explains why an individual who runs away from a bull wants to live. It is worth noting that Grice thinks that intention could be linguistic and behavioral. But what does "behavioral intention" mean? Prima facie, it just means "action". The fact that Grice identifies intention with action explains the reason why he uses "intend" and "act" in pretty much the same way; See also Grandy and Warner (2017 sec 6) for Grice's view of commonsense psychology.

So far so good. Given Grice's theory of meaning, we are in the position to define 'express' in terms of Grice's notion of speakers' intentions.

(3') A sentence *S* in a language *L* expresses a state of affairs ϕ iff *S* in *L* sentence-means ϕ.

(3.1') A sentence *S* in *L* sentence-means ϕ iff *S* in *L* speaker-means ϕ.

(3.2') A sentence *S* in *L* speaker-means ϕ iff i) a speaker by uttering *S* is to intend that the hearers come to believe in ϕ, ii) the hearers recognize the speaker's intention, and iii) the hearers come to believe in ϕ on the basis of the recognition of the speaker's intention.

However, this attempted analysis does not sound attractive. To fully explain it, I will probably have to *either* take mental states (intentions, beliefs, desires, recognitions, etc.) to be primitive notions *or* offer a reductive explanation for mental states. The former option threats the ideological parsimony of my proposed account, while the latter option is a challenging task.[34] The exact nature of mental states is a large project in philosophy of mind that I will not examine here. Either way, it seems to me that adopting Grice's theory of meaning is not a preferable option. But if it is not suitable for my account, what else? It seems to me that Lewis's reference magnetism is more

[34] Let us briefly consider what would happen if I adopt Grice's theory and offer a reductive explanation for metal states. In this case, I think that certain doctrines in philosophy of mind must be ruled out so as to be consistent with the central claim of this thesis that modality is not fundamental. For example, I cannot hold analytical behaviourism, according to which mental states are behavioral dispositions. If I adopt Grice's theory, then I will hold (1) and (2); if I hold behaviourism, then I will hold (3) as well.

(1) Sentence-meaning $=_{df}$ Speaker-meaning
(2) Speaker-meaning $=_{df}$ Mental state
(3) Mental states $=_{df}$ Behavioral disposition

However, this leads to an explanatory circuity. All I want to define overall is modality, but I end up analysing it in terms of disposition which itself is a modal notion. Disposition is generally accepted to be analysed in terms of counterfactual conditionals: an object *x* is disposed to *y* when *s* iff *x* would *y* if *x* is under the situation *s*. For example, salt is disposed to dissolve in water when it is put into water iff salt would dissolve in water if it is under the situation where it is put into water. So, I cannot hold behaviorism to explain the nature of metal states.

promising. Let us move on to the next section for a detailed survey of Lewis's account.

3. David Lewis's Reference Magnetism

An alternative approach to define 'express' is to adopt the notion of a *correct interpretation* of Lewis's (1984) reference magnetism. Given reference magnetism, I propose the following analysis to define 'express'.

> (3) A sentence S in a language L expresses a state of affairs ϕ iff for some interpretation I, I is a correct interpretation of L and S in L expresses ϕ under I.

This definition sounds more promising. To understand (3), we have to understand the notion of a correct interpretation of Lewis's reference magnetism. To introduce Lewis's reference magnetism, I shall outline Putnam's model-theoretic argument first.

In metasemantics, philosophers of language dispute over this problem: How do words get their meanings? To rephrase this, how do words refer to objects in the world?[35] According to Putnam's (1977; 1980; 1981 ch. 2) model-theoretic argument, there is no sufficient semantic determinacy to stick words onto their meanings. Putnam supports his argument by demonstrating that there are just too many ways in which assignments of meanings to words could be bizarrely made but our favored semantic theory can nevertheless be true. As a result, the semantic (referential) relations between words and meanings (objects in the world) are indeterminant. Although our linguistic conventions and behaviors are constraints on meaning-determination, they are largely dependent on our intentions or whatever mental-psychological states.[36] The semantic (referential) relations in question are still mind-dependent and thus indeterminant. Hence, Putnam's argument stands.

Let us further illustrate Putnam's argument. Consider our meaning-determining activity as a semantic theory T which consists of a set of sentences. Normally, T will

[35] In the debate between Lewis's and Putnam's, they prefer to use 'referent' of a word instead of 'meaning' of a word. But it would be permissible to take them as synonymy. For our propose let us stick to 'meaning'.

[36] Similarly, 'meaning-determination' and 'reference-fixing' are equivalent in meaning.

have more than one interpretation and each of them will map predicates and names onto objects in the world. The sort of mapping in question is to be understood as assignments of meanings. In each interpretation, meanings are assigned to each of the predicates and names of *T*. Now, the important question is: What is the best interpretation, or the correct interpretation? Putnam claims that the correct interpretation is the one that makes every sentence in *T* come out true.[37] Thus, he contends that though *T* contains some intuitively false sentences, *T* would still be logically consistent. To see this, let (D) abbreviate 'Some donkeys talk'. Undeniably, (D) is an intuitively false sentence. Nonetheless, we can arguably make (D) come out true in an interpretation by constructing a "deviant" interpretation where meanings are inappropriately assigned to words. For example, we can do it by assigning the set of birds to the predicate 'donkey' and the set of flying things to the predicate 'talk'. Given such a deviant assignment, the extensions of 'donkey' and 'talk' overlap. This makes (D) come out true under the deviant interpretation since some birds fly. On Putnam's view, there is nothing wrong with the deviant interpretation. At any rate, he probably does not think that such an interpretation is deviant at all. For him, all is required for an interpretation to be correct is merely that every sentence in *T* comes out true. However, we better say the deviant interpretation we have just made is *incorrect*. For we are pretty sure that such an interpretation by no means reflects what we really mean by 'donkey' and 'talk'. There must be something more for an interpretation to be correct than merely making every sentence in *T* comes out true.[38]

In response to Putnam's argument, Lewis's (1984) reference magnetism comes on the scene. According to Lewis, Putnam's argument can be refuted by adding an additional constraint on the notion of a correct interpretation. Instead of the constraint of making all sentences come out true, Lewis suggests that there is an *eligibility* constraint on the notion of a correct interpretation. On this view, a correct interpretation must assign as much as possible eligible meanings to predicates. More importantly, a correct interpretation must also maximize the sum of the degree of truth and the degree of eligibility, rather than doing the best of either one of them. The

[37] Here, let us ignore for a moment the distinction between uniqueness and non-uniqueness conceptions of 'correct interpretation'. We will discuss it later.

[38] For a more comprehensive illustration of Putnam's argument, see Putnam (1977; 1980; 1981 ch. 2), Lewis (1984), Sider (2001 sec. 2; 2011, 23-28) and Hale and Wright (1997).

following is a formulation of a correct interpretation of Lewis's reference magnetism:

> (RM) An interpretation *I* is a correct interpretation of a language *L* as used by a person *X* or a population *P* iff *I* maximizes the sum of the degree of truth and the degree of eligibility.

Before proceeding, I shall distinguish uniqueness and non-uniqueness conceptions of 'correct interpretation'.[39] According to the uniqueness conception, there is one and only one correct interpretation—*the* correct interpretation. According to the non-uniqueness conception, there are multiple correct interpretations. One advantage of the non-uniqueness conception is to handle cases of vagueness. For example, 'Sam is bald' will presumably have more than one correct interpretation. Since there is no clear-cut distinction between bald and not bald, there is no single correct interpretation of being bald or being not bald either. In general, the non-uniqueness conception is more realistic since natural languages are full of vague words. But things will get more complicated when it involves vagueness. So, the chief advantage of the uniqueness conception is simplicity. If one takes the uniqueness conception, (RM) should be amended into the following:

> (RM') An interpretation *I* is *the* correct interpretation of a language *L* as used by a person *X* or a population *P* iff *I* is *the one* that maximizes the sum of the degree of truth and the degree of eligibility.

In this thesis, I will stand on the non-uniqueness side and take (RM) to be an official formulation of a correct interpretation of Lewis's reference magnetism. But for simplicity, I will also assume that English has a unique correct interpretation.

(1) Fit-with-use

For an interpretation to be correct, it must make as much as possible sentences come out true. Lewis and Putnam both agree with this idea. But Lewis's constraint on

[39] I am grateful to Dan Marshall for pointing out the distinction between uniqueness and non-uniqueness conceptions of 'correct interpretation'.

'making sentences come out true' is largely different from Putnam's. It is argued that Putnam's constraint is trivial and unrestricted. As Putnam's argument showed, it is trivially easy to make the sentence 'Some donkeys talk' come out true under the deviant interpretation. On the contrary, Lewis's constraint is non-trivial and restricted. To maximize the degree of truth, Lewis contends that a correct interpretation must assign meanings to words on the basis of the linguistic behaviors of a person or a population. In other words, a correct interpretation must fit with the linguistic use of a person or a population. Let us take (1a) to be a proper understanding of Lewis's notion of fit-with-use.

