9-2022

Necessity, Essence and Analyticity: Toward an Analytic Essentialist Account of Necessity

Dongwoo Kim
The Graduate Center, City University of New York

NECESSITY, ESSENCE AND ANALYTICITY:
TOWARD AN ANALYTIC ESSENTIALIST ACCOUNT OF NECESSITY

by

DONGWOO KIM

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2022

NECESSITY, ESSENCE AND ANALYTICITY:

TOWARD AN ANALYTIC ESSENTIALIST ACCOUNT OF NECESSITY

by

DONGWOO KIM

This manuscript has been read and accepted by the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	**Melvin Fitting** Chair of Examining Committee
Date	**Nickolas Pappas** Executive Officer

Supervisory Committee:

Professor Graham Priest

Professor Michael Devitt

Professor Melvin Fitting

Professor Kit Fine

Professor David Papineau

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

NECESSITY, ESSENCE AND ANALYTICITY:

TOWARD AN ANALYTIC ESSENTIALIST ACCOUNT OF NECESSITY

by

DONGWOO KIM

Advisor: Professor Graham Priest

Some truths could not have failed to hold. Such are called *metaphysically necessary* truths. As Michael Dummett once aptly formulated, the philosophical problem about necessity is twofold: what makes necessary truths necessarily true and how do we recognize them as such? This dissertation aims to address these questions by developing and defending a novel account of necessity, which has the following three main theses: (1) the necessity of a statement about an entity is established as a consequence of a general principle implying that if the entity is a certain way then it is necessarily that way and the fact that the entity is indeed that way; (2) the general principle is analytic in the sense that it is derivable from analysis of relevant concepts; and hence (3) the necessity of the statement can be known by investigating what the entity is in fact like and by conceptual analysis. I call this new account of necessity *analytic essentialism*.

The two main questions about necessity are given both logical and philosophical treatments. A logical analysis of modal statements is given from the perspective of truthmaker semantics. The analysis is developed into a formal truthmaker semantics for modal statements, and the soundness and completeness results for a well-known family of systems of normal propositional modal logic are established. On the basis of this formal analysis, paradigmatic examples of necessary statements are examined to address the substantive question of how exactly their necessity is established and known. Along the way, the present dissertation brings together a set of recent insights from philosophical logic, the philosophy of language, the philosophy of mind and the philosophy of science to shed new light on certain under-appreciated connections among necessity, identity, definition and explanation.

ACKNOWLEDGEMENTS AND PERMISSIONS

I would first like to express my deepest gratitude to Professor Graham Priest, my dissertation supervisor, for all of the invaluable guidance and support since the very conception of the present project. Throughout the difficult time of the pandemic, he always made himself available to provide me encouragement and insightful comments on my work in progress and on multiple versions of this dissertation. I also owe a great debt of gratitude to the other members of my committee. The seed of the present project was first planted in Professor David Papineau's seminar on mental representation. Chapter 3 of this dissertation, which was originally written as a term paper for the seminar, has taken its current shape through a number of extensive discussions with him. His recent work with Marion Godman and Antonella Mallozzi on super-explanation has also been the source of chapter 6. Professor Papineau's probing comments on these and other parts of the dissertation have been invaluable. I would also like to thank Professor Kit Fine. His contribution far exceeded what is typically expected of an outside reader. Chapters 1 and 2 were developed out of my term paper for his seminar on truthmaker semantics and they have greatly benefited from his penetrating comments. He also provided many exacting comments on various parts of the dissertation and warned me away from many confusions. In fact, chapter 4 was written in response to an important criticism that Professor Fine raised in the prospectus defense. I thank Michael Devitt for his comments and discussions especially on chapters 4, 6, and 7 and also for calling my attention to some of the subtle issues concerning biological essentialism. His critical comments have greatly helped me improve the clarity of the aforementioned chapters. Finally, I am grateful to Professor Melvin Fitting for his comments and suggestions especially on technical matters of chapters 1, 2, 4, and 7. His shrewd observations and comments have helped me better understand some of the crucial issues in quantified modal logic and improve the rigor of the technical matters of this dissertation.

Many thanks are also due to Professors Saul Kripke, Romina Padro, David Rosenthal, Stephen Neale, Sergei Artemov, and Inkyo Chung for their interesting discussions and helpful suggestions at various stages of this project; Brian Porter, Ebubekir Deniz, and Esma Kayar for listening to my

half-baked ideas and providing useful feedbacks in Professor Priest's weekly meetings; the participants of the Logic and Metaphysics Workshop for providing helpful comments on talk versions of some of this material; Tomasz Zyglewicz, Sai Ying Ng, and Yale Weiss for reading and providing comments on earlier drafts of some of the chapters; Ilhwan Yu, Hanjong Soh, and Seyong Bae for vibrant discussions at the weekly research meetings during the time of the pandemic.

Finally, I wish to express my gratitude to my family. I thank my mother, father, grandmother, and sister, Kyungja Kim, Kwangsoo Kim, Soonja Jeon, and Nahyeon Kim, for their unwavering love and encouragement. And I thank my wife, Inhye Lee, and my son, Dohyeon Kim, for being the infinite source of love and joy at all times.

This dissertation contains material that I have previously published and I acknowledge the publisher's permission to reuse this material. With very minor changes, Chapter 3 was published as:

> Kim, D. (2021). Explanation and modality: on why the Swampman is still worrisome to teleosemanticists. *Synthese* 199: 2817-2839.

This paper is available from the publisher at `https://doi.org/10.1007/s11229-020-02913-8`. It is reprinted by permission of the publisher under license 5375380031040 with the following required acknowledgement:

> Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, *Synthese*, "Explanation and modality: on why the Swampman is still worrisome to teleosemanticists," Dongwoo Kim, COPYRIGHT 2021.
> (doi: 10.1007/s11229-020-02913-8)

TABLE OF CONTENTS

List of Figures and Tables

Introduction

1. Background and overview

Some truths could not have failed to hold. Such are called *metaphysically necessary*.[1] For example, consider the truth that every square is four-sided or the truth that Socrates is Socrates. There is a strong sense in which these truths are necessary. For how could there possibly be a square that does not have four sides, or how could Socrates be not Socrates himself but something else? In contrast, some things in the world could have been otherwise. Presumably, John F. Kennedy would not have died on November 22, 1963, had Lee Harvey Oswald failed to shoot him. Hence the truth that Kennedy died on November 22, 1963 is only contingent, but not necessary.

Metaphysical necessity is now widely recognized as a distinct kind of modality. However, the questions of *what makes necessary truths necessary* and of *how we recognize them as such* have been much contested in analytic philosophy. There have been two major approaches to this problem in analytic philosophy. Until 1950s, the received view was conventionalism according to which necessary truths are ones that are analytic, i.e., guaranteed to be true by linguistic conventions; hence we come to recognize their necessity by reflecting upon our use of language (Ayer, 1936; Carnap, 1947; Wittgenstein, 1956). This view, though not without defenders, is no longer popular in the contemporary philosophical community. The change of attitude is largely due to Quine's (1951; 1960) attack on the notion of analyticity and to Kripke's (1980) work on the necessity aposteriori.

Quine (1960) famously argues that it is dubious that the truth of a certain sentence can somehow be explained solely in terms of its meaning. For, in general, it relies not just on the meaning but also on what the world is like. Once this much is granted, it is only one step away from the refutation of conventionalism: no truth is analytic, *a fortiori*, no necessary truth is analytic.

[1] Throughout this dissertation, I shall use modal terms, such as 'modality', 'necessity', 'possibility', etc., to mean metaphysical notions unless otherwise indicated.

A couple of decades later, Kripke in his *Naming and Necessity* (1980) argued that certain empirical discoveries are necessary (if true at all). Empirical science is supposed to discover that water is H_2O. Given that this is true, it appears that water must necessarily be H_2O. This and other examples seem to provide counterexamples to conventionalism. For, in the case just mentioned, one certainly cannot know the truth of the identity simply by reflecting upon the meanings of the terms involved.

Though Kripke's examples were widely considered intuitively convincing, they only deepened the philosophical problem about necessity. Consider the case of water. Experience seems to tell us only that water *is* H_2O; but then how do we know that it *could not have been otherwise*? Moreover, it appears that some properties (e.g., its chemical composition) belong to water necessarily, while others only contingently. But why is this? What distinguishes necessary properties of water from contingent ones?

Inspired by Kripke's discussion and Fine's pioneering paper (1994), many recent philosophers have been attracted to essentialism, which maintains that necessary truths are those that hold by virtue of the essences of things (Lowe, 2012; Hale, 2013; Godman et al., 2020). On this view, for instance, it is necessary that water is H_2O because it is in the essence of water that its molecules are composed of hydrogen and oxygen atoms in the ratio of 2:1. Hence knowledge of this necessary truth can be obtained by investigating what water is like in its essence. More generally, essentialism holds that we can obtain knowledge of necessity by investigating the essences of things in the world.

Something important is missing, however, from both of these approaches. Conventionalism and essentialism both expound what makes a necessary truth *true*. But not much attention has been paid on the question of how exactly its *necessity* arises from the purported grounds of its *truth*. It thus seems that conventionalism and essentialism at best provide a characterization of a class of truths; but it is unclear that they can offer a genuinely explanatory account of the two main problems about necessity.[2] The present dissertation aims to address this lacuna by developing and defending a novel account of metaphysical necessity.

[2]See Fine (2002, p.265) for a similar worry.

In the dissertation, the two main questions about necessity will be given both a logical and philosophical treatment. The core insight behind the present treatment is what I call *Kripke's Principle*: a proposition is necessary just in case all the ways in which the proposition might be false are not real possibilities. This principle gives a formal, logical answer to the metaphysical problem about necessity. What makes a proposition necessary has the logical function of excluding all the ways in which the proposition might be false from the realm of real possibilities; and so we recognize the necessity of a proposition by explaining away all the ways in which the proposition might be false as unreal. In chapters 1 and 2, this logical analysis is developed into a formal truthmaker semantics for modal statements and the soundness and completeness results for a well-known family of systems of modal logic are established. These results provide a formal vindication that my analysis of modal statements is logically adequate.

On the basis of this logical analysis of modal statements, I give a substantive, philosophical treatment of the main problem. What exactly is it, in other words, that excludes the ways in which a proposition might be falsified from the realm of real possibilities? In chapters 3 and 4, I argue that the answer lies partly in analyticity and partly in what the world is like. I examine some of the paradigmatic examples of theoretical identity statements and argue that behind the necessity of each such statement lies a certain analytic principle telling us that the statement could not possibly be false if it is true at all. This principle in conjunction with the truth of the statement implies that the statement is necessary.

Chapters 5, 6 and 7 offer a critique of a rival account, namely the orthodox essentialist approach to necessity. Chapter 5 gives a brief review of the basic ideas of the essentialist approach. Then I discuss two major metaphysical accounts of essence, namely the explanation-based and definition-based accounts, respectively in chapters 6 and 7. Chapter 6 argues that the explanation-based account fails to explain necessity as having its metaphysical source in essence. Chapter 7 argues that the definition-based account does not provide a plausible epistemological account of essence. The upshot of these three chapters will thus be that the essentialist approach faces a version of what Christopher Peacocke calls *the integration challenge*: given the explanatory goal of essentialism, no metaphysics of essence can plausibly be reconciled with an epistemology of essence.

Chapter 8 brings the findings of the previous chapters together into a brief sketch of what I call the *analytic essentialist* account of necessity. On this account, the necessity of a statement about an entity is established as a consequence of two things: a general analytic principle implying that if the entity is a certain way then it could not have been otherwise and the fact that the entity is indeed that way. Hence necessity is neither a sheer creation of linguistic conventions nor a feature built into the natures of the entities in the world; rather, it is a joint product of the language and the world. Accordingly, knowledge of necessity is obtained through the joint effort of philosophical analysis and scientific investigation. Both conventionalism and essentialism have gone too far in thinking that metaphysical necessity is reducible, respectively, to analyticity and to essence. The truth cuts across the two views. It shall also be discussed how the analytic essentialist account applies not just to theoretical identity statements to necessary truths in general. Then I conclude with a brief discussion of how the analytic essentialist account differs from the orthodox essentialist approach.

It should be noted at the outset that the present dissertation does not aim to give an account of *every* kind of modal truths. Analytic essentialism is primarily intended as an account of what one might call *simple* necessary truths, i.e., truths of the form 'Necessarily P', where 'P' has no modal operators. The typical examples of such statements include 'It is necessary that water is H_2O', 'It is necessary that $\{$Socrates$\}$ contains Socrates as its sole member', etc. What is missing from the dissertation is a treatment of *nested* modal truths, i.e., truths that contain a string of modal operators. Logical and philosophical problems concerning nested modal truths inevitably involve the question of what the logic of metaphysical modality is. And it seems to me that this issue should not be prejudged before we arrive at an adequate understanding of metaphysical modality itself. I shall thus leave a proper treatment of nested modal truths as a task for another day.

2. Chapter summaries

Chapter 1. Truthmaker Semantics for Modal Logic, Part 1

What makes a necessary truth not merely true but necessarily so? In this chapter, I approach this question about necessity through the lens of truthmaker semantics. A *truthmaker* for a proposition P is a state in virtue of which P is true. A truthmaker for P is said to be *exact* if it is entirely

relevant to the truth of P. From the perspective of truthmaker semantics, the main problem can be restated thus: what are the exact truthmakers for $\Box P$ ("Necessarily P")? I formulate what I call *Kripke's Principle*: for any proposition P, "Necessarily P" is true just in case the apparent possibilities of P's being false are not real. This principle gives us a formal answer to our main problem: an exact truthmaker for "Necessarily P" has the logical function of excluding the truthmakers for "Not P" from the realm of real possibilities. I also argue that this analysis exactifies of the Kripke semantics for modal logic in the sense specified above.

Chapter 2. Truthmaker Semantics for Modal Logic, Part 2

In this chapter, I develop a formal truthmaker semantics for modalized statements and establish the soundness and completeness results for a well-known family of systems of normal modal propositional logic, namely K, D, B, 4, K4, S4 and S5. These results provide a formal vindication that the proposed analysis of modalized statements is logically adequate.

Chapter 3. Explanation and Modality: On why the Swampman is still worrisome to teleosemanticists

In this chapter, I start addressing the substantive, philosophical question of what the truthmakers for a necessary proposition really are. What exactly is it, in other words, that excludes the ways in which a proposition might be falsified from the realm of real possibilities? To this end, I examine the Swampman objection to teleosemantics as a case study. Many have thought that Davidson's (1987) Swampman scenario raises a serious problem for teleosemantics. For it describes an apparently possible scenario where there are ahistorical creatures with beliefs, which contradicts the theory. In response, Papineau (2001) argues that the Swampman scenario creates no problem for the theory—just as no one would object to the H_2O theory of water simply on the ground that it appears to be possible that water has another chemical composition, say XYZ. In this chapter, I argue against Papineau that there is a crucial difference between the two cases. In the case of the H_2O theory, I argue, the XYZ scenario can be explained away with appeal to an analytic principle that can be established from the theoretical concept of chemical substance; and like considerations

apply to other well-known cases of theoretical identity statements. In the case of teleosemantics, by contrast, it is unclear how such a general principle can be established on the ground of its basic conceptual commitments. This case study thus throws an insight as to how necessary statements and knowledge thereof are possible. In each such case, there is an underlying general principle telling us that it is necessary if true at all; so, from knowledge of its actual truth we may obtain by simple inference knowledge of its necessity.

Chapter 4. The Kripkean Explanation of Aposteriori Necessity: In the case of identity statements concerning chemical substances

This chapter further analyzes the nature of general principles that underlie necessary statements. Salmon (1979) has famously argued that it is a mistake to think that such principles are analytic by showing that some of the attempts to derive them solely from conceptual grounds fail. His argument has been influential and many have followed him in this regard. As opposed to Salmon, I argue that the underlying principles are analytic in the sense that they are derivable from relevant concepts. Taking the case of the identity of water to H_2O as the main case, I show that the underlying general principle—that if a chemical substance has a certain chemical composition then it could not have had another chemical composition—is derivable from the theoretical definition of a chemical substance as a kind of matter with a unique chemical composition.

Chapter 5. Critique of Essentialism—Preliminaries

This chapter clears up the ground before launching into criticism of the essentialist approach of necessity. On this approach, necessity has its source in essence and hence knowledge of necessity derives from that of essence. So, it faces two obvious prior questions: what is essence and how do we recognize it? This chapter presents three constraints that any prior account of essence ought to satisfy for it to be adequate for the ultimate explanatory aim of the essentialist approach. On the basis of the constraints, I formulate what I take to be the main challenge for the essentialist approach: to reconcile a plausible metaphysics of essence with a reasonable epistemology in a way that is adequate for the purpose of providing a metaphysical and epistemological account of

necessity in terms of essence.

Chapter 6. Critique of Essentialism—On the explanation-based account of essence

This chapter critically examines a version of the explanation-based account. Marion Godman, Antonella Mallozzi and David Papineau (2020) have recently argued that the essence of a kind consists in the *super-explanatory property*—a single property that is causally responsible for a multitude of commonalities shared by the instances of the kind. And they argue that this super-explanatory account of essence offers a principled account of aposteriori necessities concerning kinds. I examine two main arguments of GMP that the super-explanatory property of a kind is metaphysically necessary and argue that they both are fallacious. Along the way, a general problem also emerges that applies to any account that tries to explicate the notion of essence in terms of an explanatory relation.

Chapter 7. Critique of Essentialism—On the definition-based account of essence

This chapter concerns the definition-based account of essence which seeks to understand the notion of essence on the model of real definition. In this chapter, I argue that this account fails to provide a plausible epistemology of essence. In the course of argument, I also examine a recent objection to the definition-based account that it fails to explain necessity as having its metaphysical source in essence. It shall be seen that the objection, though not entirely groundless, is not decisive.

Chapter 8. Toward an Analytic Essentialism

I bring the findings of the previous chapters together into a novel account of necessity, which I call *analytic essentialism*. Kripke's Principle offers a logical solution to the problem of necessity: what makes a statement necessary is that which excludes all the ways in which it might be false from the realm of real possibilities; and so we know that a statement is necessary by explaining them away as unreal. What is it, then, that excludes the apparent possibilities of a statement's being false from the realm of real possibilities, and how do we explain them away? Close examinations of well-known cases of necessary statements reveal that the answer lies partly in analyticity and

partly in what entities in the world are in fact like. There are analytic general principles telling us that certain statements are necessary if true at all. These principles, when conjoined with the actual truths of the statements, yield the necessity of the statements. Hence necessity arises from the interplay of the language and the world, and knowledge of necessity is obtained through the joint effort of philosophical analysis and scientific investigation.

CHAPTER 1

Truthmaker Semantics for Modal Logic, Part 1

1. Introduction

What makes a necessary truth not merely true but necessarily so? In this chapter, I approach this question about necessity through the lens of truthmaker semantics. The aim of this chapter is to give a brief exposition of the basic ideas of truthmaker semantics and to present the informal analysis of modal operators and some auxiliary notions that motivates the formal definitions to be presented in the next chapter. A *truthmaker* for a proposition P is a state in virtue of which P is true. A truthmaker for P is said to be *exact* if it is entirely relevant to the truth of P, and *inexact* if it contains as part an exact truthmaker for P. What then are the exact truthmakers for $\Box P$ ("Necessarily P") and for $\Diamond P$ ("Possibly P)? The present proposal is based on the informal analysis that an exact truthmaker for $\Box P$ is a state that *bans* the exact truthmakers for $\neg P$ ("Not P"), and an exact truthmaker for $\Diamond P$ is a state that *allows* an exact truthmaker for P. It shall be seen that the proposal offers an *exactification* of the Kripke semantics (Kripke, 1959, 1963) in Fine's (2014) sense: it analyzes the accessibility relation between possible worlds in terms of the banning and allowing relations between the constituent states and thereby gives an account of "truth at a world" in terms of exact truthmaking.

2. Basic ideas of truthmaker semantics

The truthmaker principle states that every true proposition is made true by something (Armstrong, 2004; Williamson, 1999). That something in virtue of which a proposition is true is called a *truthmaker*, or *verifier*, for the proposition. Consider, for example, the proposition that the Empire State Building is between 33rd and 34th Streets in Manhattan. This proposition is true because of the presence of the building in the designated location. So, it can reasonably be taken as a verifier for the proposition.

What then of false propositions? According to the unilateral truthmaker principle, a false proposition is one that has no verifiers. The bilateral truthmaker principle, on the other hand, says that every false proposition is made false by something, i.e., a *falsemaker*, or *falsifier*; for example, the proposition that the Empire State Building is on 42nd Street in Manhattan is made false, say, by its absence in that area. Without further argument, we shall adopt the bilateral truthmaker principle.

Below we shall generically call a *state* whatever can play the role of a verifier or of a falsifier. It is standard to assume that states have a mereological structure; so, a state may be part of another and two states can be put together into a single state Fine (2017b). It is also important to note that we are not working with the factual notion of state, according to which a state is something that *in fact* obtains. For we would like to say, for example, that the proposition that the Empire State Building is on 42nd Street in Manhattan—which is in fact false—would be verified by the presence of the building there—a state which does not in fact obtain. Moreover, we may also allow "impossible" states, such as the presence of the Empire State Building both between 33rd and 34th Streets and on 42nd Street. Strictly speaking, we should say that a verifier for a proposition is a state that *would* make the proposition true if it obtained; and similarly for the notion of falsifier.

With the mereology of states, we can now distinguish two notions of verification. On the one hand, we say that a state is an *exact* verifier for a proposition just in case it is a verifier that is entirely relevant to the truth of the proposition, and similarly for an *exact* falsifier. On the other hand, a state is said to be an *inexact* verifier for a proposition just in case it contains an exact verifier for the proposition as its part (not necessarily proper). For example, the presence of the Empire State Building between 33rd and 34th Streets in Manhattan can reasonably be taken to be an exact verifier for the proposition that the Empire State Building is there. However, the presence of the building together with five pigeons on the top would only be considered an inexact verifier for the proposition; for some part of it—namely, the presence of five pigeons—is not relevant to the truth of the proposition. Similar examples can easily be given for exact and inexact falsifiers. For the purpose of this paper, we shall content ourselves with these illustrative examples and leave the notion of relevance at an intuitive level.

3. Formal machinery: state space

Let us formalize these basic ideas of the exact truthmaker semantics. The standard way is developed in a series of papers by Fine (2014; 2016; 2017b). A *state space* is a complete partial order $\langle \mathscr{S}, \sqsubseteq \rangle$, i.e., $\sqsubseteq$ is a reflexive, transitive and anti-symmetric binary relation on $\mathscr{S}$ such that any subset S of $\mathscr{S}$ has a least upper bound, written $\bigsqcup S$. When $S = \{s_1, ..., s_n\}$, we shall sometimes write $s_1 \sqcup ... \sqcup s_n$ to mean $\bigsqcup S$. Intuitively, $\mathscr{S}$ is the set of states and $\sqsubseteq$ is parthood relation on states. $\bigsqcup S$ is the least state that has every member of S as its part. The completeness requirement guarantees that the $\sqsubseteq$-least element uniquely exists in $\mathscr{S}$, which we shall call the *bottom state*. Given any set S and T of states, we shall let

$$S \sqcup T = \{s \sqcup t : (\exists s)(\exists t)(s \in S \text{ and } t \in T)\}.$$

Notice that $S \sqcup T$ is empty if and only if S or T is empty. In case $S = \{s\}$,

$$S \sqcup T = \{s \sqcup t : t \in T\}.$$

Example 1. Let $\mathscr{S} = \{\bot, s, t, \top\}$, where $\bot$ is the bottom state, s is the presence of the Empire State Building between 33rd and 34th Streets, t is the presence of five pigeons on the top of the building and $\top$ is the presence of both. Then we may say that s and t are both *part* of $\top$; or equivalently, $\top$ *extends* both s and t. So we have: $s \sqsubseteq \top$ and $t \sqsubseteq \top$. Also, $\top$ is the least state extending both s and t. So we have: $\top = s \sqcup t$. This state space $\Sigma_0 = \langle \mathscr{S}, \sqsubseteq \rangle$ can be represented diagrammatically thus:

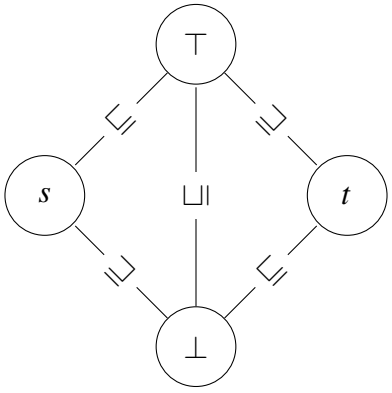

FIGURE 1.1. Diagrammatic representation of Σ_0

The parthood relation is represented in the obvious way, and s and t are considered incomparable.

Now we consider how the formal framework can be used to give a semantics for classical propositional connectives, $\neg$ (negation), $\wedge$ (conjunction) and $\vee$ (disjunction). Let $P_1, P_2, ...$ be a countably infinite list of *propositional variables*. The *well-formed formulas* (or simply, *formulas*) are constructed in the usual way, using connectives $\neg, \wedge, \vee$, respectively representing negation, conjunction and disjunction. We shall use P, Q, R, ... as metavariables for propositional variables, and A, B, C, ... for formulas in general. For any formulas A and B, $A \supset B$ is defined as $(\neg A \vee B)$.

Let Γ be a set of formulas. $At(\Gamma)$ is the set of propositional variables occurring in some formulas of Γ and $Fml(\Gamma)$ be the set of formulas whose atomic subformulas are in $At(\Gamma)$. A *model* $\mathfrak{A}$ of Γ is an ordered triple $\langle \mathscr{S}, \sqsubseteq, v \rangle$, where $\langle \mathscr{S}, \sqsubseteq \rangle$ is a state space and v is a valuation taking each state $s \in \mathscr{S}$ to a pair $\langle [s]^+, [s]^- \rangle$ of subsets of $At(\Gamma)$.[1] Intuitively, $[s]^+$ is the set of propositional variables that are exactly verified by s, and $[s]^-$ is the set of propositional variables that are exactly falsified by s. We require that for each $P \in At(\Gamma)$, there is at least one $s \in \mathscr{S}$ such that $P \in [s]^+ \cup [s]^-$.

Given a model $\mathfrak{A}$ of Γ, we define the notions of a state s *exactly verifying*, and *exactly falsifying*, a formula A in $\mathfrak{A}$ (in symbol, respectively, $\mathfrak{A}, s \Vdash^+ A$ and $\mathfrak{A}, s \Vdash^- A$) by simultaneous recursion:[2]

$$
\begin{array}{rcl}
\mathfrak{A}, s \Vdash^+ P & \Leftrightarrow & P \in [s]^+, \text{ for atomic } P; \\
\mathfrak{A}, s \Vdash^- P & \Leftrightarrow & P \in [s]^-, \text{ for atomic } P; \\
\mathfrak{A}, s \Vdash^+ \neg A & \Leftrightarrow & \mathfrak{A}, s \Vdash^- A; \\
\mathfrak{A}, s \Vdash^- \neg A & \Leftrightarrow & \mathfrak{A}, s \Vdash^+ A; \\
\mathfrak{A}, s \Vdash^+ A \wedge B & \Leftrightarrow & s = s_1 \sqcup s_2, \text{ for some } s_1, s_2 \text{ with } \mathfrak{A}, s_1 \Vdash^+ A \text{ and } \mathfrak{A}, s_2 \Vdash^+ B; \\
\mathfrak{A}, s \Vdash^- A \wedge B & \Leftrightarrow & \mathfrak{A}, s \Vdash^- A \text{ or } M, s \Vdash^- B; \\
\mathfrak{A}, s \Vdash^+ A \vee B & \Leftrightarrow & \mathfrak{A}, s \Vdash^+ A \text{ or } \mathfrak{A}, s \Vdash^+ B; \\
\mathfrak{A}, s \Vdash^- A \vee B & \Leftrightarrow & s = s_1 \sqcup s_2 \text{ for some } s_1, s_2 \text{ with } \mathfrak{A}, s_1 \Vdash^- A \text{ and } \mathfrak{A}, s_2 \Vdash^- B.
\end{array}
$$

[1] The current definition of a valuation may appear to deviate from the standard one, according to which a valuation is a (possibly partial) function from a set of propositional variables to the pairs of sets of states. Obviously, however, the two definitions are equivalent; in this connection, see the two equivalent ways of stating the semantic clauses for propositional connectives below.

[2] The following clauses were originally given in Van Fraassen (1969) and rediscovered in the truthmaker semantics literature by Fine (2017b). Here and below I use the symbol $\Leftrightarrow$ to mean *if and only if* in metalanguage, and the symbol $\Rightarrow$ to mean *if then* in the obvious way.

A moment's reflection should reveal the plausibility of these clauses. Note that these clauses specify an exact verifier for a formula recursively in terms of exact verifiers of its subformulas. We will often omit the mention of a model $\mathfrak{A}$ when it does not sacrifice clarity. Given an m-model $\mathfrak{A}$ of Γ and a formula $A \in Fml(\Gamma)$, we let

$$|A|^+ = \{s \in \mathscr{S} : s \Vdash^+ A\};$$

$$|A|^- = \{s \in \mathscr{S} : s \Vdash^- A\}.$$

Using this notation, we can rewrite the above clauses for exact verification and falsification thus:

$$\begin{aligned}
s \Vdash^+ P \quad &\Leftrightarrow \quad s \in |P|^+, \text{ for atomic } P;\\
s \Vdash^- P \quad &\Leftrightarrow \quad s \in |P|^-, \text{ for atomic } P;\\
s \Vdash^+ \neg A \quad &\Leftrightarrow \quad s \in |A|^-;\\
s \Vdash^- \neg A \quad &\Leftrightarrow \quad s \in |A|^+;\\
s \Vdash^+ A \wedge B \quad &\Leftrightarrow \quad s \in |A|^+ \sqcup |B|^+\\
s \Vdash^- A \wedge B \quad &\Leftrightarrow \quad s \in |A|^- \cup |B|^-;\\
s \Vdash^+ A \vee B \quad &\Leftrightarrow \quad s \in |A|^+ \cup |B|^+;\\
s \Vdash^- A \vee B \quad &\Leftrightarrow \quad s \in |A|^- \sqcup |B|^-.
\end{aligned}$$

A state s is an *inexact verifier* for A (notation: $s \triangleright^+ A$) in $\mathfrak{A}$ if and only if it extends an exact verifier for A in $\mathfrak{A}$, i.e.,

$$\text{there is } s' \in \mathscr{S} \text{ such that } s' \sqsubseteq s \text{ and } s' \Vdash^+ A.$$

The notion of inexact falsification (notation: $s \triangleright^- A$) is defined analogously. Notice that, for any state s and any formula A, if $s \Vdash^+ A$ then $s \triangleright^+ A$, and if $s \Vdash^- A$ then $s \triangleright^- A$.

Example 2. Let P be the proposition that the Empire State Building is between 33rd and 34th Streets and Q be the proposition that there are five pigeons on the top of the Empire State Building. Let $\Gamma = \{P \wedge Q\}$. Then we define a model $\mathfrak{A}_0$ of Γ to be $\langle \mathscr{S}, \sqsubseteq, v \rangle$, where $\Sigma_0 = \langle \mathscr{S}, \sqsubseteq \rangle$ is as defined in Example 1 and v is such that $[\bot]^+ = [\bot]^- = [s]^- = [t]^- = [\top]^+ = [\top]^- = \emptyset$ and $[s]^+ = \{P\}$ and

$[t]^+ = Q$. From the clauses above, one can easily check that $s \Vdash^+ P$, $s \Vdash^- \neg P$, $\top \triangleright^+ P$, $t \Vdash^+ Q$, $t \Vdash^+ P \vee Q$, $\top \triangleright^+ P \vee Q$, and $\top \Vdash^+ P \wedge Q$.

4. Exactification of the Boolean semantics

It will be helpful to consider the truthmaker semantics in terms of *exactification* in the sense of Fine (2014), the idea that given any inexact verifier (falsifier) for a proposition, there must be an underlying exact verifier (falsifier). A Boolean valuation in the standard semantics can itself be considered a state in a model that verifies those formulas that are true—and falsifies those that are false—under the valuation. Here the relevant notions of verification and falsification are inexact, as Fine (2014, pp.551-552) notes, because a Boolean valuation assigns truth-values to *all* formulas; so, given any formula A, the Boolean valuation—conceived as a state—may have parts that are irrelevant to the truth and falsity of A. According to exactification, therefore, each Boolean valuation can be represented as a state in a model that contains as its part an exact verifier for each formula that is true—and an exact falsifier for each formula that is false—under it.

How can this be done? Here we consider it in relation to multi-valued semantics for its importance for the subsequent discussion and for its intrinsic interest. Let $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ be a model for a set Γ of formulas. A state s is said to be *atomically sound* if and only if no propositional variable P in $At(\Gamma)$ is such that

$$s \triangleright^+ P \text{ and } s \triangleright^- P,$$

and to be *atomically unsound* otherwise. We also say that s is *atomically complete* if and only if for all propositional variable $P \in At(\Gamma)$,

$$s \triangleright^+ P \text{ or } s \triangleright^- P;$$

otherwise, s is *atomically incomplete*. One notable feature of the present semantics is that a model may have atomically inconsistent, and atomically incomplete, states. Due to this feature, there is a natural correspondence between the current semantics and the four-valued semantics as defined in Belnap (1977), where each formula can be assigned one of the following four values: True, False, Both and Neither. For each state s in a model $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ for a set Γ of formulas, we may define

the corresponding four-valued assignment φ_s for propositional variables in $At(\Gamma)$ by setting

$$\varphi_s(P) = \begin{cases} \{T\} & \text{if } s \triangleright^+ P \text{ and } s \not\triangleright^- P; \\ \{F\} & \text{if } s \not\triangleright^+ P \text{ and } s \triangleright^- P; \\ \{T,F\} & \text{if } s \triangleright^+ P \text{ and } s \triangleright^- P; \\ \emptyset & \text{if } s \not\triangleright^+ P \text{ and } s \not\triangleright^- P. \end{cases}$$

A *Belnapian* valuation $\overline{\varphi_s}$ is a valuation extending φ_s to all formulas according to the scheme as given in Belnap (1977). Then it is not difficult to see that for all formulas $A \in Fml(\Gamma)$, $s \triangleright^+ A$ if and only if $T \in \overline{\varphi_s}(A)$, and $s \triangleright^- A$ if and only if $F \in \overline{\varphi_s}(A)$. In this sense, each state s in a model corresponds to a valuation $\overline{\varphi_s}$ in the four-valued semantics.

The standard two-valued Boolean valuations can be viewed as a special kind of four-valued valuations, namely the ones that assign only $\{T\}$ or $\{F\}$ to formulas. Since such valuations are easily shown to correspond to atomically sound and complete states, Boolean valuations can be represented as atomically sound and complete state in models. More precisely: given any Boolean valuation for Γ, we can construct a model of Γ in which there is an atomically sound and complete state s such that for every formula A, A is true (false) under the Boolean valuation if and only if there is an exact verifier (falsifier) of A under s; and conversely, any atomically sound and complete state in a model for Γ uniquely determines a Boolean valuation for Γ.[3] According to the current truthmaker semantics, therefore, the notion of truth (falsity) under a Boolean valuation can be exactified as follows: a formula A is true (false) under a Boolean valuation if and only if A has an exact verifier (falsifier) under an atomically sound and complete state in a model.

Let us then consider how we may give an exact truthmaker semantical account of some of the logical notions of the Boolean semantics, such as tautology and tautological consequence. Since the former may be defined in terms of the latter in the usual way—a formula is a tautology if and only if it is a tautological consequence of the empty set of statements—we only consider how tautological consequence may be defined. Here again, we may gain a certain insight into the

[3]This result is essentially due to van Fraassen Van Fraassen (1969). See also Fine (2014) for a similar result for the Kripke semantics for intuitionistic logic.

workings to the proposed truthmaker semantics by considering the problem in relation to multi-valued semantics.

The intuitive notion of consequence is understood in terms of truth preservation: a formula A is a consequence of a set Γ of formulas if and only if A is true under every interpretation in which every formula in Γ is true. This notion becomes ambiguous in multi-valued semantics. For what exactly is meant by a formula's being true? In the four-valued semantics, for example, it may be taken to mean either $\{T\}$ or $\{T,F\}$. In multi-valued semantics, this ambiguity is resolved by defining consequence in terms of the preservation of *designated* values. In n-valued semantics, schematically, the set $\mathscr{D}$ of designated truth-values is any nonempty subset of the n possible truth-values, and A is said to be a *consequence* of Γ if and only if A is assigned a value in $\mathscr{D}$ by every n-valued valuation that assigns a value in $\mathscr{D}$ to every formula in Γ. The definition, of course, yields different notions of consequence depending on how many truth-values we have and what we take $\mathscr{D}$ to be. In the case of the four-valued semantics with $\mathscr{D} = \{\{T\},\{T,F\}\}$, for example, the definition yields the consequence of first-degree entailment, or the so-called FDE-consequence.

Now, the present truthmaker semantics provides a unifying semantical framework in which the consequences of some of the best-known multi-valued logics, such as FDE, the strong three-valued logic K3 of Kleene (1938, 1952) and the logic of paradox LP of Priest (1979), can be defined.[4] To list up the truthmaker semantical definitions, we have:

(FDE) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ for $\Gamma;A$ and for all states $s \in \mathscr{S}$, $s \rhd^+ A$ if $s \rhd^+ B$ for all $B \in \Gamma$.

(K3) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ and for all atomically sound states $s \in \mathscr{S}$, $s \rhd^+ A$ if $s \rhd^+ B$ for all $B \in \Gamma$.

(LPa) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ and for all atomically complete states $s \in \mathscr{S}$, $s \rhd^+ A$ if $s \rhd^+ B$ for all $B \in \Gamma$.

(LPb) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ and for all atomically sound states $s \in \mathscr{S}$, $s \not\rhd^- A$ if $s \not\rhd^- B$ for all $B \in \Gamma$.

[4]The present analysis is similar to the one in Angelberger, Faroldi and Korbmacher (2016). (LPb) and (TCb) below are due to the present author.

It can easily be checked that each condition defines the consequence of the corresponding logic. To briefly indicate how, let us consider (LPa) and (LPb) by way of example. There are two standard ways of defining LP-consequence in terms of Belnapian valuations. It may be defined as the preservation of the designated values $\{T\}$ and $\{T,F\}$ over all Belnapian valuations assigning only $\{T\}$, $\{F\}$, or $\{T,F\}$ to formulas. Or, equivalently, it may be also be defined as the preservation of the designated values $\{T\}$ and $\emptyset$ over all Belnapian valuations assigning only $\{T\}$, $\{F\}$, or $\emptyset$. Now, atomically complete states in a model correspond to Belnapian valuations which assign only $\{T\}$, $\{F\}$, or $\{T,F\}$ to formulas. So, for each such state s in a model, the following equivalence holds for all formulas A:

$$s \rhd^{+} A \quad \Leftrightarrow \quad \overline{\varphi_s}(A) = \{T\} \text{ or } \overline{\varphi_s}(A) = \{T,F\}.$$

Hence (LPa) is equivalent to the first definition of LP-consequence as the preservation of $\{T\}$ and $\{T,F\}$ over all Belnapian valuations that assign only $\{T\}$, $\{F\}$ and $\{T,F\}$ to formulas. Atomically sound states, on the other hand, correspond to Belnapian valuations which assign only $\{T\}$, $\{F\}$, or $\emptyset$. In this case, therefore, we have:

$$s \rhd^{-} A \quad \Leftrightarrow \quad \overline{\varphi_s}(A) = \{F\}$$

(LPb) is thus equivalent to the second definition of LP-consequence as the preservation of $\{T\}$ and $\emptyset$ over all Belnapian valuations that assign only $\{T\}$, $\{F\}$ and $\emptyset$ to formulas. Like considerations apply to (FDE) and (K3).

What about tautological consequence? Given that Boolean valuations correspond to atomically sound and complete states, the following definition would seem to be natural:

(TCa) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ for $\Gamma;A$ and for all atomically sound and complete states $s \in \mathscr{S}$, $s \rhd^{+} A$ if $s \rhd^{+} B$ for all $B \in \Gamma$.

That is, tautological consequence is the preservation of $\{T\}$ under all Belnapian valuations assigning only $\{T\}$ or $\{F\}$ to formulas, i.e., under all Boolean valuations.

Though this definition gives a straightforward truthmaker semantical account of tautological consequence, it is not entirely satisfactory for the purpose of this paper. For we will ultimately

attempt to exactify the Kripke semantics with appeal, not to possible worlds, but only to their parts (more on this in section 6 below). The trouble here is that parts of worlds, though obviously atomically sound, may not be atomically complete. We thus face the problem of giving a truthmaker semantical account of tautological consequence without appeal to atomic completeness.