(1a): An interpretation *I* maximizes the degree of truth iff *I* fits with the linguistic use of a person *X* or a population *P*.

On Lewis's view, the linguistic behaviors and conventions play an important role in assigning meanings to words. Let us illustrate it by an example. Let (C) abbreviate 'Every cat is an animal'. Now, we make two interpretations for (C) where we assign different meanings to the predicate 'cat'.

Interpretation 1 (I_1): we assign the meaning being computer to the predicate 'cat'.

Interpretation 2 (I_2): we assign the meaning being cat to the predicate 'cat'.

According to the linguistic conventions of *our* population in an ordinary sense, it seems to be a brute fact that I_1 does not reflect our use of (C). One way to think about it is that I_1 makes (C) come out false under this interpretation. (C) under I_1 is clearly false since the extensions of the set of computers and the set of animals do not overlap. Given this reason, we have a strong intuition to think that I_2 is a correct interpretation for (C). Prima facie, let us understand Lewis's notion fit-with-use as presented. (We will come up with a modified version of Lewis's notion of fit-with-use in section 3.6.)

(2) Eligibility

The main thesis of Lewis's reference magnetism is that meanings are not merely

determined by the linguistic conventions of us, but also by the objective nature of the world. As Lewis puts it,

> "Referring isn't just something we do. What we say and think not only doesn't settle what we refer to; it doesn't even settle the prior question of how it is to be settled what we refer to. Meanings—as the saying goes—just ain't in the head." (1984, 226)

> "I am inclined to favour a different kind of constraint proposed by G. H. Merrill. […] This constraint looks not to the speech and thought of those who refer, and not to their causal connections to the world, but rather to the referents themselves. Among all the countless things and classes that there are, most are miscellaneous, gerrymandered, ill-demarcated. Only an elite minority are carved at the joints, so that their boundaries are established by objective sameness and difference in nature. *Only these elite things and classes are eligible to serve as referents.* The world—any world—has the makings of many interpretations that satisfy many theories; but most of these interpretations are disqualified because they employ ineligible referents. When we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory. *It becomes once again a worthy goal to discover a theory that will come true on an eligible interpretation, and it becomes a daring and risky hope that we are well on the way toward accomplishing this*." (1984, 227 my emphasis)

Fit-with-use is just *one* of the determinant factors; the other factor is eligibility of meanings. Although there are multiple candidate meanings that are equally fitting-with-use, one meaning is more *eligible* to serve as a meaning than others. In short, a meaning *M* is eligible iff *M* is the most natural/elite kind—the one that carves the nature at the joints. To fully explicate Lewis's idea of eligibility, let us divide our explication into the notion of eligibility and the measurement of eligibility.

Let us talk about the notion of eligibility. Since Lewis interprets eligibility as naturalness, it would be permissible to think that natural properties are eligible meanings. Moreover, Lewis's naturalness is best thought of as a matter of degree, and so does eligibility. The more natural a property is, the more eligible a meaning is. More carefully, the meaning of a predicate *F* is more eligible iff the property expressed by *F* is more natural. But then what exactly is the notion of a natural property? The following quotation from Lewis's *On the Plurality of Worlds* is a nice introduction.

> "Sharing of [sparse properties] makes for qualitative similarity, they carve at the joints, they are intrinsic, they are highly specific, the sets of their instances are *ipso facto* not entirely miscellaneous, there are only just enough of them to characterise things completely and without redundancy.
>
> Physics has its short list of 'fundamental physical properties': the charges and masses of particles, also their so-called 'spins' and 'colours' and 'flavours', and maybe a few more that have yet to be discovered. […] What physics has undertaken […] is an inventory of the sparse properties of this-worldly things.
>
> If we have the abundant properties […] then we have one of them for each of the sparse properties. So we may as well say that the sparse properties are just some—a very small minority—of the abundant properties. We need no other entities, just an inegalitarian distinction among the ones we've already got. *When a property belongs to the small minority, I call it a natural property*." (1986, 60 my emphasis)

The most natural properties are also called the perfectly natural properties. They are in the elite or sparse minority of the abundant properties. Lewis suggests that the perfectly natural properties of this world are the most fundamental physical properties given by present-day physics. For example, the properties of being charge, spin, mass, quark flavors, colors, and so on, in this respect, are the perfectly natural properties of this world. Since naturalness comes in degrees, the degrees of naturalness decrease when the degrees of physical fundamentality decrease. For example, the properties of being a proton, an electron, a neutron, and so on, are less natural than the perfectly natural properties just mentioned; the properties of air, fire, water, and so on, are even lesser; and so forth.

Having elucidated the notion of eligibility, let us talk about the measurement of eligibility. For an interpretation to be correct, it must *maximize* the sum of the degree of truth and *the degree of eligibility*. But it is not immediately obvious about what it is meant by 'maximize the degree of eligibility'. The idea of maximization of the degree of eligibility can be put more carefully as follows.

> (2a) An interpretation *I* maximizes the degree of eligibility iff *I* assigns as much as possible eligible meanings to predicates.

Lewis suggests that eligibility can be measured by the definability of naturalness. A paradigm example: the meaning of being green is more eligible than the meaning of

being grue. Following Goodman (1955, ch. 3), being grue is generally defined in terms of either being green and being observed before 2100 or being blue and being not so observed yet. More carefully, an object is grue iff *either* it is green and observed before 2100 *or* it is blue and not yet observed. Given its conjunctive and disjunctive nature, it is plausible to think that 'being grue' has a longer chain of the definability in terms of naturalness. Other things being equal, it follows that the meaning of being green is *more eligible* than the meaning of being grue to serve as a meaning of the predicate 'is green'.

It might be misguided to think that less eligible or ineligible meanings like the meaning of being grue are unrelated to the objective nature of the world, and so they play no role in meaning-determination. Quite a contrary. Lewis does not intend to say that a correct interpretation *only* maximizes the eligibility of meanings. The main thesis of Lewis's reference magnetism is to maximize the *sum* of the degree of truth and the degree of eligibility. The following is Lewis's words:

> "*Ceteris paribus,* an eligible interpretation is one that maximises the eligibility of referents overall. Yet it may assign some fairly poor referents if there is good reason to. After all, 'grue' is a word of our language! *Ceteris aren't paribus*, of course; overall eligibility of referents is a matter of degree, making total theory come true is a matter of degree, the two desiderata trade off. The correct, 'intended' interpretations are the ones that strike the best balance." (Lewis 1984, 228-9)

This quotation is fairly clear except the word "trade off". I would thereby try to shed the light on this notion in what follows.

(3) Maximizing the Sum of the Degree of Truth and the Degree of Eligibility

As we have seen Lewis's quotation, the weighting of the degree of truth and the degree of eligibility is an important matter. Unfortunately, Lewis did not say much on this in his (1984). Here, I will illustrate how a true reference magnetism might maximize the sum of the degree of truth and the degree of eligibility:

> (3a) An interpretation *I* maximizes the sum of the degree of truth and the degree of eligibility iff *I* scores as high as possible.

Let us imagine that there is a scoreboard for the measurement of the degree of truth and the degree of eligibility. An interpretation starts from 0 point. Although there are infinitely many actual and possible sentences of a language, words of a language are presumably finite. An interpretation is said to score maximal points if it assigns meanings to every word of a language on the ground of the constraints of fit-with-use (1a) and eligibility (2a). This is how it works. The maximization of the degree of truth requires that the meaning-assignments under an interpretation fits with the linguistic use of a person or a population. Every time when an interpretation assigns a fitting meaning to a word that *ipso facto* reflects the linguistic use of a person or a population, then the interpretation scores one point and becomes +1. The maximization of the degree of eligibility requires that an interpretation must assign as much as possible eligible meanings to predicates. Every time when an interpretation assigns the most eligible meaning or the perfectly natural property to a predicate, then the interpretation scores one point and becomes +1. When an interpretation assigns two fitting or two most eligible meanings to words, it scores two points and becomes +2. And so forth. Let us demonstrate this idea by the following figure.

	Fit with use	Eligibility	
(I_1)	1	1	Every ***swan*** is charged.
(I_2)	2	1	Every electron is ***charged or not a cow***.
(I_3)	2	2	Every ***electron*** is ***charged***.

Suppose that there is a sentence “Every electron is charged” and call it (E). Suppose that there are three interpretations for (E) and each of them has different degrees of truth and eligibility. Note that this point-scoring is just an approximation. The idea of it is to make the notion of maximization more vivid and measurable.

In interpretation 1 (I_1), let us assign the meaning of being a *swan* to the predicate ‘is an electron’ and being charged to the predicate ‘is charged’. For the degree of truth, (E) comes out false under I_1, since I_1 makes (E) come out false under I_1, but nonetheless people assert (E). (I will explain more on this in section 3.6.) So I_1 assigns an unfitting

meaning to ‘is an electron’. But since I_1 assigns a fitting meaning to ‘is charged’, it scores 1 point and becomes +1. For the degree of eligibility, since ‘being an electron’ is more eligible than ‘being a swan’, this meaning-assignment under I_1 scores no point. But since I_1 assigns an eligible meaning to ‘is charged’, it scores 1 point and becomes +1. So far, I_1 has 1 point for the degree of truth and 1 point for the degree of eligibility. In total, I_1 scores 2 points.