A solution to this problem can be obtained by further generalizing the schematic definition of consequence to allow two sets of designated values—one for formulas in Γ and the other for A—as follows: A is a *consequence* of Γ if and only if A is assigned a value in $\mathscr{D}_1$ by every n-valued valuation that assigns a value in $\mathscr{D}_2$ to every formula in Γ.[5] Then tautological consequence can be defined by restricting the relevant valuations to the Belnapian valuations assigning only $\{T\}$, $\{F\}$, or $\emptyset$ and by taking $\mathscr{D}_1 = \{\{T\}, \emptyset\}$ and $\mathscr{D}_2 = \{\{T\}\}$, which matches up with the following truthmaker semantical condition:

(TCb) For all models $\mathfrak{A} = \langle \mathscr{S}, \sqsubseteq, v \rangle$ for $\Gamma; A$ and for all atomically sound states $s \in \mathscr{S}$, $s \not\triangleright^{-} A$ if $s \triangleright^{+} B$ for all $B \in \Gamma$.

This is easily shown to be equivalent to (TCa). Assume that (TCb) holds. Let s be an atomically sound and complete state such that $s \triangleright^{+} B$ for all $B \in \Gamma$. Then it is immediate from the assumption that $s \not\triangleright^{-} A$. Since s is atomically complete, on the other hand, either $s \triangleright^{+} A$ or $s \triangleright^{-} A$. Therefore, $s \triangleright^{+} A$. The other direction is trivial, so we omit it. On the basis of this analysis of tautological consequence, we can easily establish the soundness and completeness results of classical propositional logic with respect to the proposed truthmaker semantics.[6]

This way, the proposed truthmaker semantics for classical propositional logic provides an exact truthmaker semantical analysis of the central notions of the Boolean semantics, such as truth under a Boolean valuation, tautological consequence and tautology. In this sense, it can be said to offer an exactification of the Boolean semantics.[7]

[5]This "mixed" scheme has received much attention in the recent literature on the sorites and liar paradoxes. It should also be noted that the idea behind (TCb) below originates from the so-called strict-tolerant logic. See Cobreros *et al*. (2012) and Fitting (2021) for the basic ideas behind the mixed scheme and for the ways it relates to the strict-tolerant and other many-valued logics.

[6]To my knowledge, the proof is not available in print. This is probably because it is obvious to anyone who is familiar enough the basic formal machinery.

[7]It is worth noting that there is another project in the vicinity of exactification. Various notions of the standard semantics are usually understood in terms of truth and falsity. The project of exactification is conceived as that of giving an account of those standard notions in terms of verification and falsification. But we may also attempt to

5. Informal analysis of the modal operators

Let us now consider how to extend the truthmaker semantics to modalized propositions. First off, what are the exact verifiers and falsifiers for a necessitated proposition $\Box P$? It might be in better accordance with our abstract approach to the nature of states to ask what the logical function of an exact verifier for $\Box P$ should be. The current informal analysis is based on the following principle:

(KP) $\Box P$ *is true only if the apparent possibilities of P being false are not real.*

We call it *Kripke's Principle* as it is implicit in his criticism of mind-body identity in his Kripke (1980) as a general principle about the necessary aposteriori. Very roughly, the argument is that the identity of a brain state (e.g., c-fiber stimulation) and a mental state (e.g., pain) is not necessary because the apparent possibility of having a brain state without the corresponding mental state cannot be explained away. Clearly, (KP) is implicit in the argument.[8]

The principle can be generalized to any necessary propositions, either apriori or aposteriori. For in classical setting, we have the duality:

$$\Box P \equiv \neg\Diamond\neg P.$$

Establishing the necessity of P logically amounts to establishing the impossibility of P's being false, and this is regardless of whether P is apriori or aposteriori.[9]

Let us consider how to capture the intuitive content of (KP) in terms of states. It seems plausible that there are states with some modal implications. For an intuitive example, let s be the identity of water to H_2O and t be the state that water has the chemical composition of XYZ. Putting some philosophical complications aside, t is not a real metaphysical possibility because of s; so, we may

---

define various notions directly in terms of exact and inexact verifications and see how they relate to one another and to the standard one. See, for example, van Fraassen (1969, p.485) and Fine (2014, pp.556-557; 2017a, p.669) for various such notions of consequence.

[8] See 3.7 for a further discussion of (KP).

[9] Some might think that (KP) is redundant in case of the necessary apriori; for, if P is a necessary apriori proposition, then it would not even appear to be possible that P is false. For a simple counterexample to this, consider an elementary arithmetical proposition, say, that $537 \times 771 = 414027$. It certainly appears possible that this proposition is false. For otherwise we should be able to see its truth without calculation. In general, that a proposition a priori does not imply that it is immediately evident, i.e., does not even appear to be possibly false.

plausibly say that s renders t metaphysically impossible. For another example, let u be the moral law that one ought not kill an innocent person, w be Jack's killing an innocent person. Then it is plausible to think that w is not a morally permissible state because of u. So we may say that u makes w morally impermissible.

To capture this idea within the formal framework, we designate a subset $M \subseteq \mathscr{S}$ as the set of *modal states* and assign to each $s \in M$ a set $\beta(s)$ of states. When $t \in \beta(s)$, we say that s *bans*, or *prevents*, t; intuitively, $t \in \beta(s)$ means that s renders t impossible. We shall require the banning relation to be exact in the sense that when $t \in \beta(s)$, s must be entirely relevant to the impossibility of t.[10] The examples given above can all be considered as instances of exact banning. With the notion of exact banning in place, we can define the inexact banning relation in the usual way: a state s *inexactly bans* t (in symbol, $t \in \overline{\beta}(s)$) just in case s extends a modal state that exactly bans t, i.e.,

$$\text{there is a modal state } s' \text{ such that } s' \sqsubseteq s \text{ and } t \in \beta(s).$$

Note here that s need not itself be a modal state.

The notion of banning enables us to formulate a truthmaker semantical version of (KP):

(sKP) *$\Box P$ is true if and only if every exact falsifier for P is exactly banned by some modal state.*

Note that this is a strengthening of Kripke's Principle because it also says that banning all the exact falsifiers for P is *sufficient* to establish the necessity of P, which is natural in the light of the duality between necessity and possibility.

(sKP) is not "exact" yet, however, because it does not provide an exact verifier for $\Box P$. To formulate an exact version of (sKP), notice first that P may have more than one exact falsifier in the most general case. At least offhand, moreover, there does not have to be a single modal state

[10] It might be thought that the notion of exactness here deviates from the standard one. For exactness is standardly understood as relevance of a state in its entirety to the truth or falsity of a proposition; in the case of banning, on the other hand, exactness is explained as relevance of a state in its entirety to the impossibility of another state. Recall, however, that states are supposed to serve as exact verifiers or falsifiers for propositions. So, when t is exactly banned by s, s is entirely relevant to the impossibility of a proposition P for which t is an exact verifier, i.e., to the falsity of the modalized proposition that P is possible. This way, exactness of the banning relation can also be understood in the standard way. In this connection, it is worth noting that Fine (2017a, pp.634-635) also considers an exact notion of exclusion to provide an account of negation from the unilateral truthmaker principle. For any two states s and t, roughly, a state s *excludes* t just in case $s \sqcup t$ is impossible (according to some prior notion of impossibility). Exclusion is analogous to banning in that it is subject to conditions that we would naturally want for banning, e.g., the upward closure condition to be defined below. But it should be obvious that their intended interpretations are different.

that exactly bans all of them; it seems to be conceivable that different modal states ban different exact falsifiers for P. To accommodate these cases, let's define a *ban on the exact falsifiers for P* to be a function f from the set of exact falsifiers for P into M such that for each exact falsifier t for P,

$$t \in \beta(f(t)).$$

So f is a function that maps each exact falsifier for P to a modal state that exactly bans it. Then we may take an exact verifier for $\Box P$ to be the least upper bound of the range of f, in symbol, $\bigsqcup ran(f)$. So we propose the following:

$$s \Vdash^{+} \Box P \quad \Leftrightarrow \quad s = \bigsqcup ran(f), \text{where } f \text{ is a ban on the exact falsifiers for } P.$$

Intuitively, s is a least state containing, for each exact falsifier for P, a modal state exactly bans it.

Let's turn to $\Diamond P$. One might think that $\Diamond P$ should be true just in case not all exact verifiers for P are banned. The idea may seem reasonable in the light of the duality and (sKP). But there are two problems. First, it leaves unclear what an exact verifier for $\Diamond P$ is. The other is that just because a certain state s does not ban another state t, it does not necessarily follow that s thinks t possible; s is perhaps irrelevant to the modal status of t and hence has nothing to do with whether t is possible or not.

Instead of giving a merely negative account of $\Diamond P$, therefore, we shall associate each modal state $s \in M$ with a set $\alpha(s)$ of states. Informally, we interpret $t \in \alpha(s)$ to mean that "*s allows t*" or "*s* countenances *t*." Again, we will require the allowing relation to be exact. For some intuitive examples of exact allowing, we may think that Jack's having a certain genetic make-up exactly allows him to grow taller than 6 feet; and one might be allowed to cause a certain harm because it is a necessary means to bringing about a good result. The inexact allowing relation can be defined in the usual way. That is, t is *inexactly allowed* by s (in symbol, $t \in \overline{\alpha}(s)$) just in case s extends a modal state that exactly allows t, i.e.,

$$\text{there is a modal state } s' \text{ such that } s' \sqsubseteq s \text{ and } t \in \alpha(s').$$

With the notion of allowing at hand, we can give a positive account of the possibility operator:

$$s \Vdash^+ \Diamond P \quad \Leftrightarrow \quad \text{for some exact verifier } t \text{ for } P,\ t \in \alpha(s).$$

So, an exact verifier for $\Diamond P$ is a state that exactly allows an exact verifier for P.

From the definitions given so far and the duality between necessity and possibility, we have:

$$\begin{aligned} s \Vdash^- \Box P \quad &\Leftrightarrow \quad s \Vdash^- \neg\Diamond\neg P; \\ &\Leftrightarrow \quad s \Vdash^+ \Diamond\neg P; \\ &\Leftrightarrow \quad \text{for some exact falsifier } t \text{ for } P,\ t \in \alpha(s); \end{aligned}$$

and

$$\begin{aligned} s \Vdash^- \Diamond P \quad &\Leftrightarrow \quad s \Vdash^- \neg\Box\neg P; \\ &\Leftrightarrow \quad s \Vdash^+ \Box\neg P; \\ &\Leftrightarrow \quad s = \bigsqcup ran(f), \text{where } f \text{ is a ban on the exact verifiers for } P. \end{aligned}$$

To make all this rigorous, we define a *modalized state space* (simply, *m-space*) to be an ordered triple $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$, where $\langle \mathscr{S}, \sqsubseteq \rangle$ is a state space and μ is a function with domain $M \subseteq \mathscr{S}$ such that μ assigns to each $s \in M$ a pair $\langle \alpha(s), \beta(s) \rangle$ of subsets of $\mathscr{S}$. Then we can define a *modal model* (simply, *m-model*) $\mathscr{M}$ of Γ to be an ordered quadruple $\langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$, where $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is an m-space and v is a valuation as defined in section 2. Then we extend the notions of exact verification and falsification by adding the following clauses for necessitated formulas:

$$\begin{aligned} s \Vdash^+ \Box A \quad &\Leftrightarrow \quad s = \bigsqcup ran(f), \text{where } f \text{ is a ban of the exact falsifiers on } A; \\ s \Vdash^- \Box A \quad &\Leftrightarrow \quad \text{for some exact falsifier } t \text{ for } A,\ t \in \alpha(s). \end{aligned}$$

For any set U of states, we let

$$\mathscr{A}(U) = \{x \in \mathscr{S} : (\exists u \in U)(u \in \alpha(x))\};$$

$$\mathscr{B}(U) = \{x \in \mathscr{S} : (\exists f)(f \text{ is a ban on } U \text{ and } x = ran(f)\}.$$

Then the above clauses can be rewritten thus:

$$s \Vdash^+ \Box A \quad \Leftrightarrow \quad s \in \mathscr{B}(|A|^-);$$

$$s \Vdash^- \Box A \quad \Leftrightarrow \quad s \in \mathscr{A}(|A|^-).$$

6. Exactification of the Kripke semantics

How then does the proposed semantics for modalized propositions exactify the Kripke semantics? A possible world in the Kripke semantics can be regarded as a totality of states that obtain in it. So conceived, each possible world itself can play the role of a verifier and falsifier. Here, again, the relevant notions of verification and falsification are inexact. So, according to exactification, each possible world should be representable as a state having as its part an exact verifier for each formula that is true—and an exact falsifier for each formula that is false—in the possible world.

The Kripke semantics for modalized propositions is given in terms of the accessibility relation between possible worlds: $\Box P$ is true at a possible world w if and only if P is true at all possible worlds accessible from w, and $\Diamond P$ is true at w if and only if P is true at some possible world accessible from w. The current proposal attempts to exactify these clauses by giving an account of the accessibility relation between possible worlds in terms of the allowing and banning relations between the constituent states.

To this end, let us first consider some of the conditions on the allowing and banning relations that possible worlds and their constituent states should satisfy. For each state s (either modal or non-modal), define the *modal profile* of s as the pair $\langle \overline{\alpha}(s), \overline{\beta}(s) \rangle$ of sets of states. Intuitively, $\overline{\alpha}(s)$ is the set of states that s thinks possible, and $\overline{\beta}(s)$ is the set of states that s thinks impossible. There are a couple of conditions that are highly plausible concerning the modal profile $\langle \overline{\alpha}(s), \overline{\beta}(s) \rangle$ of a

state s. First, $\overline{\alpha}(s)$ should be *downward closed*, meaning that

$$(t \in \overline{\alpha}(s) \text{ and } t' \sqsubseteq t) \quad \Rightarrow \quad t' \in \overline{\alpha}(s).$$

Second, $\overline{\beta}(s)$ should be *upward closed*:

$$(t \in \overline{\beta}(s) \text{ and } t \sqsubseteq t') \quad \Rightarrow \quad t' \in \overline{\beta}(s).$$

Here one might argue for a stronger requirement that the exact allowing relation itself be downward closed. This requirement is objectionable, however. For, when $t \in \alpha(s)$ and $t' \sqsubseteq t$, it is conceivable that only a proper part of s is relevant to the possibility of t'; in such a case, we would have to say that t' is inexactly—but not exactly—allowed by s. So, the requirement is not acceptable as long as we want to maintain the distinction between the exact and inexact allowing relations. We address this worry by requiring the downward closure condition on $\overline{\alpha}(s)$ only. A similar worry might arise if we require the strong upward closure condition on the exact banning relation, although it seems to have much less force.[11]

There is another set of conditions concerning the modal behavior of possible worlds. Let's say that a state s is *modally sound* if and only if there is no state t such that

$$t \in \overline{\alpha}(s) \cap \overline{\beta}(s);$$

otherwise, s is *modally unsound*. Also, s is said to be *modally complete* just in case for all states t,

$$t \in \overline{\alpha}(s) \cup \overline{\beta}(s),$$

[11]It might also be objected that even the weaker requirements are not so obvious in deontic cases. For example, suppose that you are in a state s where there are two buttons, say A and B, such that you administer an electric shock to a pupil if you push either one of them, but nothing happens if you push them both. In the state s, we may reasonably think, you are allowed to push both buttons, but not one; or, you are banned from pushing one button but not from pushing both. Since pushing one button is part of pushing two, the objection goes, even the inexact allowing relation is not downward closed; nor is the inexact banning relation upward closed. The objection is mistaken, as one can easily see, because what is not allowed is to push *exactly* one button, i.e., to push A and not B or to push B and not A. And these are not part of pushing both A and B. But if you are allowed to push both A and B, however, you are of course allowed to push A and to push B.

and to be *modally incomplete* otherwise. It is reasonable to think that the possible worlds in the Kripke semantics, understood as the totalities of states, should be modally sound and complete.

In the standard Kripke semantics, the accessibility relation is intended to capture the relative possibility; so, when w' is accessible from w, it means that w' is possible relative to w. Since we intend to capture the relative possibility in terms of the allowing relation, it is natural that for any worlds w and w',

$$w' \text{ is accessible from } w \quad \Leftrightarrow \quad w' \in \overline{\alpha}(w).$$

Hence we have:

$$\begin{aligned} w' \text{ is not accessible from } w \quad &\Leftrightarrow \quad w' \notin \overline{\alpha}(w); \qquad \text{then, by the modal completeness of } w, \\ &\Leftrightarrow \quad w' \in \overline{\beta}(w), \end{aligned}$$

as one would desire.[12] Note also that the closure conditions on the inexact allowing and banning relations imply that for any states s and t,

$$\begin{aligned} t \in \overline{\alpha}(s) \quad &\Leftrightarrow \quad \text{every part of } t \text{ is in } \overline{\alpha}(s); \\ t \in \overline{\beta}(s) \quad &\Leftrightarrow \quad \text{some part of } t \text{ is in } \overline{\beta}(s). \end{aligned}$$

It thus follows: for any worlds w and w',

$$\begin{aligned} w' \text{ is accessible from } w \quad &\Leftrightarrow \quad \text{every part of } w' \text{ is in } \overline{\alpha}(w); \\ w' \text{ is not accessible from } w \quad &\Leftrightarrow \quad \text{some part of } w' \text{ is in } \overline{\beta}(w). \end{aligned}$$

Let us consider how we may derive the Kripke semantical clauses for modalized propositions from the exact counterparts. Suppose that $\Box P$ is true at a possible world w. According to the current proposal, this means that there is a state $s \sqsubseteq w$ such that $s = \bigsqcup ran(f)$ for some ban f on the exact falsifiers for P. Now, if P is false at a possible world w', then w' must extend an exact

[12] It might be thought that w' is accessible from w if and only if $w' \in \alpha(w)$, and that w' is inaccessible from w if and only if $w' \in \beta(w)$. This analysis is objectionable because it presupposes that possible worlds are themselves modal states. To get around this problem, we use the notions of inexact allowing and banning.

falsifier t' for P. Since $t' \in \beta(f(t'))$ and $f(t') \sqsubseteq s \sqsubseteq w$, $t' \in \overline{\beta}(w)$. So, w' is not accessible from w according to the current analysis of the accessibility relation. Since w' is arbitrary, it follows that every world w' at which P is false is not accessible from w. And this is exactly the condition under which $\Box P$ is true at w according to the Kripke semantics.

Suppose, on the other hand, that $\Diamond P$ is true at w. Then an exact verifier t of P must be in $\overline{\alpha}(w)$. From this, however, it does not follow that there is an accessible world w' at which P is true. For every world extending t might have as its part a state that is banned by some part of w; in such cases, there would be no world accessible from w that has an exact verifier for P as its part. To exclude such cases, it should be required that every possible world is *robust about the possibilities*: for any world w and any state t

(R) whenever $t \in \overline{\alpha}(w)$, there is a possible world w' such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$.

Intuitively, this condition says that every possibility of w should be realized in a world w' accessible from w. It thus guarantees that whenever w allows an exact verifier t of P, there must be a world w' extending t such that $w' \in \overline{\alpha}(w)$—that is, w must have an accessible world w' at which P is true.

One may object that the robustness condition is ad hoc, lacking intuitive or theoretical grounds. To first see intuitive appeal of the robustness condition. Suppose that w is not robust about the possibilities. This means that w sees a state t as a possibility but every world w' extending t is impossible relative to w. However, it is not unreasonable to think if t cannot be realized in a world possible relative to w then t should not be considered as a possibility relative to w; that is, every possibility of w should be realized in a world possible relative to w. But this is tantamount to requiring the robustness condition on w.

In addition to this intuitive ground, there is also a compelling reason to think that the robustness condition is indispensable for the purpose of exactification. Since we think of possible worlds themselves as verifiers, according to exactification, the truth of $\Diamond P$ at w requires that there is an exact verifier for $\Diamond P$ *within* w. In the Kripke semantics, on the other hand, the truth of $\Diamond P$ at w requires the existence of a potentially different possible world at which P is true. Hence any attempt to exactify the Kripke semantical clause for $\Diamond P$ faces the problem of finding a state within w that witnesses the existence of a potentially different possible world that has an exact verifier for

P. And of course the solution in its most general and abstract form consists in positing states that do the job. This is exactly what we do in postulating modal states and in imposing the robustness condition: whenever a state t is possible relative to a world w—that is, $t \in \overline{\alpha}(w)$—there is a modal state s within w such that s countenances t as part of some world w' which is itself possible relative to w. This solution, to be sure, leaves the substantive question of what modal states are. But no particular conception of modal states is required for the purpose of exactification, except that they should be subject to the robustness condition or to another condition to the same effect. In this regard, the robustness condition seems to be indispensable for the exactification of the Kripke semantics.

We can also derive the exact truthmaker semantical clauses for modalized propositions from the corresponding clauses of the Kripke semantics. Suppose that $\Box P$ is true at a world w. That is, P is true at all worlds w' accessible from w. Assume for *reductio* that no $s \sqsubseteq w$ is such that $s = \bigsqcup ran(f)$ for some ban f of the exact falsifiers for P. Since w is modally complete, some exact falsifier t of P must be in $\overline{\alpha}(w)$. By (R), then, there would be a world w' such that $t \sqsubseteq w'$ and w' is accessible from w, which implies that $\Diamond\neg P$ would be true at w; contradiction. So it follows that w contains a ban of the exact falsifiers for P.