In interpretation 2 (I_2), let us assign the meaning of being an electron to the predicate ‘is an electron’ and being *charged or not a cow* to the predicate ‘is charged. (E) come out true under I_2 since I_2 makes (E) come out true under I_2 and people assert (E). So, I_2 is considered as fitting-with-use and it assigns two fitting meanings to ‘is an electron’ and ‘is charged’. I_2 scores 2 points for the degree of truth. Nonetheless, I_2 does pretty bad for the degree of eligibility. For being *charged or not a cow* is less eligible than being charged to serve as a meaning of the predicate ‘is charged’. It is because by the chain of the definability in terms of naturalness, being *charged or not a cow* is longer than being charged. Since I_2 does not assign the most eligible meaning to ‘is charged’, it scores no point. But since I_2 assigns an eligible meaning to ‘is an election’, it scores 1 point. So far, I_2 has 2 points for the degree of truth and 1 point for the degree of eligibility. In total, I_2 scores 3 points.

In interpretation 3 (I_3), let us assign the meaning of being an electron to the predicate ‘is an electron’ and being charged to the predicate ‘is charged’. For the degree of truth, I_3 scores 2 points since it assigns two fitting meanings to ‘is an electron’ and ‘is charged’. For the degree of eligibility, I_3 also scores 2 points since it assigns two most eligible meanings to ‘is an electron’ and ‘is charged’. So far, I_3 has 2 points for the degree of truth and 2 points for the degree of eligibility. In total, I_3 scores 4 points. Thus understood, I_3 scores the maximal of the degree of truth and eligibility. That is to say, I_3 maximizes the sum of the degree of truth and the degree of eligibility. Therefore, I_3 is a correct interpretation for (E).

It is worth briefly noting that when we make interpretations for a language, say English, we do not consider a single sentence but every sentence in English. But, though, I admit that it is not clear how the counting method we illustrated will work given the infinity of sentences of English. Our illustration is merely a simplification and approximation.

4. Revisiting My Modified Account

After discussing Lewis's reference magnetism, we are now in the position to finalize my proposed account. In the beginning of section 3.1, I have proposed a modified version of Sider's account of necessity as follows:

> (*) A non-modal qualitative sentence *S* in a language *L* is a de dicto necessary truth iff *S* is a non-modal qualitative, de dicto and true sentence such that
>
> (1) There is a state of affairs ϕ such that **i)** *S* in *L* expresses ϕ and **ii)** for some sentence *S** in some language of predicate logic *L** and for some interpretation *I**, *S** in *L** expresses ϕ under *I**, and
>
> (2) *S** in *L** is a logical consequence of the set of mathematical truths in *L** under *I** and metaphysical laws in *L** under *I**.

(i)

We have already articulated (2) in the beginning of section 3.1. To articulate (1), I have sketched my use of the term 'states of affairs'. A state of affairs is a way things are or a way things are not. They are not literally true or false, but obtaining or not obtaining. In section 3.2, I have attempted to explain 'express' in terms of speakers' intentions in Grice's theory of meaning. But I gave up this approach since to take it will lead a dilemma where I have to *either* take mental states to be primitive notions *or* offer a reductive explanation for mental states. Either way, it causes problems. In section 3.3, I have outlined Lewis's reference magnetism. Now, we are ready to define 'express' in i) in terms of the notion of a correct interpretation of reference magnetism as follows.

> (3) A sentence *S* in a language *L* expresses a state of affairs ϕ iff for some interpretation *I*, *I* is a correct interpretation of *L* and *S* in *L* expresses ϕ under *I*.

The idea of (3) is to define the notion of 'express' as the notion of a correct

interpretation of reference magnetism. Intuitively, the sentence 'Every snow is white' expresses the state of affairs that every snow is white. But why not the state of affairs that some apples are red, or the state of affairs that some dogs are cute, or some others? This remains a mystery unless we are able to explain the 'expressing' relation between sentences and states of affairs. Thus, reference magnetism comes in. As we have seen in section 3.3, reference magnetism is a theory of a correct interpretation. According to reference magnetism, to say that an English sentence *S* expresses a state of affairs ϕ is to say that *S* expresses ϕ *under a correct interpretation of English*. This accounts for the 'expressing' relation in question on the basis of a correct interpretation of a language. To identify a correct interpretation of a language, we seek to achieve an interpretation that maximizes the sum of the degree of truth and the degree of eligibility. According to reference magnetism, the definition of a correct interpretation is the following:

> (RM) An interpretation *I* is a correct interpretation of a language *L* as used by a person *X* or a population *P* iff *I* maximizes the sum of the degree of truth and the degree of eligibility.

I assume that we have enough acquaintance with (RM) since we have fully explained each of essential elements of (RM) in the last section. So far, I have finished my clarification of i). Next, let us proceed to ii).

(ii)

Recall my proposed account.

> (*) A non-modal qualitative sentence *S* in a language *L* is a de dicto necessary truth iff *S* is a non-modal qualitative, de dicto and true sentence such that
>
> (1) There is a state of affairs ϕ such that **i)** *S* in *L* expresses ϕ and **ii)** for some sentence *S** in some language of predicate logic *L** and for some interpretation *I**, *S** in *L** expresses ϕ under *I**, and

(2) S^* in L^* is a logical consequence of the set of mathematical truths in L^* under I^* and metaphysical laws in L^* under I^*.

As we can see, there are several significant differences between i) and ii). First, S^* is stipulated to be stated in some language of predicate logic L^*. Presumably, S^* will be something like '$Fa \vee \neg Fa$', or '$\forall x(Cx \rightarrow Cx)$'. For example, '$F$' and '$C$' are predicates and '$a$' is a name, and they all get assigned different meanings under different interpretations, while 'x' a variable and it gets assigned to different objects under different variable-assignment functions. Second, because S^* is written in L^*, reference magnetism plays no role in determining which interpretation is a correct interpretation of L^*. So, we cannot define the notion of 'express' in ii) as the notion of a correct interpretation of reference magnetism. It is because there is no *correct* interpretation of L^* at all. That is to say, we are free to assign meanings to S^* in L^*, and any interpretation would be equally good.

A remark: In section 3.1, we have previously defined mathematical truths in L^* under I^* and metaphysical laws in L^* under I^*. A mathematical truth ψ^* in L^* under I^* expresses a mathematical truth ψ in a language L iff i) ψ is a mathematical truth in L, and ii) ψ^* under I^* expresses the same state of affairs as ψ in L. Here, let us take L to be a language whose vocabularies consist of mathematical, English and set-theoretic expressions. A metaphysical law χ^* in L^* under I^* expresses a metaphysical law χ in English iff i) χ is a metaphysical law in English, and ii) χ^* under I^* expresses the same state of affairs as χ in English. Of course, we need not repeat again how we have defined a mathematical truth ψ in L and a metaphysical law χ in English. (Readers interested in this may refer to section 3.1.) The point is that the notion of 'expressing' as used to define mathematical truths ψs* in L^* under I^* and metaphysical laws χs* in L^* under I^* is exactly the same as the notion of 'expressing' in ii). Hence, the following explanation is also applicable to them.

Here, let us introduce the basic idea to make sense of ii). Suppose that '$Fa \vee \neg Fa$' is a logical truth of predicate logic, and call it S^*. Suppose also that there is a sentence in English 'John runs or it is not the case that John runs', and call it S. Suppose also that there is a correct interpretation I of English. Let us assume that J is an

interpretation that maps being run onto '*F*' and maps John onto '*a*'. Given this setting, *S* in English under *I* expresses the same state of affairs as *S** in *L** under *J*. In other words, 'John runs or it is not the case that John runs' in English under *I* expresses the same state of affairs as '$Fa \vee \neg Fa$' in *L** under *J*. Here, interpretation *J* is thus understood in the vicinity of what logicians call 'interpretation function', even though they do not usually assign a sentence to a state of affairs as the way we do in this thesis. In general, a model of logic consists of a domain, an interpretation of names and an interpretation of predicates. In a model, an interpretation is a function that maps each name to its referent '**I(a)**' and each n-place predicate to n-place relation '**I(F)**'. To handle complex problems, logicians sometimes also need a "variable assignment"[40] $\boldsymbol{g}$ to temporarily assign variables to some objects in a domain, and a "denotation"[41] to represent the meaning of a term *t* with respect to a domain and an assignment function. At any rate, logicians almost never need to fully describe a model or fully specify the mapping of names to objects and n-place predicate to n-place relation. To describe a model for a given problem, they only need to specify the domain and then specify the meanings of names and predicates that actually appear to the problem. Similarly, we will not try to fully specify the mapping under *I**. It is because there are infinite many ways in which an interpretation *I** could be made. For our purposes, let us assume that there is one interpretation *I**, in which its mapping corresponds the mapping under a correct interpretation *I* of English, and *S** in *L** under *I** to express the same state of affairs as *S* in *L* under *I*.