Suppose now that $\Diamond P$ is true at a world w; that is, P is true at a world w' accessible from w. Since P is true at w', on the one hand, w' contains an exact verifier t of P. Since w' is accessible from w, on the other hand, every part of w' is in $\overline{\alpha}(w)$. It thus follows that t is in $\overline{\alpha}(w)$.

So, the proposed exact truthmaker semantics for modalized propositions are equivalent to the Kripke semantics, given some natural assumptions about the modal behavior of possible worlds and their constituent states. We shall later show that modally sound and complete states satisfying the robustness condition behave just like possible worlds (Theorem 8), and also that each Kripke model can be translated to an m-model in such a way that each world in the Kripke model is represented as an equivalent modally sound and complete state satisfying the robustness condition (Theorems 15). In the light of these results, the present analysis can be considered an exactification the basic notion of the Kripke semantics, namely truth at a possible world.

The emerging picture of the relation between the two semantics is quite straightforward. In the Kripke semantics, the possible worlds are conceived merely as indices that bear an accessibility relation to each another. Under the current proposal, the possible worlds are given internal structures of states, which bear the allowing and banning relations to each other. It thus gives an account of the possible worlds and the accessibility relation in terms of their internal structures and the way the constituent states are related. This seems to be well-aligned with the way we think about possible worlds in many applications. Imagine, for example, that we are tossing two coins, say A and B, simultaneously. Each coin will land either on its head or tail. So there are four 'possible worlds' in total. These possible worlds are constituted by the states of each coin. For example, the possible world $(A : \mathit{head}, B : \mathit{tail})$ would most naturally be considered as consisting of the states $(A : \mathit{head})$ and $(B : \mathit{tail})$. If we work under the notion of possibility such that it is impossible for a coin to show up both faces at the same time and such that each coin can fall on either side regardless of which side the other coin falls on, then we may think of the state $(A : \mathit{head})$ as a modal state that bans the state $(A : \mathit{tail})$ and vice versa, while it allows both $(B : \mathit{head})$ and $(B : \mathit{tail})$. In this way, the present truthmaker semantics captures the intuitive way that we think about possible worlds.

Finally, let us consider how we may give an exact truthmaker semantical account of consequence in modal logic. In the Kripke semantics, the notion of consequence is defined, for each modal logic L, as the preservation of truth in all possible worlds over all Kripke models for L. That is, a formula A is said to be a consequence of a set Γ of formulas in modal logic L if and only if every Kripke model for L is such that A is true in all worlds in which every formula in Γ is true. Now, according to the current analysis, possible worlds are conceived as modally sound and complete states that are robust about the possibilities; so, let us call such states *worldly*. For the purpose of developing a formal semantics for modal logic, we may simply assume that there are worldly states in m-spaces. Then we can give the following straightforward truthmaker semantical account of consequence in modal logic:

(*MCa*) In all m-models for $\Gamma;A$ of modal logic L and for all worldly states w, $w \rhd^{+} A$ if $w \rhd^{+} B$ for all $B \in \Gamma$.

(Note that this is not intended to be technically completely precise; see the formal definitions of relevant notions in chapter 2.)

One problem with this account is that it is at odds with the main philosophical motivation behind the truthmaker semantics, as is forcefully expressed by Fine (Fine, 2017a, p.645):

> One remarkable aspect of the present theory of truthmaker content is that possible worlds completely drop out of the picture. ... One might jokingly remark that the possible worlds approach is fine but for two features: the first is that possible worlds are worlds, i.e., complete rather than partial; and the second is that they are possible. Drop both requirements, impose a mereological structure on the resulting states, and we obtain a framework that is of much more help in developing an adequate theory of content.

I myself do not think that it is so important to get rid of possible worlds. More important, I think, is to give an account of possible worlds—and that is what I have done so far under the rubric of exactification. But I do agree with Fine that it is desirable to have a semantics for modal logic that does not necessarily assume the existence of possible worlds. So, we face the problem of finding a truthmaker semantical account of consequence in modal logic with appeal, not to possible worlds, but only to their parts.

This problem is obviously analogous to the one we faced earlier in exactifying the Boolean semantics; and it can be resolved in essentially the same way. To see how, let us first introduce a couple of auxiliary notions. Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ of Γ be an m-model. We say that a state s is a *modal boundary* if and only if s itself is modally sound and every proper extension of s is modally unsound. Then we define the set $\mathscr{S}^{\Diamond}$ of *absolute possibilities* thus:

$$\mathscr{S}^{\Diamond} = \{t \in \mathscr{S} : (\exists s \in \mathscr{S})(s \text{ is a modal boundary and } t \sqsubseteq s)\}.^{13}$$

[13]The definition is inspired by Fine's notion of a modalized state space in Fine (2017a,b). A *modalized state space* is $\langle \mathscr{S}, \sqsubseteq, S \rangle$ where $\langle \mathscr{S}, \sqsubseteq \rangle$ is a state space and S is a subset of $\mathscr{S}$ that is closed under parthood relation: for all state s and t, if $s \in S$ and $t \sqsubseteq s$ then $t \in S$. Intuitively, a modalized state space is a state space with a designated set S of states that are considered possible in the "absolute" sense. The above definition of absolute possibility essentially provides an analysis of Fine's notion of modalized state space in terms of the allowing and banning relations.

Intuitively, the modal boundaries are the largest parts of possible worlds that $\mathscr{M}$ can see; but they may not be possible worlds. And a state s is considered an *absolute possibility* if it is part of some modal boundary. Using these notions, we can define the notion of consequence in modal logic thus:

(*MCb*) In all m-models for $\Gamma;A$ of modal logic L and for all $s \in \mathscr{S}^{\Diamond}$, $s \not\triangleright^{-} A$ if $w \triangleright^{+} B$ for all $B \in \Gamma$.

(Again, see Definition 4 in section 8 for the precise definition.) The similarity of (*MCb*) to ($\mathsf{TC}b$) should be obvious: in both cases, we drop the completeness requirement and give a mixed account for consequence in terms both of verification and falsification. We shall later show that this account of consequence can be used to establish the soundness and completeness results for various systems of modal logic (Theorem 16 and Corollary 24).

7. Incompatibility conditions

Dropping possible worlds out of the picture brings some further complications in the modal behavior of states. Let us consider what, according to the present analysis of the modal operators, it means for a state s to verify axiom K, namely $\Box(A \supset B) \supset (\Box A \supset \Box B)$, where A and B are any formulas. Note first that $s \triangleright^{+} \Box(A \supset B)$ if and only if for any exact verifier t for A and for any exact falsifier u for B, $t \sqcup u \in \overline{\beta}(s)$. Suppose, on the other hand, that $s \triangleright^{+} \Box A$. This means that, for every falsifier t for A, $t \in \overline{\beta}(s)$. By the upward closure condition, then, it follows that for any exact falsifier t for A and for any exact verifier u for B, $t \sqcup u \in \overline{\beta}(s)$. So, if both $s \triangleright^{+} \Box(A \supset B)$ and $s \triangleright^{+} \Box A$, it means, informally, that s thinks that no exact falsifier for B can be put together either with an exact verifier, or with an exact falsifier, for A. In such a case—and this is what it means for s to verify $\Box(A \supset B) \supset (\Box A \supset \Box B)$—every exact falsifiers for B should also be in $\overline{\beta}(s)$.

Given this reading of axiom K, we can easily see that it is verified by every possible world—i.e., every modally sound and complete state that is robust about the possibilities. Let w be a possible world, and suppose that $w \triangleright^{+} \Box(A \supset B)$ and $w \triangleright^{+} \Box A$. So, w thinks that no exact falsifier for B can be put together either with an exact verifier, or with an exact falsifier, for A. Assume for *reductio* that $w \triangleright^{-} \Box B$, i.e., that there is an exact falsifier t for B such that $t \in \overline{\alpha}(w)$. By the robustness

condition, then, there should be another possible world w' such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. Since possible worlds determine the truth-value of every formula, either $w' \triangleright^+ A$ or $w' \triangleright^- A$, which means that w' should extend either an exact verifier, or an exact falsifier, u for A. Then, since both $t \sqsubseteq w'$ and $u \sqsubseteq w'$, it should follow that $t \sqcup u \sqsubseteq w'$. Now, recall that $w' \in \overline{\alpha}(w)$. By the downward closure condition of $\overline{\alpha}$, then, $t \sqcup u \in \overline{\alpha}(w)$. In other words, w should think that some exact falsifier for B, namely t, can be put together with an exact verifier, or an exact falsifier, u for A. But this contradicts the initial assumption. It thus follows by *reductio* that no exact falsifier for B is in $\overline{\alpha}(w)$. Since w is modally complete, we may conclude that every exact falsifier for B is in $\overline{\beta}(w)$, i.e., that $w \triangleright^+ \Box B$.

The very last step of this argument makes a crucial use of the fact that possible worlds are modally complete. So, the argument does not go through for states that are not possible worlds. For they may not be modally complete. This gives rise to a problem for the present approach because it aims to provide an account of consequence and hence of validity, in modal logic without appeal to modal completeness.

To address this problem, some further constraints need to be imposed on the modal behavior of parts of possible worlds. Let s be any state, and T and U be sets of states. We say that T and U are *s-incompatible* if and only if $T \sqcup U \subseteq \overline{\beta}(s)$. Informally, T and U are s-incompatible when s thinks that no state in T can be put together with any state in U. Now, suppose that we have a modal boundary s and a state t such that, for some propositional variable P, $\{t\}$ and $|P|^+ \cup |P|^-$ are s-incompatible; that is,

$$\{t\} \sqcup (|P|^+ \cup |P|^-) \subseteq \overline{\beta}(s).$$

Intuitively, s thinks that t can be combined neither with any exact verifiers, nor with any exact falsifiers, for P. Under the natural assumption that every possible world must contain either an exact verifier or an exact falsifier for every proposition, the incompatibility implies that s thinks that t is part of no world accessible from it. Since possible worlds are robust about the possibilities, s already has sufficient reason to think that t is not a real possibility. We may thus conclude that $t \in \overline{\beta}(s)$. More generally, then, let's say that t is *atomically s-incompatible* if and only if there is a propositional variable P such that $\{t\}$ and $|P|^+ \cup |P|^+$ are s-incompatible. Then we may

plausibly require that for any modal boundary s and any state t, if t is atomically s-incompatible, then $t \in \overline{\beta}(s)$.

Similarly, let's say that t is *modally s-incompatible* just in case there is a set U of states such that

$$\{t\} \sqcup (\mathscr{A}(U) \cup \mathscr{B}(U)) \subseteq \overline{\beta}(s).$$

Suppose that t is modally s-incompatible. Since every possible world w is modally complete, w must think, for any set U, that either something in U is possible or else everything in U is impossible. So, every possible world w must contain as part some $u \in (\mathscr{A}(U) \cup \mathscr{B}(U))$. Since s thinks that t cannot be combined any such u, s should also think that t does not belong to any possible world accessible from it. So, we may require that for any modal boundary s and any state t, if t is modally s-incompatible then $t \in \overline{\beta}(s)$.

We thus have the requirement that for any modal boundary s and a state t, $t \in \overline{\beta}(s)$ if t is either atomically or modally s-incompatible. Notice that the requirement concerns the modal behavior of a state conceived as part of a possible world. It should be noted, however, that the requirement does not by itself imply the existence of possible worlds. In the next chapter, we shall see that the requirement can be satisfied by an m-space that does not have any possible worlds in it (see Examples 3 and 4 below).

8. Concluding remarks

Truthmaker semantics offers a natural conceptual framework to address the question of what makes a necessary truth necessarily true. For, in asking this question, we are seeking what it is that in virtue of which P is necessary, i.e., what an exact verifier for $\Box P$ is. In this chapter, I proposed a formal, logical answer to the question. The intuitive basis of the proposal was Kripke's Principle: $\Box P$ is true only if the apparent possibilities of P being false are not real. On the basis of this principle, I proposed that what makes $\Box P$ true should have the logical function of excluding all the ways in which P might be falsified; more formally, an exact verifier for $\Box P$ is a smallest state that bans every exact falsifier for P.

According to the standard Kripke semantics, the truth condition for $\Box P$ is given in terms of the accessibility relation between possible worlds. However, it does not provide an account of what makes a possible world accessible or inaccessible from another world. The present truthmaker semantical analysis gives an account of the accessibility relation between possible worlds in terms of the banning and allowing relation between their constituent states. The basic idea is that some states have modal implications; we called such states *modal states*. A possible world w is accessible to another possible world w' when every constituent state of w' is allowed by some modal state in w; and w is inaccessible to w' when some constituent state of w' is banned by some modal state in w. On the basis of this account, we have shown that the truthmaker semantical analysis of modal operators is equivalent to that of the standard Kripke semantics. The present analysis thus offers an exactification of the Kripke semantics: it analyzes the accessibility relation between possible worlds in terms of the allowing and banning relations between their constituent states and thereby gives an account of "truth at a world" in terms of exact truthmaking.

It should be noted that the discussion so far offers only a formal, logical answer to the problem about necessity. What makes a proposition necessary should have the logical function of excluding all the ways in which the proposition might be false from the realm of real possibilities. Hence we recognize the necessity of a proposition by explaining away all the ways in which the proposition might be false as unreal possibilities. However, this analysis leaves unanswered the substantive, philosophical question of what exactly it is that excludes the apparent possibilities of a proposition's being false from the realm of real possibilities. This will be the main issue of chapters 3 and 4 below. Before turning to that issue, however, I shall first show that the present analysis can be developed into an adequate formal truthmaker semantics for modal operators.

CHAPTER 2

Truthmaker Semantics for Modal Logic, Part 2

1. Introduction

This chapter gives a formal exposition of the proposed semantics and establishes the soundness and completeness results for a well-known family of systems of normal modal propositional logic, namely K, D, B, 4, K4, S4 and S5. These results provide a formal vindication that the proposed analysis of modal statements is logically adequate. Despite the risk of being repetitive, I shall re-state the key definitions, so that the technical exposition can be read independently of the informal discussions of the previous chapter.

2. Formal exposition

A *state space* is defined to be a complete partial order $\langle \mathscr{S}, \sqsubseteq \rangle$, where $\mathscr{S}$ is any set and $\sqsubseteq$ is a reflexive, transitive and anti-symmetric binary relation on $\mathscr{S}$ such that every $S \subseteq \mathscr{S}$ has a least upper bound (in symbol, $\bigsqcup S$). For any sets S and T of states, we let

$$S \sqcup T = \{s \sqcup t : (\exists s)(\exists t)(s \in S \text{ and } t \in T)\}.$$

Notice that $S \sqcup T = \emptyset$ if and only if either S or T is empty. In case $S = \{s\}$,

$$\{s\} \sqcup T = \{s \sqcup t : t \in T\},$$

where $s \sqcup t = \bigsqcup\{s,t\}$.

An *m-space* is defined to be an ordered triple $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$, where $\langle \mathscr{S}, \sqsubseteq \rangle$ is a state space and μ is a function from $M \subseteq \mathscr{S}$ into $\mathscr{P}(\mathscr{S}) \times \mathscr{P}(\mathscr{S})$. Each $s \in M$ is called a *modal state* and μ assigns to each $s \in M$ a pair $\langle \alpha(s), \beta(s) \rangle$ of subsets of $\mathscr{S}$. When $t \in \alpha(s)$, we say that t is *exactly allowed* by s; and when $t \in \beta(s)$, t is said to be *exactly banned* by s. In writing $\alpha(s)$ and $\beta(s)$, s is

assumed to be a modal state. For each state $s \in \mathscr{S}$, the *modal profile* of s is $\langle \overline{\alpha}(s), \overline{\beta}(s) \rangle$, where

$$\overline{\alpha}(s) = \{t \in \mathscr{S} : (\exists s' \in \mathscr{S})(s' \sqsubseteq s \text{ and } t \in \alpha(s'))\};$$

$$\overline{\beta}(s) = \{t \in \mathscr{S} : (\exists s' \in \mathscr{S})(s' \sqsubseteq s \text{ and } t \in \beta(s'))\}.$$

When $t \in \overline{\alpha}(s)$, t is said to be *inexactly allowed* by s; and when $t \in \overline{\beta}(s)$, we say that t is *inexactly banned* by s. A state s is *modally sound* if and only if $\overline{\alpha}(s) \cap \overline{\beta}(s) = \emptyset$, and *modally complete* if and only if $\overline{\alpha}(s) \cup \overline{\beta}(s) = \mathscr{S}$.

Given an m-space $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$, a state s is a *modal boundary* of Σ if and only if s is modally sound and every proper extension of s is modally unsound. Then we define the set $\mathscr{S}^{\Diamond}$ of *absolute possibilities* of Σ as follows:

$$\mathscr{S}^{\Diamond} = \{t \in \mathscr{S} : (\exists s \in \mathscr{S})(s \text{ is a modal boundary and } t \sqsubseteq s)\}.$$

Given a set $U \subseteq \mathscr{S}$, a *ban* on U is a function f from U into M such that for all $u \in U$, $u \in \beta(f(u))$. We also let

$$\mathscr{A}(U) = \{x \in \mathscr{S} : (\exists u \in U)(u \in \alpha(x))\};$$

$$\mathscr{B}(U) = \{x \in \mathscr{S} : (\exists f)(f \text{ is a ban on } U \text{ and } x = ran(f)\}.$$

Definition 1. An m-space $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is said to be *normal* if and only if it satisfies the following conditions:

(**N1**) $\mathscr{S}^{\Diamond} \neq \emptyset$;

(**N2**) The inexact allowing relation is downward closed: for all states s, t and t',

$$t \in \overline{\alpha}(s) \text{ and } t' \sqsubseteq t \quad \Rightarrow \quad t' \in \overline{\alpha}(s).$$

(**N3**) The inexact banning relation is upward closed: for all states s, t and t',

$$t \in \overline{\beta}(s) \text{ and } t \sqsubseteq t' \quad \Rightarrow \quad t' \in \overline{\beta}(s).$$

(**N4**) For any modal boundary s of Σ and any $t \in \mathscr{S}^{\Diamond}$, if $t \in \overline{\alpha}(s)$, then there is a modal boundary t' such that $t \sqsubseteq t'$ and $t' \in \overline{\alpha}(s')$.

(**N5i**) For all $U \subseteq \mathscr{S}$, $\mathscr{A}(U) \cup \mathscr{B}(U) \neq \emptyset$.

(**N5ii**) For any modal boundary s of Σ and any $t \in \mathscr{S}^{\Diamond}$, if t is modally s-incompatible, i.e., if there is a set U of states such that

$$\{t\} \sqcup (\mathscr{A}(U) \cup \mathscr{B}(U)) \subseteq \overline{\beta}(s),$$

then $t \in \overline{\beta}(s)$.

Let me note a few things about the definition. First, (**N1**) ensures that there are modally sound states in Σ. Second, (**N4**) is a version of robustness condition on the modal boundaries of Σ; but it does not require the modal boundaries to be modally complete. Third, the reason for having (**N5i**) is to technically ensure a certain behavior that we would naturally want from formal models (see Proposition 1 below); it also simplifies (**N5ii**). Finally we can check that (**N5i**) is equivalent to

(**N5i***) For all states s, there exists a state t such that $s \in \alpha(t) \cup \beta(t)$.

That is, every state is either exactly allowed or banned by some states. To see this, first assume that (**N5i**) holds. Let s be any state. By (**N5i**), either $\mathscr{A}(\{s\})$ is nonempty, in which case there is some t such that $s \in \alpha(t)$, or $\mathscr{B}(\{s\})$ is nonempty, in which case there is some t such that $s \in \beta(t)$. Either way, therefore, $s \in \alpha(t) \cup \beta(t)$ for some t. Hence (**N5i***) holds. Conversely, suppose that (**N5i***) holds. Let U be any set of states. Now either each state in U is exactly banned by some state or not. In the former case, $\mathscr{B}(U)$ is not empty. In the latter case, some state in U is not exactly banned by any state; by (**N5i***), then, u is exactly allowed by some state and so $\mathscr{A}(U)$ is not empty. Either way, therefore, $\mathscr{A}(U) \cup \mathscr{B}(U) \neq \emptyset$. So (**N5i**) holds.

Let Γ be a set of formulas. We shall let:

$$At(\Gamma) = \{P : \ P \text{ is a propositional variable occurring in some formula in } \Gamma\};$$

$$Fml(\Gamma) = \{A : \ \text{every atomic subformula } P \text{ of } A \text{ is in } At(\Gamma)\}.$$

Given a set Γ of formulas, we define a *m-model* of Γ to be an ordered quadruple $\langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$, where $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is an m-space and v is a valuation that takes each state $s \in \mathscr{S}$ to a pair $\langle [s]^+, [s]^- \rangle$ of subsets of $At(\Gamma)$. We shall require that for each $P \in At(\Gamma)$, there is some $t \in \mathscr{S}$ such that $P \in [t]^+ \cup [t]^-$.