To justify the assumption in question, let us go through an intuitively plausible example. Suppose that we want to make a claim that '$Fa \vee \neg Fa$' in *L** under an interpretation *J* expresses the same state of affairs as 'John runs or it is not the case that John runs' in English under a correct interpretation *I* of English. Probably, this claim will still be justified even though we do not specify the meanings of '*F*' and '*a*'. For it is quite obvious that there *is* an interpretation function that maps '*a*' to an object *John* and '*F*' to a property *being run*, such that '$Fa \vee \neg Fa$' in *L** under an interpretation *J* expresses the same state of affairs as 'John runs or it is not the case that John runs'

[40] $\boldsymbol{g}$ is a variable assignment $=_{df}$ $\boldsymbol{g}$ is a function that assigns each variable to objects in a domain.

[41] $\mathbf{[\alpha]_{D,g}}$ is the denotation of $\alpha =_{df}$ For some term α, i) if α is a name or a predicate, then $\mathbf{[\alpha]_{D,g} = I(\alpha)}$, ii) if α is a variable, then $\mathbf{[\alpha]_{D,g} = g(\alpha)}$.

in English under *I*. Of course, the mapping under *J* is very simple that we can easily write it out if we want to. But to handle complex sentences in English—like 'Every C_0 is a city', the exact mapping of interpretation functions *I** must be made in a more sophisticated way that we cannot even write it out. However, there is really no reason why we cannot assume that there *is* an interpretation *I** and *S** in *L** under *I** expresses the same state of affairs as 'Every C_0 is a city' in English under a correct interpretation *I* of English. Hence, let us assume that there is *I**. We hold this assumption in exactly the same way that we hold the assumption that there is an interpretation *J* and '$Fa \vee \neg Fa$' in *L** under *J* expresses the same state of affairs as 'John runs or it is not the case that John runs' in English under a correct interpretation *I* of English. So, given this reason, I take it to be a justifiable assumption.

To explain (1), all we need is i) a sentence *S* in a language *L* expresses ϕ, and ii) there is an interpretation *I**, whatever it might be, and *S** in *L** under *I** expresses ϕ. We have defined the notion of 'express' in i) as the notion of a correct interpretation of reference magnetism. So, we have explained i). Next, we have defined the notion of 'express' in ii) as an interpretation function of a model in the vicinity of the notion of logician interpretation function. And if what I have said in the previous paragraph provides a conclusive reason to justify the assumption, then we have also explained ii). To explain (2), we have defined mathematical truths ψs* in *L** under *I**, metaphysical laws χs* in *L** under *I** and the notion of logical consequence. Putting all together, I have defended my proposed analysis (*). As we can see, my analysis provides a modified reductive account of necessity. On this account, analytic, natural-kind and micro-reduction truths are necessary iff they express the same states of affairs as logical truths. Since logical truths are necessary truths, and analytic, natural-kind and micro-reduction truths are necessary truths as well.

Before moving on, I shall mention some reasons why Lewis's reference magnetism is worthy holding. My proposed analysis (3) will be flawed unless reference magnetism is a plausible theory of meaning. Firstly, reference magnetism can account for the genuine meanings of our words and give rigorous constraints on interpretations for languages. It solves the problem of semantic indeterminacy and avoids deviant interpretations. This is a remarkable contribution to philosophy of language. Secondly, reference magnetism can arguably account for the genuine

meanings of the terms whose meanings diverge radically in philosophical disputes, like the notion of knowledge and the notion of personal identity. (See Weatherson 2003; Sider 2009) The philosophical significance of reference magnetism is far-reaching. Thirdly, together with Compositionality Principle, reference magnetism can be extended to account for the genuine meanings of sentences. The basic idea of Compositionality Principle is that the meaning of a sentence *S* is determined by the meanings of every expression in *S* plus the syntactic structure of *S*. Given these theoretic merits, there is good reason to think that reference magnetism is a plausible theory of meaning.

So: given that reference magnetism is worth holding, we are justified in believing (3). So far, I have defined 'express' in i) as the notion of a correct interpretation. Also, I have showed that 'express' in ii) is best understood as an interpretation function of a model in the vicinity of the notion of logician interpretation function. At the core of my proposed account is that no term whatsoever in (1)-(3) is modally defined. Therefore, I have defended my proposed account. Next, I will employ it to account for the necessity of a micro-reduction truth.

5. Analyzing the Necessity of a Micro-reduction Truth

Let us work on the necessity of a micro-reduction truth. Let 'C_0' be a complex predicate which describes the actual microphysical states of New York City in every detail. On any reasonable account, 'Every C_0 is a city' is a necessary truth. Let 'micro-reduction truth' be a term of such a necessary truth. In what follows, I will clarify how my proposed account addresses the necessity of a micro-reduction truth.

Suppose that there is a sentence in some language of predicate logic L^* '$\forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \dots C_nx))$', where '$C_0x$', '$C_1x$', etc., are formulas, and 'C_0', 'C_1', etc., are complex predicates, and 'x' is a variable. This sentence is a logical truth, no matter what meanings of 'C_0', 'C_1', and so on amount to. Here, we could assume any reductive or nonreductive account of the notion of a logical truth, except for the modal account according to which a logical truth is a necessary truth. Otherwise, this causes circularity. For present purposes, let us assume a model-theoretical account of a logical truth according to which a sentence *S* is a logical truth iff *S* is true in all models.

According to the model-theoretical account, '$\forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \ldots C_nx))$' is true in all models, and thus a logical truth.

Now, we can employ my proposed analysis to account for the necessity of a micro-reduction truth. To improve readability, let us have abbreviations for *S* and *S**.

S =$_{df}$ 'Every C_0 is a city'

*S** =$_{df}$ $\ulcorner \forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \ldots C_nx))\urcorner$

Suppose that *S* is written in English and *S** is written in some language of predicate logic. Here is my proposed account.

(*) A non-modal qualitative sentence *S* in a language *L* is a de dicto necessary truth iff *S* is a non-modal qualitative, de dicto and true sentence such that

(1) There is a state of affairs ϕ such that **i)** *S* in *L* expresses ϕ and **ii)** for some sentence *S** in some language of predicate logic *L** and for some interpretation *I**, *S** in *L** expresses ϕ under *I**, and

(2) *S** in *L** is a logical consequence of the set of mathematical truths in *L** under *I** and metaphysical laws in *L** under *I**.

To define 'express' in i), we need (3).

(3) A sentence *S* in English expresses a state of affairs ϕ iff for some interpretation *I*, *I* is a correct interpretation of English and *S* in English expresses ϕ under *I*.

To define the notion of a correct interpretation of English, we need (RM).

(RM) An interpretation *I* is a correct interpretation of English as used by a person *X* or a population *P* iff *I* maximizes the sum of the degree of truth and the degree of eligibility.

Since we have already explained (2) before, let us go to (1) directly. Here, I will legitimate my analysis (1) by working on two things. i) I will illustrate how to make a *correct interpretation I* of English for *S* on the basis of reference magnetism. ii) I will justify that there *is* an interpretation *I**, and *S** in *L** expresses ϕ under *I**. After that, I will clarify the assumptions I need in my analysis (*).

A Correct Interpretation of English for *S*

To find out a correct interpretation of English for *S*, I shall illustrate reference magnetism by giving examples. Suppose that English has a unique correct interpretation. As stated in (RM), a correct interpretation of English is an interpretation that maximizes the sum of the degree of truth and the degree of eligibility. However, there is no point in assigning meanings to every sentence of English, in order to simply show how to work out a correct interpretation of English. For the sake of simplicity, let us only illustrate how to assign meanings to *a few* relevant words on the basis of reference magnetism. My intention is merely to show how a true reference magnetism might work. It should be also clear that when one interprets a language, one makes interpretations for every sentence of a language, rather than making interpretations for one single sentence.

Now, our objective is to find out a correct interpretation of English for *S*. To do this, we attempt to make interpretations of English. Suppose that in interpretation 1 (I_1) we assign being a city to the predicate 'is a city' and being C_0 to 'is C_0', and so on. As we stipulated, 'C_0' is a complex predicate for being the actual microphysical states of New York City. Presumably, 'C_0' contains many different atomic predicates inside, say 'is a proton', 'is an electron', 'is 3 kilograms', 'is unit negatively charged' and so forth. Suppose also that in I_1 we assign being a proton to the predicate 'is a proton' and being an electron to 'is an electron', and so on.

Interpretation 1
Predicate 'is a city' → being a city

Predicate 'is C_0' → being C_0

Predicate 'P_1', 'P_2', 'P_3'…'P_n'→ being M_1, M_2, M_3…M_n

Atomic predicate of C_0 'is a proton' → being a proton
Atomic predicate of C_0 'is an electron'→ being an electron
Atomic predicate of C_0 'p_1, p_2, p_3…p_n' → being m_1, m_2, m_3…m_n
} being C_0

In order to bring out the contrast, let us make another interpretation. Suppose that in interpretation 2 (I_2) we assign *being a book* to the predicate 'is a city' and being C_0 to 'is C_0', and so on. Once again, we stipulate that 'C_0' is a predicate for being the actual microphysical states of New York City. Again, there are many different atomic predicates inside 'C_0', say 'is a proton', 'is an electron' and so on. Suppose also that in I_2 we assign *being a kangaroo* to the predicate 'is a proton' and *being a dolphin* to 'is an electron', and so on. Suppose that I_1 and I_2 are only different in this regard. They both assign meanings to the rest of English words by the ordinary standard.

Interpretation 2

Predicate 'is a city' → being a book

Predicate 'is C_0' → being C_0

Predicate 'P_1', 'P_2', 'P_3'…'P_n'→ being M_1, M_2, M_3…M_n

Atomic predicate of C_0 'is a proton' → being a kangaroo
Atomic predicate of C_0 'is an electron'→ being a dolphin
Atomic predicate of C_0 'p_1, p_2, p_3…p_n' → being m_1, m_2, m_3…m_n
} being C_0

Now, it is time for weighting of the degree of truth and the degree of eligibility. In terms of the degree of truth, I_1 achieves a higher degree than I_2. Suppose that some competent speaker of English asserts the sentence 'Every proton is a subatomic particle', and call it (P). Clearly, (P) is true under I_1. Since (P) is true under I_1 and English speakers assert (P), I_1 fits with our use of (P). But I_2 does not fit with our use

of (P) since it makes (P) come out false and English speakers assert (P).[42] (P) is false under I_2 because when we assign being a kangaroo to the predicate 'is a proton', nothing is both in the extension of kangaroos and in the extension of subatomic particles. For the same reason, I_2 also makes other sentences containing 'city', 'proton', 'electron' come out false. Therefore, I_1 achieves a higher degree of truth than I_2. In terms of the degree of eligibility, I_1 still achieves a higher degree than I_2. Presumably, being a city and being a book are at a more or less similar degree of eligibility. It is hard to tell which one is more eligible than the other. However, it is easy to see that the meaning of being a proton is *more eligible* than the meaning of being a kangaroo to serve as a meaning of the predicate 'is a proton'. Likewise, the meaning of being an electron is *more eligible* than the meaning of being a dolphin to serve as a meaning of the predicate 'is an electron'. Given this, I_1 achieves a higher degree of eligibility than I_2.

To conclude, I_1 is a correct interpretation of English for S on the basis of reference magnetism. Given this, S in English under I_1 expresses the state of affairs ϕ.

Interpretation *I for *S****

Moving on to ii). Recall that S^* is a logical truth of the form:

$$\ulcorner \forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \ldots C_nx)) \urcorner$$

where 'C_0x', 'C_1x', etc., are formulas, and 'C_0', 'C_1', etc., are complex predicates, and 'x' is a variable. As we stipulated, S^* is written in some language of predicate logic L^*. Again, since S^* is written in L^*, reference magnetism cannot help here. There is basically no *correct* interpretation of L^* for S^*.

To solve this problem, I have previously suggested that interpretation I^* is best understood as an interpretation function of a model in the vicinity of the notion of logician interpretation function. Since we cannot fully specify the exact mapping of an interpretation function I^* for S^*, let us assume that there is an interpretation I^* in which

[42] Here we judge whether an interpretation fits with use on the basis of (Fitting) and (¬Fitting) to be outlined in section 3.6.

its mapping corresponds to the mapping of a correct interpretation I_1 of English, and S^* in L^* under I^* expresses the same state of affairs as S in English under I_1. My justification for this assumption is that there *is* an interpretation I^* indeed. (Recall our analogy with '$Fa \vee \neg Fa$'.) I admit that the mapping of I^* must be highly sophisticated and even infinitely long that we cannot ever write it out. But again, I do not see any reason why we cannot assume that there *is* an interpretation I^* and S^* in L^* expresses ϕ under I^*. I assume that if we have a computer program to compute it, then we will eventually find that there is such an interpretation I^*. Here, the idea is analogous with the situation where we try to find out a password for a mobile phone. No matter how complex the password is going to be, if we have a computer program to run though all possibilities, then the program will eventually work it out. For the same reason, we are legitimate to assume that *there is* an interpretation I^*.

So far, I have validated my analysis (1) by demonstrating two things. i) I have illustrated how to make a *correct interpretation I* of English for S on the basis of reference magnetism. ii) I have justified that there *is* an interpretation I^*, and S^* in L^* expresses ϕ under I^*. Next, let us clarify why my proposed analysis (*) needs to assume a Compositionality Principle.

Compositionality Principle

In what follows, I will consider an assumption that I need in my proposed account. The assumption is a Compositionality Principle. Surely, we interpret the meanings of each word of a language in virtue of a correct interpretation of reference magnetism and an interpretation function of a model. But such knowledge alone does not suffice to tell us anything about the meanings of sentences of a language. We need an extra principle to compose the meanings of words into the meanings of sentences. To do this, Compositionality Principle comes in.

> (CP) For any sentence S in a language L, the meaning of S in L is determined by its syntactic structure and the meanings of its constituents.

Given (CP), the meanings of each word of a language via a correct interpretation of

reference magnetism or an interpretation function of a model can be extended to account for meanings of sentences of a language. As we have seen, I have defined 'express' in i) in terms of the notion of a correct interpretation of reference magnetism and 'express' in ii) in terms of interpretation function of logic. Given these two things, we can in principle know meanings of each word in English and some language of predicate logic. Given (CP), it makes them up into meanings of sentences. Thus understood, we need three things to work out states of affairs expressed by sentences: reference magnetism, an interpretation function, and a Compositionality Principle.

To sum up, I have shown how my proposed analysis accounts for the necessity of a micro-reduction truth. On my view, 'Every C_0 is a city' is a necessary truth iff there is a state of affairs ϕ such that i) there is a correct interpretation I of English and 'Every C_0 is a city' expresses ϕ under I and ii) there is an interpretation I^* of L^*, '$\forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \dots C_nx))$' in L^* expresses ϕ under I^*.

Before proceeding, it might be a to mention some reasons why my proposed account is worthy holding. Compared with Sider's account of modality, there are a number of theoretical merits in favor of my proposed account. First, my proposed account gives a non-circular reduction of de dicto necessity and provides a modest improvement on the notion of de dicto necessity. Second, in comparison with Sider's account, my proposed account is equally parsimonious in terms of both ontological and ideological commitments. So far, there is no additional ontological entity or primitive notion added. Hence, if one finds that the metaphysical landscape of Sider's account of modality is attractive, one should find that my proposed account is attractive too. Third, I want to argue that my proposed account is *simpler* than Sider's account. It might be argued that Sider's analysis of the necessity of a micro-reduction truth is obscure and highly sophisticated in terms of its technical details, whereas my proposed account is simple and easy to understand. Why are analytic, natural-kind and micro-reduction truths necessary truths? My answer: they express the same states of affairs as logical truths, and since logical truths are necessary truths, they are necessary truths as well. Although I admit that there are some subtle analyses and technical idioms inevitably presented in my proposed account (in the light of the demand of clarity and precision), it is still much simpler than Sider's account in general. I believe that other things being equal, simplicity is a good candidate of theoretical merits. Last

but not least, the main advantage of my proposed account is that it can avoid the multiple realizability problem disclosed for Sider's account because it does not appeal to Sider's metaphysical semantics to explain the necessity of a micro-reduction truth. Thus, there are cogent reasons in favor of my proposed account.

Analytic Truths and Natural-kind Truths

Before finishing this section, it is worth mentioning that my proposed account can also give an account of the necessity of an analytic truth and the necessity of a natural-kind truth.[43] Consider the following sentences.

(I) Every bachelor is an unmarried man.
(II) $\forall x(Bx \rightarrow Bx)$.

(III) Every water molecule is a H_2O molecule.
(IV) $\forall x(Wx \rightarrow Wx)$.

Here, we can give a reductive analysis of the necessities of sentences (I) and (III), just in the same way that we have done for the necessity of a micro-reduction truth. I assume that we all know how it works. To avoid repetitiousness, let us simplify my analysis. Here is the basic idea. The sentence (I)/(III) is a necessary truth iff there is a state of affairs ϕ such that i) there is a correct interpretation I of English and (I)/(III) in English expresses ϕ under I and ii) for some logical truth (II)/(IV) in some language of predicate logic L^*, for some interpretation I^* of L^*, (II)/(IV) in L^* expresses ϕ under I^*. We probably need not go through again how to make a correct interpretation of English on the basis of reference magnetism, and how we are justified to assume that there is an interpretation I^* and (II)/(IV) in L^* expresses ϕ under I^*. I think that

[43] The set of natural-kind truths is a subset of necessary a posteriori truths. In Kripke's (1972) examples, necessary a posteriori truths are richer than that. For example, 'necessarily, Trump is human' is a necessary a posteriori truth. I am afraid that my proposed account cannot explain this kind of de re modal truth. Moreover, I am afraid that my account is also silent about 'Necessarily, Hesperus is Phosphorus'. As we have seen in section 3.1, my proposed account is only applicable to non-modal qualitative sentences and leaves de re necessity aside.

it would be permissible to leave it out.