Definition 2. Given an m-model $\mathscr{M}$ of Γ, the notions of exact verification and falsification (written $\mathscr{M}, s \Vdash^+ A$ and $\mathscr{M}, s \Vdash^- A$, respectively) can be defined recursively as follows:

$$
\begin{array}{rcl}
\mathscr{M}, s \Vdash^+ P & \Leftrightarrow & P \in [s]^+, \text{ for atomic } P; \\
\mathscr{M}, s \Vdash^- P & \Leftrightarrow & P \in [s]^-, \text{ for atomic } P; \\
\mathscr{M}, s \Vdash^+ \neg A & \Leftrightarrow & \mathscr{M}, s \Vdash^- A; \\
\mathscr{M}, s \Vdash^- \neg A & \Leftrightarrow & \mathscr{M}, s \Vdash^+ A; \\
\mathscr{M}, s \Vdash^+ A \wedge B & \Leftrightarrow & s = s_1 \sqcup s_2, \text{ for some } s_1, s_2 \text{ with } \mathscr{M}, s_1 \Vdash^+ A \text{ and } \mathscr{M}, s_2 \Vdash^+ B; \\
\mathscr{M}, s \Vdash^- A \wedge B & \Leftrightarrow & \mathscr{M}, s \Vdash^- A \text{ or } \mathscr{M}, s \Vdash^- B; \\
\mathscr{M}, s \Vdash^+ A \vee B & \Leftrightarrow & \mathscr{M}, s \Vdash^+ A \text{ or } \mathscr{M}, s \Vdash^+ B; \\
\mathscr{M}, s \Vdash^- A \vee B & \Leftrightarrow & s = s_1 \sqcup s_2 \text{ for some } s_1, s_2 \text{ with } \mathscr{M}, s_1 \Vdash^- A \text{ and } \mathscr{M}, s_2 \Vdash^- B. \\
\mathscr{M}, s \Vdash^+ \Box B & \Leftrightarrow & s = \bigsqcup ran(f), \text{ where } f \text{ is a ban on the exact falsifiers for } B; \\
\mathscr{M}, s \Vdash^- \Box B & \Leftrightarrow & t \in \alpha(s) \text{ for some exact falsifier } t \text{ for } B.
\end{array}
$$

We shall often omit the mention of a model $\mathscr{M}$ if it does not sacrifice clarity. Recall, if we let

$$|A|^+ = \{s \in \mathscr{S} : s \Vdash^+ A\}$$

$$|A|^- = \{s \in \mathscr{S} : s \Vdash^- A\},$$

then the clauses for exact verification and falsification can be simplified thus:

$$
\begin{aligned}
s \Vdash^+ P \quad &\Leftrightarrow \quad s \in |P|^+, \text{ for atomic } P;\\
s \Vdash^- P \quad &\Leftrightarrow \quad s \in |P|^-, \text{ for atomic } P;\\
s \Vdash^+ \neg A \quad &\Leftrightarrow \quad s \in |A|^-;\\
s \Vdash^- \neg A \quad &\Leftrightarrow \quad s \in |A|^+;\\
s \Vdash^+ A \wedge B \quad &\Leftrightarrow \quad s \in |A|^+ \sqcup |B|^+\\
s \Vdash^- A \wedge B \quad &\Leftrightarrow \quad s \in |A|^- \text{ or } s \in |B|^-;\\
s \Vdash^+ A \vee B \quad &\Leftrightarrow \quad s \in |A|^+ \text{ or } s \in |B|^+;\\
s \Vdash^- A \vee B \quad &\Leftrightarrow \quad s \in |A|^- \sqcup |B|^-;\\
s \Vdash^+ \Box A \quad &\Leftrightarrow \quad s \in \mathscr{B}(|A|^-);\\
s \Vdash^- \Box A \quad &\Leftrightarrow \quad s \in \mathscr{A}(|A|^-).
\end{aligned}
$$

The notions of inexact verifications and falsifications (in symbol, respectively, $s \triangleright^+ A$ and $s \triangleright^- A$) in an m-model $\mathscr{M}$ are defined as follows:

$$
\begin{aligned}
\mathscr{M}, s \triangleright^+ A \quad &\Leftrightarrow \quad (\exists t)(t \sqsubseteq s \text{ and } \mathscr{M}, t \Vdash^+ A);\\
\mathscr{M}, s \triangleright^- A \quad &\Leftrightarrow \quad (\exists t)(t \sqsubseteq s \text{ and } \mathscr{M}, t \Vdash^- A).
\end{aligned}
$$

In an m-model $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ of Γ, a state $s \in \mathscr{S}$ is said to be *atomically sound* if and only if there is no $P \in At(\Gamma)$ such that $s \triangleright^+ P$ and $s \triangleright^- P$. s is *atomically complete* if and only if for all $P \in At(\Gamma)$, $s \triangleright^+ P$ or $s \triangleright^- P$.

Definition 3. Given a set Γ of formulas, we define a *normal m-model of* Γ to be $\langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$, where $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is a normal m-space and v is a valuation such that:

(**V1**) Every $s \in \mathscr{S}^{\Diamond}$ is atomically sound.

(**V2**) For any modal boundary s of Σ and $t \in \mathscr{S}^{\Diamond}$, if t is atomically s-incompatible, i.e., if there is a propositional variable $P \in At(\Gamma)$ such that

$$\{t\} \sqcup (|P|^+ \cup |P|^-) \subseteq \overline{\beta}(s),$$

then $t \in \overline{\beta}(s)$.

Example 3. Define a state space $\langle \mathscr{S}, \sqsubseteq \rangle$ as follows:

- $\mathscr{S} = \{\bot, s, t, \top\}$;
- $\sqsubseteq = \{\langle \bot, x\rangle : x \in \mathscr{S}\} \cup \{\langle x, \top\rangle : x \in \mathscr{S}\} \cup \{\langle x, x\rangle : x \in \mathscr{S}\}$.

It should be obvious that this defines a state space. Then we define an m-space Σ_1 by letting μ be a function with domain $\mathscr{S}$ such that:

$$\alpha(\bot) = \{\bot\}, \quad \beta(\bot) = \emptyset;$$

$$\alpha(s) = \{s\}, \quad \beta(s) = \emptyset;$$

$$\alpha(t) = \{t\}, \quad \beta(t) = \emptyset;$$

$$\alpha(\top) = \emptyset, \quad \beta(\top) = \mathscr{S}.$$

Σ_1 can be represented by the following diagram:

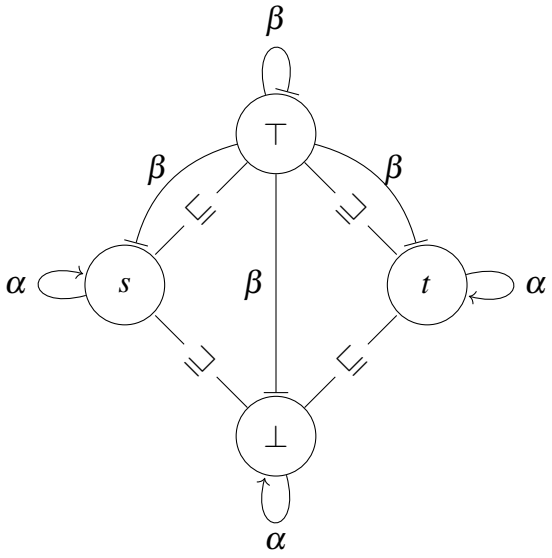

FIGURE 2.1. Diagrammatic representation of Σ_1

Here the parthood relation is represented in the obvious way with the assumptions that each state is part of itself and that it is transitive (e.g., $\bot \sqsubseteq \top$). The exact allowing and banning relations are respectively indicated by normal arrows and by bar-arrows. We check that Σ_1 is a normal m-space. It should be obvious that the modal boundaries of this model are s and t; so, $\mathscr{S}^{\Diamond} = \{\bot, s, t\}$.

So (**N1**) is met. (**N2**) and (**N3**) are clearly met also. To check that (**N4**) is satisfied, it suffices to note that $\overline{\alpha}(s) = \{\bot, s\}$ and $\overline{\alpha}(t) = \{\bot, t\}$. (**N5i**) is satisfied because every $x \in \mathscr{S}$ is hit by some arrow (i.e., either exactly allowed or banned by some state). Finally (**N5ii**) is vacuously satisfied, because $\overline{\beta}(s) = \overline{\beta}(t) = \emptyset$ and because for any $x \in \mathscr{S}^{\Diamond}$ and for any $U \subseteq \mathscr{S}$, $\{x\} \sqcup (\mathscr{A}(U) \cup \mathscr{B}(U))$ is nonempty by (**N5i**).

Now let $\Gamma = \{P \wedge Q\}$. We define a normal m-model $\mathscr{M}_1 = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ for Γ by setting $[s]^+ = \{P\}$ and $[t]^+ = \{Q\}$; and for the other $x \in \mathscr{S}$, $[x]^+ = [x]^- = \emptyset$. So, $|P|^+ \cup |P|^- = \{s\}$ and $|Q|^+ \cup |Q|^- = \{t\}$. (**V1**) is clearly met. Again, (**V2**) is vacuously satisfied because $\overline{\beta}(s) = \overline{\beta}(t) = \emptyset$.

It is worth noting here that no states behave like possible worlds in $\mathscr{M}_1$. Recall that a possible world can plausibly be considered a state that is both modally sound and complete and atomically sound and complete. And there is no such state in $\mathscr{S}$. $\top$ is modally unsound. s is not only modally incomplete, but it is also atomically incomplete; and similarly for t and $\bot$.

Example 4. For a bit more complex example, let:

$$\mathscr{S} = \{\bot, s, t, w, w', u, \top\};$$

$$\sqsubseteq = \{\langle \bot, x\rangle : x \in \mathscr{S}\} \cup \{\langle x, \top\rangle : x \in \mathscr{S}\} \cup \{\langle x, u\rangle : x \in \mathscr{S} \setminus \{\top\}\} \cup \{\langle s, w'\rangle, \langle t, w'\rangle\} \cup \{\langle x, x\rangle : x \in \mathscr{S}\}.$$

$\langle \mathscr{S}, \sqsubseteq \rangle$ clearly is a state space. Then we let μ be a function with domain $\{w, u\}$ such that

$$\alpha(w) = \{w', s, t, \bot\}, \quad \beta(w) = \{\top\};$$
$$\alpha(u) = \{w\}, \quad \beta(u) = \{w, u\}.$$

Then $\Sigma_2 = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ can be represented in FIGURE 2 below.

Σ_2 clearly satisfies (**N1**)-(**N3**). The modal boundaries of Σ are w and w', and $\mathscr{S}^{\Diamond} = \{w, w', s, t, \bot\}$. (**N4**) is also met because everything in $\overline{\alpha}(w)$ is part of w'. Notice also that every state is hit by some arrow. So, (**N5i**) is satisfied. To verify that (**N5ii**) is satisfied, we only need to consider $\overline{\beta}(w) = \{\top\}$ because $\overline{\beta}(w') = \emptyset$. We check that for each $y \in \mathscr{S}^{\Diamond}$, if $y \notin \overline{\beta}(w)$ then there is no set U of states such that

$$\{y\} \sqcup (\mathscr{A}(U) \cup \mathscr{B}(U)) \subseteq \overline{\beta}(w) \qquad (*)$$

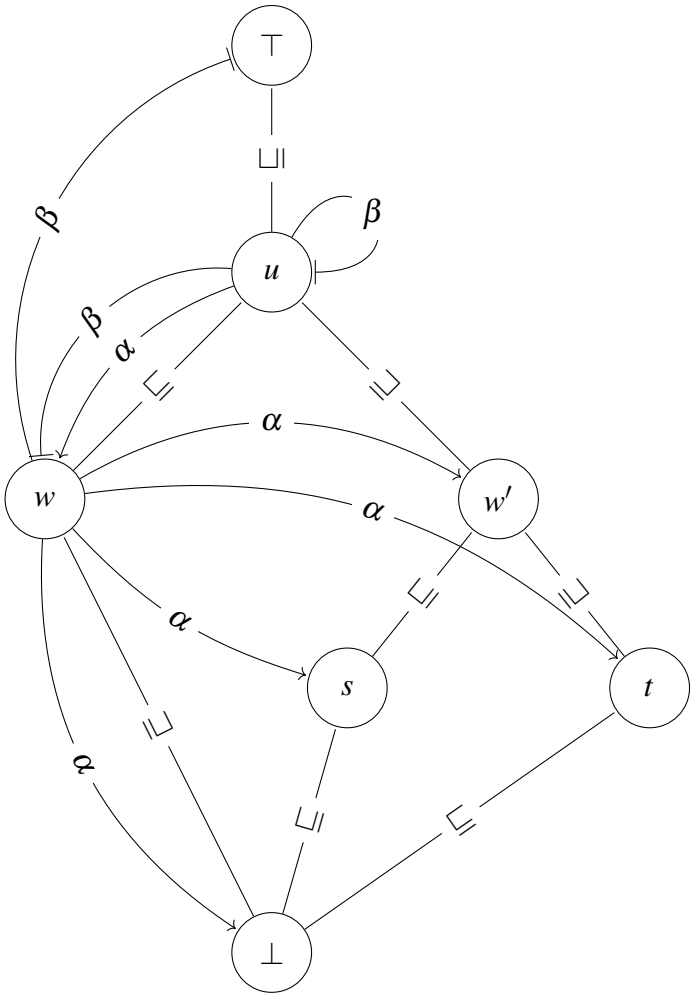

FIGURE 2.2. Diagrammatic representation of Σ_2

Let $y = w'$. Notice that if there is a set U satisfying (*) then $\mathscr{A}(U) \cup \mathscr{B}(U) = \{\top\}$. Since $\top$ is not a modal state, there is no $U \subseteq \mathscr{S}$ such that $\top \in \mathscr{A}(U)$. For all $U \subseteq \mathscr{S}$, moreover, either $\mathscr{B}(U) = \{\bot\}$ (if $U = \emptyset$), or $\mathscr{B}(U) \subseteq \{w, u\}$ (otherwise). This is because w and u are the only states that exactly bans a state and because $w \sqsubseteq u$. Hence there is no $U \subseteq \mathscr{S}$ such that $\mathscr{A}(U) \cup \mathscr{B}(U) = \{\top\}$. Therefore, no $U \subseteq \mathscr{S}$ satisfies (*) when $y = w'$. The same consideration applies to all the other states in $\mathscr{S}^{\Diamond}$. Hence (**N5ii**) is satisfied.

Let $\Gamma = \{P \wedge Q\}$, and define a normal m-model $\mathscr{M}_2$ based on Σ_2 by setting $[s]^+ = \{P\}$ and $[t]^+ = \{Q\}$, and, for the other $x \in \mathscr{S}$, $[x]^+ = [x]^- = \emptyset$. So, $|P|^+ \cup |P|^- = \{s\}$ and $|Q|^+ \cup |Q|^- = \{t\}$. (**V1**) is clearly met. To verify that (**V2**) is also met, note first that $\overline{\beta}(w') = \emptyset$. So we only need

to consider $\overline{\beta}(w)$. For any $x \in \mathscr{S}^{\Diamond}$, $\{x\} \sqcup (|P|^{+} \cup |P|^{-}) \subseteq \overline{\beta}(w)$ only if $x = \top$; and $\top \in \overline{\beta}(w)$. $|Q|^{+} \cup |Q|^{-}$ can be treated symmetrically. So (**V2**) is met.

Again, it is not difficult to see that no states behave like a possible world in $\mathscr{M}_2$. $\top$ and u are modally unsound. w, s and t are neither modally complete nor atomically complete. w' is atomically complete, but it is not modally complete.

Definition 4. For any set Γ of formulas and a formula A, A is a *consequence of* Γ *in* a normal m-model $\mathscr{M}$ for $\Gamma;A$ iff for all $s \in \mathscr{S}^{\Diamond}$ such that $s \rhd^{+} B$ for all $B \in \Gamma$, $s \not\rhd^{-} A$. A is a *consequence of* Γ *on* a normal m-space Σ iff A is consequence of Γ in all normal m-models based on Σ. A is *a consequence of* Γ *with respect to* a class $\mathfrak{S}$ of normal m-space iff A is a consequence of Γ on all normal m-spaces in $\mathfrak{S}$. In each case, A is said to be *valid* in the corresponding sense if $\Gamma = \emptyset$.

Or, equivalently, A is a consequence of Γ with respect to a class $\mathfrak{S}$ of normal m-spaces if and only if for all normal m-models $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ for $\Gamma;A$ based on normal m-spaces in $\mathfrak{S}$ and for all $s \in \mathscr{S}^{\Diamond}$ such that $s \rhd^{+} \Gamma$, $s \not\rhd^{-} A$. It should be obvious that this definition is a straightforward formalization of (*MCb*) in section 6.

3. Some basic results

Here we establish some basic results. The upshots will be, first, that every $s \in \mathscr{S}^{\Diamond}$ is consistent (Theorem 4), and second that the axiom K is valid in every normal m-model, i.e., valid with respect to the class of normal m-spaces (Theorem 5).

Proposition 1. *Let* Γ *be a set of formulas and* $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ *be a normal m-model of* Γ. *For each* $A \in Fml(\Gamma)$, $|A|^{+} \cup |A|^{-} \neq \emptyset$.

PROOF. By induction on A. The base case holds because we require that for every propositional variable $P \in At(\Gamma)$, $|P|^{+} \cup |P|^{-} \neq \emptyset$. For the induction step, assume that $|B|^{+} \cup |B|^{-} \neq \emptyset$ and that $|C|^{+} \cup |C|^{-} \neq \emptyset$. Let $A = \neg B$. Then

$$\begin{aligned} |A|^{+} \cup |A|^{-} &= |\neg B|^{+} \cup |\neg B|^{-} \\ &= |B|^{-} \cup |B|^{+}, \end{aligned}$$

which is nonempty by the I.H. Let $A = B \wedge C$. Notice:

$$|A|^+ \cup |A|^- = (|B|^+ \sqcup |C|^+) \cup (|B|^- \cup |C|^-).$$

So, it suffices to show that if $|B|^- = |C|^- = \emptyset$ then $|B|^+ \sqcup |C|^+ \neq \emptyset$. Suppose that $|B|^- = |C|^- = \emptyset$. Then, by the I.H., both $|B|^+$ and $|C|^+$ are nonempty. Hence $|B|^+ \sqcup |C|^+ \neq \emptyset$. The case where $A = B \vee C$ can be treated dually. Let $A = \Box B$. $|A|^+ \cup |A|^- = \mathscr{A}(|B|^-) \cup \mathscr{B}(|B|^-)$, which is nonempty by (**N5i**). □

Notice that Proposition 1 shows that every formula in $Fml(\Gamma)$ has either an exact verifier or falsifier in $\mathscr{M}$, which one would naturally want. This is why we have (**N5i**) in the formal definition of normal m-model.

Lemma 2. *Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model of Γ. Let s be a modal boundary and t be any state. For any formulas $A \in Fml(\Gamma)$, if $\{t\} \sqcup (|A|^+ \cup |A|^-) \subseteq \overline{\beta}(s)$ then $t \in \overline{\beta}(s)$.*

PROOF. Induction on A. The base case holds by (**V2**). Assume as the induction hypothesis that the property holds for B and C.

Let $A = \neg B$. Suppose that $\{t\} \sqcup (|A|^+ \cup |A|^-) \subseteq \overline{\beta}(s)$. Then, since $|A|^+ = |\neg B|^+ = |B|^-$ and $|A|^- = |\neg B|^- = |B|^+$, it follows that $\{t\} \sqcup (|B|^+ \cup |B|^-) \subseteq \overline{\beta}(s)$. By the I.H., then $t \in \overline{\beta}(s)$.

Let $A = B \wedge C$. Assume that $\{t\} \sqcup (|A|^+ \cup |A|^-) \subseteq \overline{\beta}(s)$. Since $|A|^+ = |B \wedge C|^+ = |B|^+ \sqcup |C|^+$ and $|A|^- = |B \wedge C|^- = |B|^- \cup |C|^-$, we have:

$$\{t\} \sqcup ((|B|^+ \sqcup |C|^+) \cup (|B|^- \cup |C|^-)) \subseteq \overline{\beta}(s),$$

which is equivalent to:

$$(\{t\} \sqcup (|B|^+ \sqcup |C|^+)) \cup (\{t\} \sqcup |B|^-) \cup (\{t\} \sqcup |C|^-) \subseteq \overline{\beta}(s).$$

Now, suppose that $|B|^+ \sqcup |C|^+ = \emptyset$. Then either $|B|^+ = \emptyset$ or $|C|^+ = \emptyset$. In the former case, it follows from Proposition 1 that $|B|^-$ is not empty; and $\{t\} \sqcup (|B|^+ \cup |B|^-) = \{t\} \sqcup |B|^- \subseteq \overline{\beta}(s)$. In the latter case, similarly, $\{t\} \sqcup (|C|^+ \cup |C|^-) \subseteq \overline{\beta}(s)$. Either way, it follows from the I.H. that $t \in \overline{\beta}(s)$.