Before we finish this section, let us explore an alternative approach to account for the necessity of an analytic truth and the necessity of a natural-kind truth. Here, I just describe this approach, but I myself do not advocate it. The alternative approach allows us to get rid of the stipulation that *S** is written in some language of predicate logic. We only need such a stipulation for explaining the necessity of a micro-reduction truth. It is precisely because in stating *S** we do not want to confine ourselves to the natural languages that we had. This accommodates the possibility that the fundamental story of the world goes beyond the scope of any natural language. After all, this is the crucial point of analyzing the necessity of a micro-reduction truth. But things become easier in accounting for the necessity of an analytic truth and the necessity of a natural-kind truth. We can simply take both *S* and *S** to be written in English.

(I) Every bachelor is an unmarried man.
(II') Every bachelor is a bachelor.

(III) Every water molecule is a H_2O molecule.
(IV') Every water molecule is a water molecule.

Let us take (I) as our working example. If (I) works, there is no reason why others do not. To account for the necessity of (I), one can simply argue that (I) expresses the same state of affairs as (II') under a correct interpretation of English. They express the same state of affairs since the predicate ‘is a bachelor’ expresses the same property as the predicate ‘is an unmarried man’ under a correct interpretation of English. Of course, we also need a Compositionality Principle to compose meanings of words into meanings of sentences. Finally, we might need to hold a structured theory of states of affairs to identify the sameness of (I) and (II'). Roughly speaking, according to some general version of structured theories of states of affairs, two states of affairs are identical iff they have the same structure and constituents in the same order. Let us assume for the sake of argument that states of affairs have objects and properties as constituents, and the ordered n-tuples represent the (alleged) structure of states of

affairs. Consider the following sentences.[44]

(1) The state of affairs that Marilyn Monroe is an actress.

(1*) <*a*, being *F*>

(2) The state of affairs that Norma Jean Baker is an actress.

(2*) <*a*, being *F*>

In (1*) and (2*), '*a*' is the referent of both Marilyn Monroe and Norma Jean Baker, and '*F*' is the property of being an actress. According to the toy account of structured states of affairs we assume here, the states of affairs (1) and (2) are identical since they contain the same object and property as constituents in the same order. For the names 'Marilyn Monroe' and 'Norma Jean Baker' refer to the very same object. On this view, the states of affairs (1) and (2) are the same state of affairs under different *names*.

Likewise, the states of affairs (3) and (4) are identical since they contain the same property as constituents in the same order.

(3) The state of affairs that every bachelor is an unmarried man.

(3*) <$\forall x$ <<*x*, being *F*>→<*x*, being *F*>>>

(4) The state of affairs that every bachelor is a bachelor.

(4*) <$\forall x$ <<*x*, being *F*>→<*x*, being *F*>>>

In (3*) and (4*), '*F*' is the property of being a bachelor. As we can see, (3) and (4) identical because the predicates 'is a bachelor' and 'is an unmarried man' actually express the same property. This claim can be justified by reference magnetism.

[44] This example is adopted from the entry "State of Affairs" in Stanford Encyclopedia of Philosophy.

Therefore, the states of affairs (3) and (4) are the same state of affairs under different *predicates*.

So: it proves that (I) expresses the same state of affairs as (II'). Again, to describe this approach is just for bringing out an alternative option to account for the necessity of an analytic truth and the necessity of a natural-kind truth. It is important to note that the alternative approach does not work for the necessity of a micro-reduction truth, since we do not want *S** to be written in English or any natural language. For there is no guarantee that natural languages that we had can cover the fundamental story of the world. I myself stand for my proposed account. i) For it is more flexible and general to account for the necessity of a micro-reduction truth. ii) And it requires no additional assumption for states of affairs to be structured and have constituents (since it is highly disputable), but still reach to the same conclusion.

6. An Objection to My Modified Account

It is argued that Lewis's reference magnetism is vague in terms of the notion of 'fit-with-use'. Previously, we said that an interpretation *I* maximizes the degree of truth iff the meaning-assignments under *I* fits with the linguistic use of a person *X* or a population *P*. But how exactly can we determine whether or not an interpretation fits with the linguistic behaviors of a person or a population? I admit that this objection is plausibly correct. The notion of fit-with-use is vague because so far we do not have a precise condition under which we can judge whether or not an interpretation fits with use. On this basis, I would like, in an exploratory spirit, to sketch a minor modification by specifying the conditions of fitting and non-fitting.

Revising the Notion of Fit-with-use

Suppose that we are going to identify a correct interpretation of English on the basis of reference magnetism. Suppose also that there is a correct interpretation *I* of English. To recapitulate the notion of a correct interpretation,

> (RM) An interpretation *I* is a correct interpretation of English as used by a person *X* or a population *P* iff *I* maximizes the sum of the degree of truth and the degree

of eligibility.

To maximize the degree of truth, Lewis contends that a correct interpretation must assign meanings to words on the basis of the linguistic behaviors of a person *X* or a population *P*. This suggests the following:

(1a) An interpretation *I* maximizes the degree of truth iff *I* fits with the linguistic use of a person *X* or a population *P*.

Here, we go to the degree of truth. Since we only want to show how to measure the degree of truth, let us pass over the part of eligibility for the sake of convenience. It may be helpful to have a working example. To simplify, let us suppose that there are only three sentences in our simplified English.

(1) Every swan is a bird.

(2) Socrates is a kangaroo.

(3) Plato has two legs, two hands, one nose, and is an ancient Greek.

Now, we are going to judge whether the interpretation *I* fits with the linguistic behaviors of a person's use of or a population's use of (1), (2) and (3). In principle, an interpretation can presumably manage the use of a population. This just involves more information and gets more complicated. For simplicity, let us stick to the use of a person. To make such a judgment, we need to have a *precise condition* under which an interpretation is considered as fitting with the use of a person. Here is a suggestion.[45]

(Fitting) An interpretation *I* fits with a person *X*'s use of a sentence *S* iff either

(i) *S* is true under *I* and *X* asserts *S*,

[45] I am indebted to Dan Marshall for this suggestion.

(ii) *S* is false under *I* and *X* never asserts *S*, or

(iii) *S* is true under *I* and *X* never asserts *S*.

To judge whether or not the interpretation I fits with a person's use of a sentence, we need a truth theory to determine the truth-conditions of sentences. Let us assume that there is a true truth theory and it generates T-theorems to determine the truth-conditions of sentences. Let us also assume that, given facts of the world, an interpretation of a language *L* can determine the truth-conditions for all sentences in *L* by the true truth theory we assume. Therefore, we shall assume that the followings are true:

T-theorem: *S* is true iff *P*.

(D) Given facts of the world, an interpretation of a language *L* can determine the truth-conditions for all sentences in *L*.

Suppose that the person in question is a competent speaker of English, Mary. Suppose that Mary asserts (1), and for (1) we assign *being a swan* to 'is a swan' and *being a bird* to 'is a bird' under *I*. Given condition (i), (1) is true under *I* and Mary asserts (1). Hence, we say that *I* fits with Mary's use of (1).

T-theorem 1: 'Every swan is a bird' is true iff every swan is a bird.

Suppose that Mary never asserts (2) and for (2) we assign *being Socrates* to 'Socrates' and *being a kangaroo* to 'is a kangaroo' under *I*. Given condition (ii), (2) is false under *I* and she never asserts (2). Given the world, Socrates is not a kangaroo. Hence, we say that *I* fits with her use of (2).

T-theorem 2: 'Socrates is a kangaroo' is true iff Socrates is a kangaroo.

Suppose that Mary never asserts (3) and for (3) we assign the object *Plato* to 'Plato',

being a leg to ‘is a leg’, *being a hand* to ‘is a hand’, *being a nose* to ‘is a nose’, *being an ancient* to ‘is an ancient’ and *being a Greek* to ‘is a Greek’ under *I*. Given condition (iii), (3) is true under *I* and she never asserts (3). Hence, we say that *I* fits with her use of (3).

> T-theorem 3: ‘Plato has two legs, two hands, one nose, and is an ancient Greek’ is true iff Plato has two legs, two hands, one nose, and is an ancient Greek.

The general idea of Fitting is simple. It contains two parts. The first part is that we employ a true truth theory to determine the truth-conditions of sentences and we assume that given facts of the world, an interpretation of a language *L* can determine the truth-conditions for all sentences in *L* by the true truth theory we employ. Therefore, we are in the position to know whether a sentence under an interpretation is true or false. The second part deals with facts about whether or not a person asserts a sentence. Let us assume that we can know these facts by observation. (Let us ignore Epistemology for a moment.) Here, we assume that a person’s asserting or not asserting a sentence counts as a part of the person’s linguistic behaviors.