Suppose, on the other hand, that $|B|^+ \sqcup |C|^+ \neq \emptyset$. Let $b \in |B|^+$. Then,

$$\{t \sqcup b\} \sqcup |C|^+ \subseteq \overline{\beta}(s),$$

because $\{t \sqcup b\} \sqcup |C|^+ \subseteq \{t\} \sqcup (|B|^+ \sqcup |C|^+)$. Notice also that

$$\{t \sqcup b\} \sqcup |C|^- \subseteq \overline{\beta}(s),$$

because $\{t\} \sqcup |C|^- \subseteq \overline{\beta}(s)$ and $\overline{\beta}(s)$ is upward closed. By the I.H., then, $t \sqcup b \in \overline{\beta}(s)$. Since b was an arbitrary element of $|B|^+$, it follows that $\{t\} \sqcup |B|^+ \subseteq \overline{\beta}(s)$. Then, by the I.H., $t \in \overline{\beta}(s)$. The case where $A = B \vee C$ can be treated dually.

Let $A = \Box B$. Suppose that $\{t\} \sqcup (|\Box B|^+ \cup |\Box B|^-) \subseteq \overline{\beta}(s)$. This means:

$$\{t\} \sqcup (\mathscr{A}(|B|^-) \cup \mathscr{B}(|B|^-)) \subseteq \overline{\beta}(s).$$

By (**N5ii**), $t \in \overline{\beta}(s)$. □

Lemma 3. *Let $\mathscr{M}$ be a normal m-model of Γ and A be any formula. For all states s,*

$$s \triangleright^+ \Box A \quad \Leftrightarrow \quad |A|^- \subseteq \overline{\beta}(s).$$

PROOF. The left-to-right direction is obvious. To see the other direction, suppose that $|A|^- \subseteq \overline{\beta}(s)$. Then, for each $t \in |A|^-$, there is $s_t \sqsubseteq s$ such that $t \in \beta(s_t)$. Then let f be a function that maps each $t \in |A|^-$ to s_t. f is a ban on $|A|^-$ and $\bigsqcup ran(f) \sqsubseteq s$. □

It is worth noting that this proof requires the axiom of choice.

Theorem 4. *Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model of Γ. For all $A \in Fml(\Gamma)$ and for all $s \in \mathscr{S}^{\Diamond}$, either $s \not\triangleright^+ A$ or $s \not\triangleright^- A$.*

PROOF. By induction on A. It suffices to check that if s is both atomically and modally sound, then either $s \not\triangleright^+ A$ or $s \not\triangleright^- A$. So, let s be an atomically and modally sound state.

The base case holds because s is atomically sound.

Let $A = \neg B$. By the I.H., either $s \not\triangleright^+ B$ or $s \not\triangleright^- B$. In the former case, $s \not\triangleright^- \neg B$, and in the latter case $s \not\triangleright^+ \neg B$.

Let $A = B \wedge C$. If $s \triangleright^+ B$ and $s \triangleright^+ C$, then it follows from the I.H. that $s \not\triangleright^- B$ and $s \not\triangleright^- C$ and hence that $s \not\triangleright^- B \wedge C$. Otherwise, $s \not\triangleright^+ B \wedge C$. The case where $A = B \vee C$ can be treated dually.

Finally, $A = \Box B$. If $s \triangleright^+ \Box B$ and $s \triangleright^- \Box B$, then it would follow from Lemma 3 that $|B|^- \subseteq \overline{\beta}(s)$. Also, $|B|^- \cap \overline{\alpha}(s) \neq \emptyset$; but this is impossible because s is modally sound. □

Theorem 5. *Let Γ be a set of formulas and $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model of Γ. For any $s \in \mathscr{S}^{\Diamond}$, $s \not\triangleright^- \Box(A \supset B) \supset (\Box A \supset \Box B)$ (where $A, B \in Fml(\Gamma)$).*

PROOF. Pick any $s \sqsubseteq \mathscr{S}^{\Diamond}$, and let t be a modal boundary containing s. It suffices to verify that $t \not\triangleright^- \Box(A \supset B) \supset (\Box A \supset \Box B)$. Assume, for contradiction, that $t \triangleright^- \Box(A \supset B) \supset (\Box A \supset \Box B)$. Then $t \triangleright^+ \Box(A \supset B)$, $t \triangleright^+ \Box A$ and $t \triangleright^- \Box B$. By Lemma 3, this means:

$$|A|^+ \sqcup |B|^- \subseteq \overline{\beta}(t), \quad |A|^- \subseteq \overline{\beta}(t), \quad \text{and } |B|^- \cap \overline{\alpha}(t) \neq \emptyset.$$

Let $u \in |B|^- \cap \overline{\alpha}(t)$. Then $\{u\} \sqcup (|A|^+ \cup |A|^-) \subseteq \overline{\beta}(s)$. By Lemma 2, then, $u \in \overline{\beta}(t)$. But then $u \in \overline{\alpha}(t) \cap \overline{\beta}(t)$, which contradicts the modal soundness of t. □

4. World model

In section 2, we have seen a couple of examples of normal m-models whose modal boundaries are neither modally nor atomically complete. These examples show that normal m-models do not necessarily contain possible worlds. For possible worlds should be both modally and atomically complete. In this section, we shall introduce a special subclass of normal m-models whose modal boundaries behave just like possible worlds.

Let $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ be an m-space. We say that a subset $\mathscr{R}$ of $\mathscr{S}$ is *worldly* if and only if $\mathscr{R}$ is nonempty and, for all $w \in \mathscr{R}$,

(W1) w is modally sound and complete; and

(W2) for all $t \in \mathscr{S}^{\Diamond}$, if $t \in \overline{\alpha}(w)$ then there is a state $w' \in \mathscr{R}$ such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$.

Notice here that (W2) is a version of robustness condition.

Definition 5. An m-space $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is said to be a *world space* (or, simply, *w-space*) if and only if it satisfies (**N2**), (**N3**), and

(**W**) the modal boundaries are worldly.

It should be clear that the definition is a straightforward formalization of the informal analysis of possible worlds (§5) that possible worlds are modally sound and complete states that are robust about the possibilities.

Proposition 6. *W-spaces are m-spaces.*

PROOF. Let Σ be a w-space. So it satisfies (**N2**), (**N3**) and (**R**). It is clear that (**R**) implies (**N1**) and (**N4**). (**N5i**) is also met because the modal boundaries are modally complete. To verify (**N5ii**), let w be a modal boundary and $t \in \mathscr{S}^{\Diamond}$. Suppose that there is a set U of states such that

$$\{t\} \sqcup (\mathscr{A}(U) \cup \mathscr{B}(U)) \subseteq \overline{\beta}(w).$$

We need to show that $t \in \overline{\beta}(w)$. Since w is modally complete, it suffices to show that $t \notin \overline{\alpha}(w)$. Assume, for contradiction, that $t \in \overline{\alpha}(w)$. Then it would follow from (**R**) that there is a modal boundary w' such $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. Since w' would also be modally complete, w' should extend some $u \in \mathscr{A}(U) \cup \mathscr{B}(U)$. Now, consider $t \sqcup u$. Notice that $t \sqcup u \sqsubseteq w'$ and $t \sqcup u \in \overline{\beta}(w)$. By (**N3**), then, $w' \in \overline{\beta}(w)$. Hence $w' \in \overline{\alpha}(w) \cap \overline{\beta}(w)$, contradicting the modal soundness of w. Therefore, $t \notin \overline{\alpha}(w)$. □

Definition 6. Let Γ be a set of formulas. We define a world model $\mathscr{M}$ to be $= \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$, where $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$ is a w-space and v is a valuation satisfying:

(**C**) each modal boundary is atomically sound and complete.

In a world model, therefore, each modal boundary either verifies or falsifies every propositional variable in the inexact sense.

Proposition 7. *W-models are normal m-models.*

PROOF. Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a w-model. It is obvious that (**C**) implies (**V1**). For if every modal boundary is atomically sound then everything below it must also be atomically sound. To check that $\mathscr{M}$ satisfies (**V2**), let w be a modal boundary and $t \in \mathscr{S}^{\Diamond}$. Assume that there is a propositional variable $P \in At(\Gamma)$ such that

$$\{t\} \sqcup (|P|^{+} \cup |P|^{-}) \subseteq \overline{\beta}(w).$$

We need to show that $t \in \overline{\beta}(w)$. Again, it suffices to show that $t \notin \overline{\alpha}(w)$. Suppose, for contradiction, that $t \in \overline{\alpha}(w)$. By (**N4**), there would be a modal boundary w' such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. By (**C**), then, w' would be atomically complete. Hence some part u of w' would belong to $(|P|^{+} \cup |P|^{-})$. Then it would follow from the assumption that $t \sqcup u \in \overline{\beta}(w)$. Since $t \sqcup u \sqsubseteq w'$, it would follow that $w' \in \overline{\beta}(w)$, contradicting the modal soundness of w. □

Theorem 8. *Let $\mathscr{M}$ be a world model of Γ and w be a modal boundary. Then for all formulas A, either $w \triangleright^{+} A$ or $w \triangleright^{-} A$, but not both.*

PROOF. By induction on A. The base case holds because of (**C**). The cases for the Boolean connectives are straightforward, so we omit them here. Let $A = \Box B$. We argue by cases. Suppose first that $|B|^{-} \subseteq \overline{\beta}(w)$. By Lemma 3, $w \triangleright^{+} \Box B$. Since w is modally sound, moreover, $|B|^{-} \cap \overline{\alpha}(w) = \emptyset$. So, $w \not\triangleright^{-} \Box B$. Now suppose that $|B|^{-} \not\subseteq \overline{\beta}(w)$. By Lemma 3, then, $w \not\triangleright^{+} \Box B$. Since w is modally complete, $|B|^{-} \cap \overline{\alpha}(w) \neq \emptyset$; so, $w \triangleright^{-} \Box B$. □

This theorem shows that the modal boundaries in a w-model consistently determine the truth value of every formula and hence are plausibly considered possible worlds.

5. Completeness result

We now establish the soundness and completeness results for the minimal system K of normal modal propositional logic with respect to the proposed truthmaker semantics. This result will be extended to stronger systems in the next section. Below we shall assume that the standard Kripke semantics is already familiar the present readership.

To establish the soundness result, we show that every normal m-model can be "completed" to a w-model, which can in turn be translated to an equivalent Kripke model. This shows that no theorem of K has a countermodel in the proposed truthmaker semantics. To prove the completeness result, on the other hand, we show that every Kripke model can be translated to an equivalent w-model, *a fortiori*, to an equivalent normal m-model. This shows that every formula, if not a theorem of K, has a countermodel in the proposed truthmaker semantics.

Translation from normal m-models to Kripke models

We first consider how to translate normal m-models to Kripke models. Let $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ be a normal m-space with $M = dom(\mu)$ and $\mu(s) = \langle \alpha(s), \beta(s) \rangle$. A *completion* of Σ is a w-space $\Sigma' = \langle \mathscr{S}, \sqsubseteq, \mu' \rangle$ with $M' = dom(\mu')$ and $\mu' = \langle \alpha'(s), \beta'(s) \rangle$ such that $M \subseteq M'$ and such that for each $s \in M$, $\alpha(s) \subseteq \alpha'(s)$ and $\beta(s) \subseteq \beta'(s)$. Intuitively, a completion Σ' is obtained from Σ by extending μ so as to make the modal boundaries modally complete.

One natural way of obtaining a completion is by "closing off," so to speak, the modal boundaries; that is, by letting each modal boundary be a modal state that bans every state that it does not allow (in the inexact sense) under the original μ. To make this precise, let $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ be a normal m-space with $dom(\mu) = M$ and $\mu(s) = \langle \alpha(s), \beta(s) \rangle$. Define μ^* as follows: $M^* = M \cup \{w : w \text{ is a modal boundary of } \Sigma\}$ and μ^* agrees with μ for all $s \in M$ except that for all modal boundaries w of Σ, $\beta^*(w) = \mathscr{S} \setminus \overline{\alpha}(w)$. Define the *closure* Σ^* of Σ to be $\langle \mathscr{S}, \sqsubseteq, \mu^* \rangle$. Then we can prove:

Proposition 9. *For every normal m-space* Σ, Σ^* *is a completion of* Σ.

PROOF. Let $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$ be a normal m-space with $dom(\mu) = M$ and $\mu(s) = \langle \alpha(s), \beta(s) \rangle$. Let $\Sigma^* = \langle \mathscr{S}, \sqsubseteq, \mu^* \rangle$ be as just defined. It is obvious from the definition of μ^* that for all $s \in M$, $\alpha(s) \subseteq \alpha^*(s)$ and $\beta(s) \subseteq \beta^*(s)$. So, it only remains to show that $\Sigma^* = \langle \mathscr{S}, \sqsubseteq, \mu^* \rangle$ is a w-space.

We first check that Σ^* satisfies (**N2**), i.e., that for all $s \in \mathscr{S}$, $\overline{\alpha^*}(s)$ is downward closed. Notice that for all states $s \in \mathscr{S}$, $\overline{\alpha^*}(s) = \overline{\alpha}(s)$ and $\overline{\alpha}(s)$ is downward closed because Σ is a normal m-space.

We now show that Σ^* satisfies (**N3**), i.e., that for all $s \in \mathscr{S}$, $\overline{\beta^*}(s)$ is upward closed. We argue by cases:

Case 1: s is an absolute possibility that is not a modal boundary of Σ. In this case, $\overline{\beta^*}(s) = \overline{\beta}(s)$. Hence $\overline{\beta^*}(s)$ is upward closed.

Case 2: s is a modal boundary of Σ. In this case, $\overline{\beta^*}(s) = \mathscr{S} \setminus \overline{\alpha}(s)$, which is upward closed because $\overline{\alpha}(s)$ is downward closed.

Case 3: s is not an absolute possibility of Σ. Suppose that $t \in \overline{\beta^*}(s)$. Pick any $t' \in \mathscr{S}$ with $t \sqsubseteq t'$. Now, either $t \in \overline{\beta}(s)$ or $t \in \overline{\beta^*}(w)$ for some modal boundary $w \sqsubseteq s$. Either way, $t' \in \overline{\beta^*}(s)$; for both $\overline{\beta}(s)$ and $\overline{\beta^*}(w)$ are upward closed.

Before turning to (**W**), observe that Σ and Σ^* have exactly the same modal boundaries. For the modally sound states of Σ remain modally sound in Σ^*, and similarly for the modally unsound states of Σ. Now we show that Σ^* satisfies (**W**). It is clear from the construction that the modal boundaries of Σ^* are modally sound and complete. Now let w be a modal boundary of Σ^* and $t \in \mathscr{S}^{\Diamond}$. Suppose that $t \in \overline{\alpha^*}(w)$. Recall that $\overline{\alpha^*}(w) = \overline{\alpha}(w)$. So, $t \in \overline{\alpha}(w)$. Since Σ satisfies (**N4**), it follows that there is a modal boundary w' of Σ such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. Since Σ and Σ^* have exactly the same modal boundary and $\overline{\alpha}(w) = \overline{\alpha^*}(w)$, it follows that there is a modal boundary w' of Σ^* such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha^*}(w)$. □

Now let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model. We define a *completion* of $\mathscr{M}$ to be a w-model $\langle \mathscr{S}, \sqsubseteq, \mu', v' \rangle$ such that $\langle \mathscr{S}, \sqsubseteq, \mu' \rangle$ is a completion of $\langle \mathscr{S}, \sqsubseteq, \mu \rangle$ and v' is such that (1) for each s, $[s]^+ \subseteq [s]'^+$ and $[s]^- \subseteq [s]'^-$, and (2) v' satisfies (**C**). Given any normal m-model $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$, let v^* be a valuation such that v^* agrees with v for all $s \in \mathscr{S}$ except that for all modal boundaries w of Σ^*, $[w]^{*-} = \{P \in At(\Gamma) : w \not\Vdash^+ P\}$; that is, v^* is obtained from v by letting each modal boundary exactly falsify all the propositional variables that it does not verify under the original v. Then define the *closure* $\mathscr{M}^*$ of $\mathscr{M}$ to be $\langle \mathscr{S}, \sqsubseteq, \mu^*, v^* \rangle$, where μ^* is as defined above. Then it is immediate from the definition that $\mathscr{M}^*$ satisfies (**C**). By Proposition 9, we thus have:

Proposition 10. *For every normal m-model $\mathscr{M}$, $\mathscr{M}^*$ is a completion of $\mathscr{M}$.*

Lemma 11. *Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model of Γ and let $\mathscr{M}' = \langle \mathscr{S}, \sqsubseteq, \mu', v' \rangle$ be a completion of $\mathscr{M}$. Then, for all $A \in Fml(\Gamma)$:*

$$\begin{aligned} \mathscr{M}, s \triangleright^+ A \quad &\Rightarrow \quad \mathscr{M}', s \triangleright^+ A; \\ \mathscr{M}, s \triangleright^- A \quad &\Rightarrow \quad \mathscr{M}', s \triangleright^- A. \end{aligned}$$

PROOF. By straightforward induction on A. □

Now we show that every w-model has an equivalent Kripke model. Let Γ be a set of formulas and $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a w-model of Γ. Define the corresponding Kripke model $K(\mathscr{M}) = \langle W^{\mathscr{M}}, R^{\mathscr{M}}, \Phi^{\mathscr{M}} \rangle$ of Γ as follows:

- $W^{\mathscr{M}}$ = the modal boundaries of $\mathscr{M}$;
- $R^{\mathscr{M}} = \{\langle w, w' \rangle \in W_{\mathscr{M}} \times W_{\mathscr{M}} : w' \in \overline{\alpha}(w)\}$;
- $\Phi^{\mathscr{M}}(w, P) = \begin{cases} T, & w \triangleright^+ P; \\ F, & \text{otherwise.} \end{cases}$

This clearly defines a Kripke model of Γ. The notion of a formula A's *being true at* a world w in a Kripke model, (in symbol, $w \models A$), is defined in the usual way.

Theorem 12. *Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a w-model of Γ and $K(\mathscr{M}) = \langle W^{\mathscr{M}}, R^{\mathscr{M}}, \Phi^{\mathscr{M}} \rangle$ be the corresponding Kripke model. For all formulas A and for all $w \in W^{\mathscr{M}}$,*

$$\begin{aligned} \mathscr{M}, w \triangleright^+ A \quad &\Leftrightarrow \quad w \models A \text{ in } K(\mathscr{M}); \\ \mathscr{M}, w \triangleright^- A \quad &\Leftrightarrow \quad w \not\models A \text{ in } K(\mathscr{M}). \end{aligned}$$

PROOF. By Theorem 8, it suffices to verify the first equivalence. By induction on A. Let P be a propositional variable. For all $w \in W^{\mathscr{M}}$:

$$\begin{aligned} w \triangleright^+ P \quad &\Leftrightarrow \quad \Phi(w, A) = T; \\ &\Leftrightarrow \quad w \models P. \end{aligned}$$

Let $A = \neg B$, for some B. Then

$$\begin{aligned} w \triangleright^+ \neg B &\Leftrightarrow w \triangleright^- B; && \text{then, by the I.H.,} \\ &\Leftrightarrow w \not\models B; \\ &\Leftrightarrow w \models \neg B. \end{aligned}$$

Let $A = B \wedge C$, for some B and C. Then

$$\begin{aligned} w \triangleright^+ B \wedge C &\Leftrightarrow w \triangleright^+ B \text{ and } w \triangleright^+ C; && \text{then, by the I.H.,} \\ &\Leftrightarrow w \models B \text{ and } w \models C; \\ &\Leftrightarrow w \models B \wedge C; \end{aligned}$$

The case where $A = B \vee C$ can be proved dually.