Next, let us describe the notion of ¬Fitting. Imagine that there is an interpretation *J* where we assign *being an egg* to the predicate ‘is a bird’. Obviously, interpretation *J* makes (1) come out false, but suppose that Mary asserts (1) nevertheless. Hence, we say that (1) is false under *J* and she asserts it. How would my proposed modification account for such an unusual phenomenon? Well, we count it as non-fitting. We say that interpretation *J* fails to fit with her use of (1). Here, let us propose (¬Fitting):

> (¬Fitting) An interpretation *I* fails to fit with a person *X*’s use of a sentence *S* iff (i) *S* is false under *I* and *X* asserts *S*.

Imagine an extraordinary situation where Mary always asserts false sentences like (2) ‘Socrates is a kangaroo’, and (2) is clearly false under *I*. For these phenomena, we can think of Mary’s abnormal use of language as a disobedience to the rule of use. A language under an interpretation provides certain rules of use, and they are not only syntactic rules but also perhaps conventional or behavioral rules. As a language user,

Mary can choose to obey or disobey the rules of use. When she asserts (2) and (2) is clearly false under *I*, it seems to suggest that she *disobeys* the rule of use of (2) under *I*. Thus understood, it would be permissible to rephrase (¬Fitting) into (¬Obeying):

> (¬Obeying) An interpretation *I* fails to fit with a person *X*'s use of a sentence *S* iff *S* is false under *I* and a person *X* disobeys the rule of use of *S*.

Although (¬Fitting) and (¬Obeying) use different descriptions to specify the condition of non-fitting, they actually talk about the same thing.

This is a final remark. Lewis's notion of fit-with-use is best thought of as a matter of degrees. Thus, we want (Fitting) to accommodate this too. Here, we suggest (D-fitting):

> (D-Fitting) An interpretation *I* maximizes the degree of fitting with the linguistic use of a person *X* or a population *P* iff *I* scores the maximal point of (¬Fitting) or as high as possible.

We assume that an interpretation starts from 0 point of (¬Fitting). Since there are infinitely many actual and possible sentences of a language that count as (Fitting), it would be unwise to calculate it by addition. The point will definitely be infinite. Instead, we suggest using deduction. Every time when an interpretation counts as (¬Fitting), it loses one point and becomes -1. To put it in terms of (¬Obeying), every time when a person or a population disobeys the rule of use of a sentence, it loses one point and becomes -1. It goes without say that if an interpretation counts as (¬Fitting) or (¬Obeying) twice, it loses two points and becomes -2. And so forth. That is to say, the maximal or ideal degree of fitting-with-use is 0 point. This means that an interpretation has no (¬Fitting) or (¬Obeying) whatsoever. It is ideally and perfectly fitting. However, when we interpret multitudinous sentences of a language, it is rare to have the perfectly fitting. To compare two or more interpretations, a winning interpretation is an interpretation with the highest point.

Although this modified version of the notion of fit-with-use is not a fully developed version, this modification has specified the conditions under which an

interpretation is considered as fitting or non-fitting with the use of a person. And we have done our work. I hope that this simple modification has made the notion of fit-with-use clear.

CONCLUSION

I.

This thesis explores the metaphysical nature of necessity. In chapter 1, I have discussed Theodore Sider's (2011) new reductive account of metaphysical necessity and in particular focused on his analysis of the necessity of a micro-reduction truth. In chapter 2, I have presented and clarified the multiple realizability problem posed by Jonathan Schaffer (2013). In brief, he shows that Sider's metaphysical semantics cannot handle multiple realizability. His critique offers conclusive reason to think that there must be something wrong with Sider's metaphysical semantics. Since Sider's analysis of the necessity of a micro-reduction truth is mostly constitutive of his metaphysical semantics, something has gone wrong in either one (or both) if Schaffer's critique is successful. At any rate, it motivates a modified version of Sider's account of necessity.

The most important result of this thesis is presented in chapter 3. Given the multiple realizability problem, I have shown that the necessity of a micro-reduction truth can be reductively defined without appeal to Sider's metaphysical semantics. I have propounded and defended a modified version of Sider's account of necessity. On my account, analytic, natural-kind and micro-reduction truths are necessary if and only if they express the same states of affairs as logical truths. Since logical truths are necessary truths, analytic, natural-kind and micro-reduction truths are necessary truths as well. I have accounted for the necessity of a micro-reduction truth as follows. Suppose that the English sentence 'Every C_0 is a city' is a necessary truth. To address such a necessity, I have argued that 'Every C_0 is a city' is a necessary truth if and only if

> There is a state of affairs ϕ such that i) 'Every C_0 is a city' in English expresses ϕ and ii) there is an interpretation I^* of L^* and '$\forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \ldots C_nx))$' in L^* expresses ϕ under I^*.

To legitimate my proposed analysis, I have defended two central claims in section 3.5. (1) I have defined 'express' in i) in terms of the notion of a correct interpretation of Lewis's reference magnetism and have shown how to make a correct interpretation I

of English for 'Every C_0 is a city' on the basis of reference magnetism. (2) I have argued that it is reasonable to assume that there is an interpretation I^* in some predicate logic L^*, in which its mapping corresponds to the mapping of a correct interpretation I of English, and '$\forall x(C_0x \rightarrow (C_0x \vee C_1x \vee \dots C_nx))$' in L^* expresses ϕ under I^*.

I hope that I have successfully defended my claims (1) and (2). I also hope that my proposed analysis has provided a sufficiently clear and plausible explanation for the necessity of a micro-reduction truth, the necessity of an analytic truth and the necessity of a natural-kind truth.

II.

Let me finish this thesis by propounding some open problems that I have not fully explored here and I want to study in future work.

In this thesis, I leave aside an open problem of how two states of affairs are identical to each other. What is the individuation-condition of a state of affairs? As Mark Textor (2016, section 6) puts it, "a theory of states of affairs must answer the question how a state of affairs can "involve" objects and properties (relations) and combine them, if the objects don't exemplify the properties (stand in the relations)". For example, if one believes that states of affairs have a certain structure, then it is philosophically important to have an in-depth study of how a state of affairs is structured and its constituents bind together. In this thesis, I cannot spare myself dealing with this question.

Another question concerns a reductive analysis of de re necessity. It is superfluous for me to say that my modified account is an account of de dicto necessity and applies only to non-modal qualitative sentences. A completed account of necessity should be able to account for both de dicto and de re modality. To give an account of all modal sentences, I might base my modified account on Lewis's counterpart theory and define abstract possible worlds and individuals non-modally. I hope to study this project in the near future.

Finally, I am curious about whether my modified account can be somehow taken to be an account of the connection between the fundamental and the non-fundamental. I am optimistic about that. The very rough idea I can give is the following: to say that a non-fundamental sentence S_N holds in virtue of a fundamental sentence S_F is to say

that S_N expresses the same state of affairs as S_F. More carefully,

> (#) A non-fundamental sentence S_N holds in virtue of a fundamental sentence S_F iff there is a state of affairs ϕ such that i) S_N in non-fundamental language expresses ϕ and ii) there is an interpretation I^* of L^* and S_F in L^* expresses ϕ under I^*.

Suppose that we have a language of predicate logic L^* and an interpretation I^* that maps each predicate to a perfectly natural property or relation; and L^* contains enough predicates and names such that: for each object O, I^* maps a name in L^* to O, and for each property or relation P, I^* maps a predicate in L^* to P. If so, then we have a powerful language as acting like Sider's fundamental language. Given this language, we can in principle tell the non-fundamental story of the world by using only fundamental language.

An apparent problem is that my sketchy analysis will also face the multiple realizability problem posed Schaffer. This is how it can address the problem. For example, I can say that (1) the sentence 'Moore has hands' expresses the same state of affairs as the sentence '*M* instantiates *H*', where M is the actual microstate identified with Moore in L^* and H is the property of having hands in L^*. Frist, I assume that the notion of instantiation is a fundamental notion. This differs from Sider's metaphysical semantics. Second, the expressing relation is the biconditional: 'Moore has hands' ↔ '*M* instantiates *H*'. Third, my primary concern is how to specify *H*. Schaffer has proposed the triple conditions of adequacy for Sider's metaphysical semantics. Is my sketchy analysis able to meet the triple conditions of adequacy? I am optimistic about this but there must be some revisions, though I will not here attempt these revisions. I hope to work more on this in future.

BIBLIOGRAPHY

Adams, Robert. (1974). Theories of Actuality. *Noûs*, 8: 211–31; reprinted in Loux [1979], 190–209.

Armstrong, David. M. (1989). *A Combinatorial Theory of Possibility*. Cambridge University Press.

Ayer, Alfred Jules. (1936). *Language, Truth and Logic*. London: Victor Gollancz Ltd, 1946. 2nd edition.

Braddon-Mitchell, David & Jackson, Frank. (2007). *The Philosophy of Mind and Cognition*. Blackwell.