Finally, let $A = \Box B$ for some B.

$$\begin{aligned} w \triangleright^+ \Box B &\Leftrightarrow |B|^- \subseteq \overline{\beta}(w) && (1) \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(w' \triangleright^- B \Rightarrow w' \in \overline{\beta}(w)); && (2) \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(w' \notin \overline{\beta}(w) \Rightarrow w' \not\triangleright^- B); && \text{then, since } w' \text{ is modally sound and complete,} \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(w' \in \overline{\alpha}(w) \Rightarrow w' \not\triangleright^- B); && \text{then, by Theorem 8,} \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(w' \in \overline{\alpha}(w) \Rightarrow w' \triangleright^+ B); && \text{then, by the definition of } R^{\mathscr{M}}, \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(R^{\mathscr{M}}(w,w') \Rightarrow w' \triangleright^+ B); && \text{then, by the I.H.,} \\ &\Leftrightarrow (w' \in W^{\mathscr{M}})(R^{\mathscr{M}}(w,w') \Rightarrow w' \models B); \\ &\Leftrightarrow w \models \Box B. \end{aligned}$$

We check the equivalence between (1) and (2). Assume (1). Let $w' \in W_{\mathscr{M}}$. Suppose that $w' \triangleright^- B$, i.e., that w' extends some $t' \in |B|^-$. Since $t' \in \overline{\beta}(w)$, $w' \in \overline{\beta}(w)$. Conversely, assume (2). Let $t \in |B|^-$. Assume, for contradiction, that $t \notin \overline{\beta}(w)$. Since w is modally complete, $t \in \overline{\alpha}(w)$. So

there is $w' \in W^{\mathscr{M}}$ such that $t \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. But then (2) implies that $w' \in \overline{\beta}(w)$. So $w' \in \overline{\alpha}(w) \cap \overline{\beta}(w)$, which contradicts the modal soundness of w. So, $|B|^- \subseteq \overline{\beta}(w)$. □

Translation from Kripke models to normal m-models

We consider the converse embedding of Kripke models into normal m-models. There is a sense in which the Kripke models form a special subclass of w-models. Let $K = \langle W, R, \Phi \rangle$ be a Kripke model of Γ, where W is a nonempty set, R is a binary relation on W, and Φ is a valuation that assigns a truth-value to every $P \in At(\Gamma)$ at each $w \in W$. For each $w \in W$, let $R_w = \{w' \in W : R(w, w')\}$.

Define the corresponding w-space $\Sigma(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K \rangle$ as follows:

- $\mathscr{S}^K = W \cup \{\bot, \top\}$.
- $\sqsubseteq^K = \{\langle x, x\rangle : x \in \mathscr{S}\} \cup \{\langle \bot, x\rangle : x \in \mathscr{S}\} \cup \{\langle x, \top\rangle : x \in \mathscr{S}\}$
- μ^K is a function with domain W such that for all $w \in W$,

$$\alpha(w) = \emptyset \text{ and } \beta(w) = \mathscr{S}, \qquad \text{if } R_w = \emptyset;$$

$$\alpha(w) = R_w \cup \{\bot\} \text{ and } \beta(w) = \mathscr{S} \setminus \alpha(w), \qquad \text{if } R_w \neq \emptyset.$$

and such that $\alpha(\top) = \beta(\top) = \{\top\}$.

That is, $\Sigma(K)$ is obtained from K by adding the bottom and top states, placing in between the members of W as mutually incomparable states and then setting up the allowing and banning relations in the obvious way.

Example 5. Define a Kripke frame $K_1 = \langle W, R \rangle$ as follows:

- $W = \{w_1, w_2\}$
- $R = \{\langle w_1, w_2\rangle\}$

We may represent K with the following diagram, where thick triangle arrows are used to indicate the accessibility relation R:

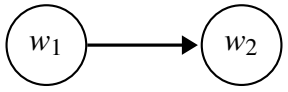

FIGURE 2.3. Diagrammatic representation of K_1

To obtain the corresponding w-space $\Sigma(K_1)$, we first let $\mathscr{S}^{K_1} = W \cup \{\bot, \top\} = \{w_1, w_2, \bot, \top\}$. Then we let w_1 and w_2 be mutually incomparable states between $\bot$ and $\top$. To set up μ^{K_1}, we first let $dom(\mu^{K_1}) = W \cup \{\top\}$. Consider w_1. Since $R_{w_1} = \{w_2\}$, we let w_1 allow $\alpha(w_1) = \{\bot, w_2\}$ and $\beta(w_1) = \{\top\}$; and since $R_{w_2} = \emptyset$, let $\alpha(w_2) = \emptyset$ and $\beta(w_2) = \{w_1, w_2, \bot, \top\}$. Finally, we set $\alpha(\top) = \beta(\top) = \{\top\}$. Then $\Sigma(K_1)$ can be diagrammatically represented thus:

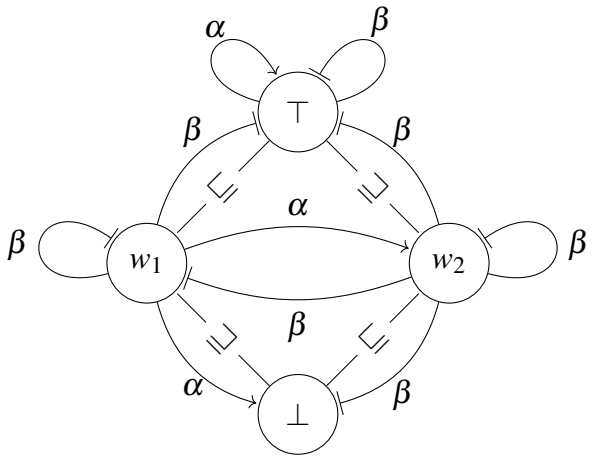

FIGURE 2.4. Diagrammatic representation of $\Sigma(K_1)$

We now define the corresponding w-model $\mathscr{M}(K)$ by adjoining to $\Sigma(K)$ a valuation=v^K defined as follows: for all $w \in W$ and $P \in At(\Gamma)$,

$$P \in [w]^+ \quad \Leftrightarrow \quad \Phi(w,P) = T;$$

$$P \in [w]^- \quad \Leftrightarrow \quad \Phi(w,P) = F;$$

In a sense, the Kripke models can be thought of as constituting a special subclass of w-models in which each world consists, as it were, of "one great fact."

Proposition 13. *Let $K = \langle W, R, \Phi \rangle$ be a Kripke model. Then $\Sigma(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K \rangle$ is a w-space and hence a normal m-space.*

PROOF. Let $K = \langle W, R, \Phi \rangle$ be a Kripke model. Let $\Sigma(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K \rangle$ be defined as above. Clearly $\langle \mathscr{S}^K, \sqsubseteq^K \rangle$ is a state space. It is also immediate from the definition of μ^K that $\Sigma(K)$ satisfies (**N2**) and (**N3**). To check that $\Sigma(K)$ satisfies (**W**), notice first that W forms the modal boundaries of

$\Sigma(K)$. And it is clear from the definition that every $w \in W$ is modally sound and complete. Now let $w \in W$ and $s \in (\mathscr{S}^K)^{\Diamond}$. Suppose that $s \in \overline{\alpha}(w)$. We need to check that there is a modal boundary $w' \in \mathscr{S}^K$ such that $s \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. Notice that when $s \in \overline{\alpha}(w)$, either $s = b$ or $s = w^*$ for some $w^* \in R_w$. In the latter case, we can let w^* be the desired w' itself. In the former case, notice first that, by construction, $R_w \neq \emptyset$. Picking any $w' \in R_w$, therefore, we have: $b \sqsubseteq w'$ and $w' \in \overline{\alpha}(w)$. Hence $\Sigma(K)$ satisfies (**W**). □

Lemma 14. *Let $K = \langle W, R, \Phi \rangle$ be a Kripke model. Then $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K, v^K \rangle$ is a w-model and hence a normal m-model.*

PROOF. It suffices to show that $\mathscr{M}(K)$ satisfies (**C**). Observe that in the original Kripke model, either $w \models P$, in which case $w \triangleright^+ P$, or $w \not\models P$, in which case $w \triangleright^- P$, but never both. □

Theorem 15. *Let $K = \langle W, R, \Phi \rangle$ be a Kripke model of Γ and $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K, v^K \rangle$ be the corresponding w-model. For any $w \in W$ and any $A \in Fml(\Gamma)$,*

$$w \models A \text{ in } K \quad \Leftrightarrow \quad w \triangleright^+ A \text{ in } \mathscr{M}(K);$$

$$w \not\models A \text{ in } K \quad \Leftrightarrow \quad w \triangleright^- A \text{ in } \mathscr{M}(K).$$

PROOF. By Theorem 8, it suffices to verify the first equivalence. By induction on A. The base case is immediate from the construction and the cases for the Boolean connectives are straightforward; so we omit them. Let $A = \Box B$. To verify the first equivalence, suppose first that $w \models \Box B$ in K. By Lemma 3, it suffices to show that $|B|^- \subseteq \overline{\beta}(w)$ in $\mathscr{M}(K)$. Assume, for contradiction, that $|B|^- \not\subseteq \overline{\beta}(w)$. Since w is modally complete, $|B|^- \cap \overline{\alpha}(w) \neq \emptyset$. By (**W**), then, there would be a $w' \in W$ such that $w' \in \overline{\alpha}(w)$ and $w' \triangleright^- B$. By the I.H., then, $w' \in R_w$ and $w' \not\models B$ in K. But this contradicts the assumption that $w \models \Box B$. Conversely, suppose that $w \triangleright^+ \Box B$ in $\mathscr{M}(K)$. We show that for all $w' \in W$, if $w' \not\models B$ in K then $w' \notin R_w$. Pick an arbitrary w' and assume that $w' \not\models B$ in K. By the I.H., $w' \triangleright^- B$ in $\mathscr{M}(K)$. Since $w \triangleright^+ \Box B$, it follows that $w' \in \overline{\beta}(w)$; therefore, $w' \notin R_w$. □

We now turn to consequence. For any set Γ formulas and a formula A, we shall write $\Gamma \models A$ to mean that in all Kripke models $K = \langle W, R, \Phi \rangle$ of $\Gamma; A$ and all worlds w in W, if $w \models B$ for all $B \in \Gamma$

then $w \models A$.[1] We say that $\Gamma \mid\!\models A$ if and only if A is a consequence of Γ with respect to the class of normal m-spaces. In other words, $\Gamma \mid\!\models A$ if and only if for all normal m-models $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ of $\Gamma; A$ and for all $s \in \mathscr{S}^{\Diamond}$ such that $s \triangleright^{+} \Gamma$, $s \not\triangleright^{-} A$. It is immediate from Definition 4 that A is valid with respect to the class of normal m-models just in case $\emptyset \mid\!\models A$.

Theorem 16. *For all sets* Γ *of sentences and a sentence* A*,*

$$\Gamma \models A \quad \Leftrightarrow \quad \Gamma \mid\!\models A.$$

PROOF. Suppose that $\Gamma \not\mid\!\models A$. For some normal m-model $\mathscr{M}$ and for some state $s \in \mathscr{S}^{\Diamond}$, $s \triangleright^{+} B$ for all $B \in \Gamma$ and $s \triangleright^{-} A$. Let $\mathscr{M}'$ be a completion of $\mathscr{M}$. Consider the corresponding Kripke model $K(\mathscr{M}')$. By Lemma 11 and Theorem 12, it follows that there is a world $w \in W^{\mathscr{M}}$ such that $w \models B$ for all $B \in \Gamma$ and $w \not\models A$ in $K(\mathscr{M})$; that is, $\Gamma \not\models A$.

Conversely, suppose that $\Gamma \not\models A$. For some Kripke model $K = \langle W, R, \Phi \rangle$ and some world $w \in W$, $w \models B$ for all $B \in \Gamma$ and $w \not\models A$. Let $\mathscr{M}(K) = \langle \mathscr{S}^{K}, \sqsubseteq^{K}, \mu^{K}, v^{K} \rangle$ be the corresponding w-model as defined above. It follows from Theorem 15 that in $\mathscr{M}(K)$, $w \triangleright^{+} B$ for all $B \in \Gamma$ and $w \triangleright^{-} A$. Since $w \in (\mathscr{S}^{K})^{\Diamond}$, $\Gamma \not\mid\!\models A$. □

Let's write $\Gamma \vdash_{\mathsf{K}} A$ to mean that A is derivable from Γ in the system K. Given the soundness and completeness results for the Kripke semantics, we have:

Corollary 17. *For all sets* Γ *of sentences and a sentence* A*,*

$$\Gamma \vdash_{\mathsf{K}} A \quad \Leftrightarrow \quad \Gamma \mid\!\models A.$$

6. Truthmaker semantical analysis of modal axioms

One of the main advantages of the Kripke semantics is that it offers the "reduction" of various modal axioms to conditions on frames. The result is well-known as the correspondence theorem, summarized in TABLE 1 below (adopted from [4]). The current approach analyzes the accessibility

[1] This is sometimes called the *local* consequence relation in the context of the Kripke semantics.

relation between possible worlds in terms of the allowing and banning relations on states. So it is natural to seek a truthmaker semantical analogue of the correspondence theorem.

TABLE 2.5. Some well-known modal axioms and the corresponding frame conditions

Modal Axiom	Scheme	Frame Condition
D	$\Box A \supset \Diamond A$	Serial
T	$\Box A \supset A$	Reflexive
4	$\Box A \supset \Box\Box A$	Transitive
B	$A \supset \Box\Diamond A$	Symmetric
5	$\Diamond A \supset \Box\Diamond A$	Euclidean

Let's say that a Kripke model $K = \langle W, R, \Phi \rangle$ is *based on* a Kripke frame $\langle W, R \rangle$. Given a class $\mathfrak{F}$ of Kripke frames, we also say that K is *in* $\mathfrak{F}$ if and only if K is based on a frame in $\mathfrak{F}$. TABLE 2 provides the standard names of some well-known classes of frames.

TABLE 2.6. Some well-known classes of Kripke frames

Class	Frame Conditions
D	Serial
T	Reflexive
4	Transitive
B	Symmetric
5	Euclidean
S4	Reflexive, Transitive
S5	Reflexive, Transitive, Symmetric

We shall show that for each of the classes of Kripke frames, there is a class of normal m-spaces that is equivalent to it. Let's say that a normal m-model $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ is *based on* a normal m-space $\Sigma = \langle \mathscr{S}, \sqsubseteq, \mu \rangle$. A normal m-model $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ *is in* a class $\mathfrak{S}$ of normal m-spaces if and only if $\mathscr{M}$ is based of a normal m-space in $\mathfrak{S}$. We say that a class $\mathfrak{F}$ of Kripke frames is *equivalent* to a class $\mathfrak{S}$ of normal m-spaces if and only if (1) for each Kripke model K in $\mathfrak{F}$, the corresponding normal m-model $\mathscr{M}(K)$ is in $\mathfrak{S}$, and (2) for each normal m-model $\mathscr{M}$ in $\mathfrak{S}$, the corresponding Kripke model $K(\mathscr{M}^*)$ is in $\mathfrak{F}$, where $\mathscr{M}^*$ is the closure of $\mathscr{M}$ as defined above. Notice here that we pick a particular way of completing the modal boundaries of a normal m-model for reasons that will become clearer later on (see the comment to the proof of Theorem 20 below).

Now we shall list five conditions on normal m-spaces:

(TM-D) For all modal boundaries w, $\overline{\alpha}(w) \neq \emptyset$.

(TM-T) For all modal boundaries w, $w \in \overline{\alpha}(w)$.

(TM-4) For all modal boundaries w and w' and for any state t, if $t \notin \overline{\alpha}(w)$ and $t \in \overline{\alpha}(w')$, then $w' \in \overline{\beta}(w)$.

(TM-B) For all modal boundaries w and w', if $w \notin \overline{\alpha}(w')$, then $w' \in \overline{\beta}(w)$.

(TM-5) For all modal boundaries w and w' and for any state t, if $t \in \overline{\alpha}(w)$ and $t \notin \overline{\alpha}(w')$, then $w' \in \overline{\beta}(w)$.

The conditions on normal m-spaces may seem quite complex at first glance. However, it is not so difficult to see that each condition offers an analysis of the corresponding frame condition in terms of the allowing and banning relations. (TM-T), for example, says that every modal boundary w allows itself. From the way we translated normal m-models to Kripke models, we can easily see that any normal m-model satisfying (TM-T) translates to a Kripke model in **T**. Using these conditions, we define the classes of normal m-models that correspond to those of Kripke frames:

TABLE 2.7. Some classes of normal m-spaces and their defining conditions

Class	Conditions on normal m-spaces
$\mathfrak{S}(\mathbf{D})$	(TM-D)
$\mathfrak{S}(\mathbf{T})$	(TM-T)
$\mathfrak{S}(\mathbf{4})$	(TM-4)
$\mathfrak{S}(\mathbf{B})$	(TM-B)
$\mathfrak{S}(\mathbf{5})$	(TM-5)
$\mathfrak{S}(\mathbf{S4})$	(TM-T), (TM-4)
$\mathfrak{S}(\mathbf{S5})$	(TM-T), (TM-4), (TM-B)

Then we can show that each of these classes is equivalent to the corresponding class of Kripke frames.

Theorem 18. $\mathfrak{S}(\mathbf{D})$ *is equivalent to* **D**.

PROOF. Let a normal m-model $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be in $\mathfrak{S}(\mathbf{D})$. Let $\mathscr{M}^* = \langle \mathscr{S}, \sqsubseteq, \mu^* \rangle$ be the closure of $\mathscr{M}$. Then consider the corresponding Kripke model $K(\mathscr{M}^*) = \langle W^{\mathscr{M}^*}, R^{\mathscr{M}^*}, \Phi^{\mathscr{M}^*} \rangle$. It should be obvious from the construction that, for all $w \in W^{\mathscr{M}}$, $R_w \neq \emptyset$; hence $K(\mathscr{M})$ is in **D**.

Conversely, suppose that a Kripke model $K = \langle W, R, \Phi \rangle$ is in **D**. Then the corresponding normal m-model $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K, v^K \rangle$ clearly satisfies (TM-D). □

Similarly, it is also easy to check (so we omit the proof):

Theorem 19. $\mathfrak{S}(\mathbf{T})$ *is equivalent to* **T**. □

Theorem 20. $\mathfrak{S}(\mathbf{4})$ *is equivalent to* **4**.

PROOF. Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model in $\mathfrak{S}(\mathbf{4})$. Let $\mathscr{M}^* = \langle \mathscr{S}, \sqsubseteq, \mu^*, v^* \rangle$ be the closure of $\mathscr{M}$. We show that $\mathscr{M}^*$ also satisfies (TM-4). Pick any modal boundary w and w', and t be any state. Suppose that $t \notin \overline{\alpha}^*(w)$ and $t \in \overline{\alpha}^*(w')$. Since μ^* agrees with μ concerning what's allowed by each state, it follows that $t \notin \overline{\alpha}(w)$ and $t \in \overline{\alpha}(w')$. Since $\mathscr{M}$ is assumed to satisfy (TM-4), $w' \in \overline{\beta}(w)$. Therefore, $w' \in \overline{\beta}^*(w)$. Now consider the corresponding Kripke model $K(\mathscr{M}^*) = \langle W^{\mathscr{M}^*}, R^{\mathscr{M}^*}, \Phi^{\mathscr{M}^*} \rangle$. Suppose that $R^{\mathscr{M}^*}(w_1, w_2)$ and $R^{\mathscr{M}^*}(w_2, w_3)$. In $\mathscr{M}^*$, $w_2 \in \overline{\alpha}^*(w_1)$ and $w_3 \in \overline{\alpha}^*(w_2)$. So, $w_2 \notin \overline{\beta}^*(w_1)$. By (TM-4), either $w_3 \in \overline{\alpha}^*(w_1)$ or $w_3 \notin \overline{\alpha}^*(w_2)$. Since $w_3 \in \overline{\alpha}^*(w_2)$, it follows that $w_3 \in \overline{\alpha}^*(w_1)$. Therefore, $R^{\mathscr{M}^*}(w_1, w_3)$.

Conversely, suppose that a Kripke model $K = \langle W, R, \Phi \rangle$ is in **4**. Let $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K, v^K \rangle$ be the corresponding normal m-model. We show that $\mathscr{M}(K)$ satisfies (TM-4). Let w_1 and w_2 be modal boundaries and t be any state. Suppose that $w_2 \notin \overline{\beta}(w_1)$ and that $t \in \overline{\alpha}(w_2)$. We need to show that $t \in \overline{\alpha}(w_1)$. Since $\mathscr{M}(K)$ is a w-model, $w' \in \overline{\alpha}(w_2)$. Since we also assumed that $t \in \overline{\alpha}(w_2)$, there is also a modal boundary w_3 such that $t \sqsubseteq w_3$ and $w_3 \in \overline{\alpha}(w_2)$. In the original Kripke model K, then $R(w_1, w_2)$ and $R(w_2, w_3)$. Since K is assumed to be in **4**, it follows that $R(w_1, w_3)$. In $\mathscr{M}(K)$, therefore, $w_3 \in \overline{\alpha}(w_1)$. Hence $t \in \overline{\alpha}(w)$. □

Theorem 21. $\mathfrak{S}(\mathbf{B})$ *is equivalent to* **B**.