Cameron, Ross P. (2010). "The Grounds of Necessity." *Philosophy Compass* 5 (4):348-358.

Choi, Sungho and Fara, Michael. "Dispositions", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/dispositions/>.

Contessa, Gabriele. (2013). "Does your metaphysics need structure?" *Analysis* 73 (4): 715-21.

Correia, Fabrice & Schnieder, Benjamin. (2012). "Grounding: an opinionated introduction." In Fabrice Correia & Benjamin Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press. pp. 1.

Davidson, Donald. (1967). "Truth and meaning." *Synthese* 17 : 304-23.

— (1984). *Inquiries Into Truth And Interpretation*. Oxford University Press.

Dorr, Cian (2005). "What we disagree about when we disagree about ontology." In Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford University Press. pp. 234-86.

— (2013). "Reading writing the book of the world." *Philosophy and Phenomenological Research* 87 (3): 717-24.

— (2016). "To Be F Is To Be G." *Philosophical Perspectives* 30 (1):39-134.

Dorr, Cian & Hawthorne, John. (2013). "Naturalness". In Karen Bennett & Dean Zimmerman (eds.), Oxford Studies in Metaphysics: Volume 8. Oxford University Press. pp. 1.

Fine, Kit. (2001). "The Question of Realism." *Philosopher's Imprint* 1: 1–30.

— (2010). "Some Puzzles of Ground." *Notre Dame Journal of Formal Logic* 51: 97–118.

— (2012). "Guide to Ground." In Fabrice Correia & Benjamin Schnieder (eds.), *Metaphysical Grounding*. Cambridge University Press. pp. 37--80.

— (2013). "Fundamental truth and fundamental terms." *Philosophy and Phenomenological Research* 87 (3): 725-32.

Foster, John. (1976). "Meaning and Truth Theory," in Gareth Evans and John McDowell, eds., *Truth and Meaning*, Oxford: Clarendon Press.

Goodman, Nelson. (1955). *Fact, Fiction, and Forecast*. Harvard University Press.

Grandy, Richard E. and Warner, Richard. "Paul Grice", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL =

<https://plato.stanford.edu/archives/win2017/entries/grice/>.
Crane, Tim & Farkas, Katalin. (eds.) (2004). *Metaphysics: A Guide and Anthology*. Oxford University Press.
Grice, H. Paul. (1957). "Meaning." *Philosophical Review* 66 (3):377-388.
— (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.
Hale, Bob & Wright, Crispin. (1997). "Putnam's model-theoretic argument against metaphysical realism." In Bob Hale & Crispin Wright (eds.), *A Companion to the Philosophy of Language*. Blackwell. pp. 427--57.
Hirsch, Eli. (2013). "The metaphysically best language." *Philosophy and Phenomenological Research* 87 (3): 709-16.
King, Jeffrey. "Structured Propositions", The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/propositions-structured/>.
— (2007). *The Nature and Structure of Content*. Oxford University Press.
Kripke, Saul. (1972). "Naming and Necessity." In Donald Davidson and Gilbert Harman (eds.), Semantics of Natural Language, 253–355, 763–9. Dordrecht: D. Reidel. Revised edition published in 1980 as *Naming and Necessity* (Harvard University Press, Cambridge, MA).
Lewis, David. (1973) *Counterfactuals*. Cambridge: Harvard University Press.
— 1983, "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61: 343–377.
— (1984). "Putnam's paradox". *Australasian Journal of Philosophy* 62 (3):221 – 236.
— (1986). *On the Plurality of Worlds*. Wiley-Blackwell.
— (1986) *Philosophical Papers, Volume II*. New York: Oxford University Press.
— (1994) "Humean Supervenience Debugged," *Mind* 103: 473–390.
Lycan, William. (2008) *Philosophy of Language: A Contemporary Introduction*. 2nd ed. Routledge Contemporary Introductions to Philosophy. New York: Routledge.
Marshall, Daniel Graham. (2016). "A Puzzle for Modal Realism." *Philosophers' Imprint* 16.
McKay, Thomas and Nelson, Michael. "Propositional Attitude Reports", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/>.
Melia, Joseph. (2003). *Modality*. Mcgill-Queen's University Press.
Merricks, Trenton. (2013). Three comments on writing the book of the world. *Analysis* 73 (4): 722-36.
Miller, Alexander. (2007). *Philosophy of Language*. 2nd ed. Fundamentals of Philosophy (London, England). London; New York: Routledge.
Nolan, Daniel. (2011). "The extent of metaphysical necessity." *Philosophical Perspectives* 25: 313-39.
O' Connor, Timothy, and Montgomery, Nickolas. "Review of writing the book of the world by Theodore sider." *Notre dame philosophical reviews*. Accessed April 18, 2017. http://ndpr.nd.edu/news/41449-writing-the-book-of-the-world/.
Plantinga, Alvin (1974). *The Nature of Necessity*. Clarendon Press.
— (1976). "Actualism and possible worlds." *Theoria* 42 (1-3):139-160.

Pollock, John L. (1984). *The foundations of philosophical semantics*. Princeton University Press.
Putnam, Hilary (1977). "Realism and Reason." *Proceedings and Addresses of the American Philosophical Association* 50 (6):483-498.
— (1980). "Models and reality." *Journal of Symbolic Logic* 45 (3):464-482.
— (1981). *Reason, Truth and History*. Cambridge University Press.
— (1975). "The meaning of 'meaning'." *Minnesota Studies in the Philosophy of Science* 7:131-193.
Quine, W. V. O. (1936). "Truth by Convention". In *Journal of Symbolic Logic*. 77-106.
— (1951). "Two Dogmas of Empiricism." *Philosophical Review* 60 (1):20–43.
Robertson, Teresa and Atkins, Philip. "Essential vs. Accidental Properties." The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/essential-accidental/>.
Roca-Royes, Sonia. (2011a). "Essential Properties and Individual Essences." *Philosophy Compass* 6 (1):65-77.
— (2011b). "Essentialism vis-à-vis Possibilia, Modal Logic, and Necessitism." *Philosophy Compass* 6 (1):54-64.
Rosen, Gideon. (2010). "Metaphysical Dependence: Grounding and Reduction." In Bob Hale & Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford University Press. pp. 109-36.
Schaffer, Jonathan (2009). "On what grounds what." In David Manley, David J. Chalmers & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press. pp. 347-383.
— (2013). "Metaphysical semantics meets multiple realizability." *Analysis* 73 (4): 736-51.
Schwarz, Wolfgang. (2014). "Against Magnetism." *Australasian Journal of Philosophy* 92 (1):17-36.
Sider, Theodore. (1993). *Naturalness, intrinsicality, and duplication*. Ph.D. Dissertation, University of Massachusetts.
— (2001a). "Criteria of personal identity and the limits of conceptual analysis." *Philosophical Perspectives* 15 (s15):189-209.
— (2001b). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.
— (2002). "The Ersatz Pluriverse." *Journal of Philosophy* 99, no. 6: 279-315.
— (2003). "Reductive theories of modality." In Michael J. Loux & Dean W. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*. Oxford University Press. pp. 180-208.
— (MS) Beyond the Humphrey Objection. Unpublished manuscript. Available at http://tedsider.org/papers/counterpart_theory.pdf.
— (2009). "Ontological realism." In David John Chalmers, David Manley & Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press. pp. 384--423.
— (2011). *Writing the book of the world*. Oxford: Clarenson Press.
— (2013a). "Against parthood." In *Oxford studies in metaphysics.*, ed. Karen

Bennett, Dean W. Zimmerman. Vol. Volume 8, 237-293. Oxford: Oxford University Press.
— (2013b). “Précis of writing the book of the world.” *Philosophy and Phenomenological Research* 87 (3): 706-8.
— (2013c). “Replies to Dorr, Fine, and Hirsch.” *Philosophy and Phenomenological Research* 87 (3): 733-54.
— (2013d). “Summary.” *Analysis* 73 (4): 713-5.
— (2013e). “Symposium on writing the book of the world.” *Analysis* 73 (4): 751-70.
Soames, Scott. (2010). *Philosophy of Language*. Princeton: Princeton University Press.
Speaks, Jeff. “Theories of Meaning”, The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/meaning/>.
Szabó, Zoltán Gendler. “Compositionality”, The Stanford Encyclopedia of Philosophy (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/compositionality/>.
Tarski, Alfred. (1936). “The concept of truth in formalized languages.” In A. Tarski (ed.), *Logic, Semantics, Metamathematics*. Oxford University Press. pp. 152--278.
Textor, Mark. “States of Affairs”, The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/states-of-affairs/>.
Weatherson, Brian. (2003). What good are counterexamples? *Philosophical Studies* 115 (1):1-31.
Williams, J. Robert G. (2007). “Eligibility and inscrutability.” *Philosophical Review* 116 (3):361-399.
— (2015). “Lewis on Reference and Eligibility.” In Barry Loewer & Jonathan Schaffer (eds.), *A companion to David Lewis*. Wiley-Blackwell. pp. 367-382.