PROOF. Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model in $\mathfrak{S}(B)$. Let $\mathscr{M}^* = \langle \mathscr{S}, \sqsubseteq, \mu^*, v^* \rangle$ be the closure of $\mathscr{M}$. We first show that $\mathscr{M}^*$ is also in $\mathfrak{S}(B)$. Pick any modal boundaries w and w'. Suppose that $w \notin \overline{\alpha}^*(w')$. Then, by construction, $w \notin \overline{\alpha}(w')$. So, $u \in \overline{\beta}(w)$. Therefore, $w \in \overline{\beta}^*(w')$. Now, consider the corresponding Kripke model $K(\mathscr{M}^*) = \langle W^{\mathscr{M}^*}, R^{\mathscr{M}^*}, \Phi^{\mathscr{M}^*} \rangle$. We need to show that $K(\mathscr{M}^*)$ is in **B**. Pick any two worlds w_1 and w_2. Suppose that $R^{\mathscr{M}^*}(w_1, w_2)$. Assume, for

contradiction, that $R^{\mathscr{M}^*}(w_2, w_1)$ does not hold. In $\mathscr{M}^*$, then, $w_2 \in \overline{\alpha}^*(w_1)$ and $w_1 \notin \overline{\alpha}^*(w_2)$. By (TM-B), then, $w_2 \in \overline{\beta}^*(w_1)$ But this contradicts the modal soundness of w_1. Hence $R^{\mathscr{M}^*}(w_2, w_1)$ holds. So, $K(\mathscr{M}^*)$ is in **B**.

Conversely, suppose that a Kripke model $K = \langle W, R, \Phi \rangle$ is in **B**. Let $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K, v^K \rangle$ be the corresponding m-space. Let w and w' be modal boundaries. Suppose that $w \notin \overline{\alpha}(w')$. Then $R(w', w)$ does not hold in K. Since K is assumed to be in **B**, $R(w, w')$ also fails in K. In $\mathscr{M}(K)$, therefore, $w' \notin \overline{\alpha}(w)$. Since $\Sigma(K)$ is a w-space, $w' \in \overline{\beta}(w)$. □

Theorem 22. $\mathfrak{S}(\mathbf{5})$ *is equivalent to* **5**.

PROOF. Let $\mathscr{M} = \langle \mathscr{S}, \sqsubseteq, \mu, v \rangle$ be a normal m-model in $\mathfrak{S}(5)$. Let $\mathscr{M}^* = \langle \mathscr{S}, \sqsubseteq, \mu^* \rangle$ be the closure. We first show that $\mathscr{M}^*$ is also in $\mathfrak{S}(5)$. Pick modal boundaries w and w'. Suppose that $t \in \overline{\alpha}^*(w)$ and $t \notin \overline{\alpha}^*(w')$. In $\mathscr{M}$, then, $t \in \overline{\alpha}(w)$ and $t \notin \overline{\alpha}(w')$. Since $\mathscr{M}$ is assumed to be in $\mathfrak{S}(5)$, it follows that $w' \in \overline{\beta}(w)$ in $\mathscr{M}$. Therefore, $w' \in \overline{\beta}^*(w)$. Consider the corresponding Kripke model $K(\mathscr{M}^*) = \langle W^{\mathscr{M}^*}, R^{\mathscr{M}^*}, \Phi^{\mathscr{M}^*} \rangle$. Suppose that $R^*(w_1, w_2)$ and $R^*(w_1, w_3)$. We need to show that $R^*(w_2, w_3)$. In $\mathscr{M}^*$, then, $w_2 \in \overline{\alpha}^*(w_1)$ and $w_3 \in \overline{\alpha}^*(w_1)$. From the former, we have: $w_2 \notin \overline{\beta}^*(w_1)$. Since $\mathscr{M}^*$ in in $\mathfrak{S}(5)$, either $w_3 \notin \overline{\alpha}^*(w_1)$ or $w_3 \in \overline{\alpha}^*(w_2)$. Since $w_3 \in \overline{\alpha}^*(w_1)$, it follows that $w_3 \in \overline{\alpha}^*(w_2)$. In $K(\mathscr{M}^*)$, therefore, $R(w_2, w_3)$. Hence $K(\mathscr{M}^*)$ is in **5**.

Conversely, suppose that a Kripke model $K = \langle W, R, \Phi \rangle$ is in **5**. Let $\mathscr{M}(K) = \langle \mathscr{S}^K, \sqsubseteq^K, \mu^K \rangle$ be the corresponding normal m-model. Let w_1 and w_2 be modal boundaries and t be any state. Suppose that $t \in \overline{\alpha}(w_1)$ and $t \notin \overline{\alpha}(w_2)$. We verify that $w_2 \in \overline{\beta}(w_1)$. Since $\mathscr{M}(K)$ is a w-model, $t \in \overline{\beta}(w_2)$. Now let w_3 be a world such that $t \sqsubseteq w_3$ and $w_3 \in \overline{\alpha}(w)$. Since $t \in \overline{\beta}(w_2)$ and $t \sqsubseteq w_3$, $w_3 \in \overline{\beta}(w_2)$. In the original Kripke model K, therefore, $R(w_1, w_3)$ holds and $R(w_2, w_3)$ fails. Since K is assumed to be in **5**, it follows that $R(w_1, w_2)$ does not hold. So, $w_2 \in \overline{\beta}(w)$. □

Corollary 23. $\mathfrak{S}(\mathbf{S4})$ *and* $\mathfrak{S}(\mathbf{S5})$ *are respectively equivalent to* **S4** *and* **S5**. □

Comment: It should be clear from the inspection that the proofs of Theorems 20-22 make use of the definition of μ^*. Hence we may say that $\mathfrak{S}(\mathbf{4})$, $\mathfrak{S}(\mathbf{B})$, $\mathfrak{S}(\mathbf{5})$ are equivalent, respectively, to **4**, **B**, **5** *modulo the "closing off" construction*. This suggests that the conditions on normal m-models need to be fine-tuned depending on how they are to be completed. In contrast, observe that the

proofs of Theorems 18 and 19 do not depend on the definition of μ^*. So it seems to be for modal formulas of degree 2 (or higher) that we need to take into account how normal m-models are to be completed. These considerations naturally lead to questions of some technical interest, such as whether there are conditions on normal m-models that translate to transitivity, symmetry and Euclidean condition on Kripke frames "categorically," i.e., independently of how to complete normal m-models, and, if there are no such categorical conditions, whether it holds for all modal axioms of degree 2 or higher.

With these results, we can easily establish the soundness and completeness results for some well-known family of systems of normal propositional modal logic. Let D, T, B, $\mathsf{K4}$, $\mathsf{S4}$ and $\mathsf{S5}$ be the systems of normal propositional modal logic that are characterized by (i.e., sound and complete with respect to) **D**, **T**, **4**, **B**, **K4**, **S4** and **S5**, respectively. Let's say that a system S is *sound* with respect to a class $\mathfrak{S}$ of normal m-models if and only if every theorem of S is valid with respect to $\mathfrak{S}$. S is said to be *complete* with respect to $\mathfrak{S}$ if and only if every formula that is valid with respect to $\mathfrak{S}$ is a theorem of S. Then we have:

Corollary 24. *The systems* D, T, B, $\mathsf{K4}$, $\mathsf{S4}$ *and* $\mathsf{S5}$ *are sound and complete with respect to* $\mathfrak{S}(\boldsymbol{D})$, $\mathfrak{S}(\mathbf{T})$, $\mathfrak{S}(\mathbf{4})$, $\mathfrak{S}(\mathbf{B})$, $\mathfrak{S}(\mathbf{S4})$ *and* $\mathfrak{S}(\mathbf{S5})$, *respectively.*

PROOF. We shall consider the system D; the other cases are similar. We first check that D is sound with respect to $\mathfrak{S}(\mathbf{D})$. Assume, for contradiction, that some theorem A of D is not valid with respect to $\mathfrak{S}(\mathbf{D})$. Let $\mathscr{M}$ be a normal m-model in $\mathfrak{S}(\mathbf{D})$ in which A is exactly falsified by some absolute possibility. Then it would follow from Theorems 12 and 18 that there is a Kripke model in **D**—namely $K(\mathscr{M}^*)$—in which A is not true at some world. But this contradicts the standard soundness result for D with respect to the class **D** of Kripke frames.

Conversely, we now check that D is complete with respect to $\mathfrak{S}(\mathbf{D})$. Let A be a formula that is not a theorem of D. Then it follows from the standard completeness result for D with respect to **D** that there is a Kripke model K in which A is not true at some world. Then it follows from Theorems 15 and 18 that there is a normal m-model in $\mathfrak{S}(\mathbf{D})$—namely $\mathscr{M}(K)$—in which A is exactly falsified by some absolute possibility. Hence A is not valid with respect to $\mathfrak{S}(\mathbf{D})$. □

7. Concluding remarks

In this chapter, I have developed a bilateralist truthmaker semantics for normal modal propositional logic. The soundness and completeness results for a well-known family of the systems of normal modal propositional logic were established. This provides a formal vindication that my truthmaker semantical analysis of modal statements are adequate.

One major advantage of the present semantics over the standard Kripke semantics is that it does not require the existence of possible worlds, but only of their parts. It also enables us to understand possible worlds or their parts as structured entities as opposed to simple points of evaluation without any internal structure. Given also the natural correspondences between the present semantics and some of the best known multi-valued logics as discussed in the previous chapter, moreover, the present semantics promises to offer a unifying formal framework for multi-valued modal logics; but a proper investigation into this is a task for another day.

CHAPTER 3

Explanation and Modality: On why the Swampman is still worrisome to teleosemanticists

1. Introduction

What makes necessary truths necessary? In chapters 1 and 2, I approached this problem from the perspective of truthmaker semantics. I proposed Kripke's Principle as a formal, logical answer to this problem: what makes a proposition necessary is that which excludes all the ways in which it might be falsified. This leaves the substantive, philosophical question unanswered: what exactly is it that excludes the ways in which a proposition might be falsified? This chapter aims to address this question; and to this end, I turn to the Swampman objection to teleosemantics.

Many have thought that Davidson's Swampman scenario raises a serious problem for teleosemantics. For it appears to be possible from the scenario that there are completely ahistorical creatures with beliefs, which contradicts the theory. In a series of papers (2001, 2006, 2016), David Papineau argues that the scenario is not even the start of an objection to teleosemantics as a scientific reduction of belief. No one would object to the H_2O theory of water simply on the basis of an apparent possibility that water might have another chemical composition, say XYZ—then why think that the Swampman scenario creates any problem for teleosemantics? In this chapter, I argue against Papineau that there is a crucial difference between the two cases. In this chapter, I am concerned with two *diagnostic* questions, viz., whether or not the Swampman scenario poses any problem for teleosemantics and, if so, what exactly it is. Any substantive claims about teleosemantics must remain tentative.

The plan for this chapter is as follows. After briefly reviewing the basic ideas of teleosemantics and the Swampman objection (sections 2 and 3), I first argue that teleosemantics as a reductive

* With minor changes, this chapter was previously published as: Kim, D. (2021). Explanation and modality: on why the Swampman is still worrisome to teleosemanticists. *Synthese* 199: 2817-2839.

account of belief is committed to the modal claim that beliefs are necessarily selectional states (sections 4 and 5). I then argue from Kripke's Principle that in order to establish the truth of the modal claim, teleosemanticists should explain away the apparent possibility of the Swampman's having beliefs as unreal (sections 6 and 7). How can this be done?; how in general can an apparent possibility be explained away as unreal? In section 8, I first discuss this question with respect to some of the well-known cases of reductive accounts. In the case of the H_2O theory of water, I argue, the XYZ scenario can be explained away by appeal to the general principle that every chemical substance has its chemical composition as a necessary property; and this principle is analytic in the sense that it is derivable from the theoretical definition of chemical substance. Like considerations apply to other well-known reductive theories as well, such as the kinetic molecular theory of heat. Then in section 9, I discuss whether the same strategy works for teleosemantics. I argue that while this can be done by making conceptual commitments to success semantics and the etiological conception of biological function, the trouble here is that, unlike the case of the H_2O theory, the major intuition about the Swampman scenario goes directly against these theoretical analyses of representation and biological function. I thus conclude that the Swampman scenario raises a question about the adequacy of the two conceptual foundations of teleosemantics. The difference between teleosemantics and the other reductive theories lies in their conceptual foundations.

The present analysis is of great relevance to the main problem of this dissertation. For it concerns what exactly is required to explain an apparent possibility away. In each case of a necessary statement, according to the present analysis, part of what excludes the apparent counterpossibilities away as unreal is an analytic general principle implying that the statement is necessary if true at all. The analysis present analysis also affords a straightforward account of how knowledge of necessity is possible. We sometimes know by conceptual analysis a general principle implying that a certain statement is necessary if it is true at all. In such a case, knowledge that the statement is necessary can be arrived at by simple inference from knowledge that the statement is in fact true. The present analysis of the Swampman objection thus throws an insight as to how necessary statements and knowledge thereof are possible.

2. The two pillars of teleosemantics

It is part of the nature of beliefs that they are representational; that is, they have truth conditions necessarily. Teleosemantics is a family of theories that try to explain the representational nature in terms of biological functions. The basic idea is that beliefs are states whose function is to guide actions in appropriate ways (Papineau, 2016, p.97). Suppose that you believe that mushrooms with a ring on the stem are poisonous. This belief may prompt you to act in ways that would be beneficial when such mushrooms are in fact poisonous; you would, for example, avoid consuming them. In this way, teleosemantics understands beliefs as functional states that elicit responses that are appropriate just in case they are true.

From this characterization, there already arise two questions. First, why is this an account of the *representational nature* of beliefs? In other words, what is the connection between a state's having the function of guiding actions and its having a truth condition? The standard teleosemantic answer is given by *success semantics*: the truth condition of a belief is determined by the condition under which responsive behaviors that it prompts are appropriate. So, going back to the previous example, the belief has the truth condition that mushrooms with a ring on the stem are poisonous precisely because of its function of prompting responsive behaviors that would be appropriate just in case the condition obtains. In this sense, according to teleosemantics, the representational nature of beliefs is grounded in their action-guiding function.

The second question concerns what it is for a state of an organism to have a certain function. It can be answered in more than one way, each giving a version of teleosemantics. For the purpose of this chapter, I shall be concerned only with what I take to be the most common answer, namely the *etiological conception*: a biological feature (of an organism) has the function of producing an effect E if it is chosen by a process of selection in virtue of producing E.[1] Consider, for example, long necks of giraffes. From my research, there seems to be no consensus among evolutionary biologists as to why long necks were selected. It was originally proposed by Lamarck and endorsed later by Darwin that long necks have been selected because they allowed giraffes to reach leaves on higher branches of tree. If this is right, then according to the etiological conception, the biological function

[1] See, for example, Cummins (1975) and Nanay (2014) for alternative accounts of biological function.

of long necks is to obtain food beyond the reach of other animals. Another hypothesis is that long necks have evolved as their legs are getting longer, so as to ensure that giraffes can drink water. Under this hypothesis, their function would be to ensure the access to water. So, according to the etiological conception, the function of a biological state is determined by the selectional history behind it. Here let me note that selectional history need not be confined to that of natural selection; it could also be widened to include selectional histories of non-genetically inherited traits (say, via some learning processes).

With these two pillars—success semantics and the etiological conception of biological function—we can state the core tenet of teleosemantics more precisely. Call a biological state S of an organism a *selectional state covarying with P* just in case

(T1) it reliably prompts a class C of responsive behaviors;

(T2) the responses C succeed in achieving a biological end just in case P obtains; and

(T3) the state S has been historically selected for because of (T1) and (T2).

Then, according to teleosemantics, a belief that P is a selectional state covarying with P.[2]

Notice that the teleosemantic thesis is indeed a *consequence* of the two pillars of teleosemantics. Success semantics states that for an organism to have a belief that P is to be in a state whose function is to elicit responses that are appropriate just in case P holds. For a state to have such a function, according to the etiological conception, it must be selected for producing such responses. So it follows that beliefs are selectional states.

3. The Swampman problem

It should be clear that teleosemantics understands selectional history as part of the grounds of the representational nature of beliefs. To see this more clearly, consider again the belief that mushrooms with a ring on the stem are poisonous. According to teleosemantics, the belief has the content because it is chosen for prompting the responsive behaviors that are appropriate just in case such mushrooms are poisonous. However, it also seems possible that the same internal state is selected because those responsive behaviors are appropriate just in case such mushrooms

[2]Here I am lazily using, and will continue to use, 'P' ambiguously. The first occurrence of 'P' should be replaced by a sentence and the second should be considered as a name of the proposition expressed by the sentence.

kill competitors. Then it would follow from (T1)-(T3) that in this alternative scenario, the truth condition of the internal state would be that mushrooms with a ring on the stem kill competitors. So, according to teleosemantics, the content of the belief is determined in part by the selectional history.[3]

So it seems natural to take teleosemantics as implying that no belief is possible without selectional history; in other words, there cannot be any ahistorical creatures with beliefs. This consideration invites one of the most influential objections to teleosemantics, which is inspired by Donald Davidson's (1987, p.443) thought-experiment:

> Suppose lightning strikes a dead tree in a swamp. ... [E]ntirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference.

It is assumed that the Swampman has no history whatsoever. So, none of its internal states satisfies condition (T3). According to teleosemantics, therefore, it has no beliefs.[4] However, many philosophers find this consequence to be counterintuitive. For we in the ignorance of its extraordinary origin would naturally attribute beliefs to it and interact with it on that basis, just as we would do for Davidson himself. Hence something must have gone wrong with teleosemantics, or so many have argued.

A standard teleosemantic response is that we should override the pre-theoretic intuition in favor of the explanatory and unifying power of the theory. This line is taken by Millikan (1996), Neander (1996) and Papineau (1996), just to name a few. However, it seems that the opponents

[3]I believe that my remark is in agreement with Mark Greenberg (2005, p.314): "On biological teleological theories, in contrast to non-normative historical or biological theories, certain historical facts are alleged to be determinants of content because they are a source of function. So biological teleological theories have a distinctive explanation of why historical facts are basic determinants of content."

[4]It is worth noting that in his original paper, Davidson himself endorses the conclusion, suggesting that a belief can be formed only "in a context that would give it the right" truth condition or any truth condition at all (ibid, p.414). To a certain extent, I am sympathetic to Davidson's view. For the purpose of this chapter, my personal view is less important than the analysis of the problem; so I will not be concerned so much with this alternative view.

take the lack of a better answer as an indication that the Swampman scenario poses a daunting problem for teleosemantics.

More recently, Papineau (2001; 2006; 2016) attempts to resolve this frustrating situation by providing a more nuanced response.[5] He claims that the Swampman scenario is not even relevant to teleosemantics. More specifically, he makes three interrelated claims:

(1) Teleosemantics is not intended as an analysis of the concept of belief, but as a scientific reduction of the representational nature of belief.

(2) As such, teleosemantics can be made consistent with the intuition about the Swampman scenario.

(3) Regardless of (2), moreover, the Swampman scenario poses no problem for teleosemantics.

As opposed to these claims, I shall argue:

- There is a tension between (1) and (2); more specifically, as a reductive account of the representational nature of beliefs, teleosemantics is not consistent with the major intuition about the Swampman scenario.
- The scenario raises a conceptual issue for teleosemantics; more specifically, it raises questions concerning the adequacy of the conceptual foundations of teleosemantics, namely, success semantics and the etiological conception of biological function.

The upshot will then be that teleosemanticists should tackle the Swampman objection head-on if they want to understand it as an explanatory account of the representational nature of beliefs.

4. Teleosemantics as a scientific reduction

Papineau (2016, pp.114-15) admits that the Swampman objection would pose a daunting problem if it is understood as an analysis of the ordinary concept of belief. For we certainly do not seem to mean a selectional state covarying with *P* by the term 'belief that *P*'. However, he insists

[5] As we shall see below (§6), Papineau's argument was partly anticipated in Neander (1996).

that the core teleosemantic thesis should not be understood as a piece of conceptual analysis; it is rather intended as a *scientific reduction*.[6]

The primary example of a scientific reduction is that of water to H_2O. In this case, according to Papineau, water is scientifically reduced to H_2O in the sense that:

> We start with the folk role associated with 'water'—odourless, tasteless, colourless, potable. We take 'water' to refer to whichever theoretically significant liquid fills this role. And then science tells us that H_2O is in fact the liquid which does this. (Papineau 2001, p.279)

The basic idea seems clear. We pick out a liquid by the superficial—sensible and functional—properties that are typically associated with the term 'water'. Then we empirically discover that it is a chemical substance with the chemical composition of H_2O.

It might appear that Papineau's notion of scientific reduction coincides with Kripke's (1980) reading of theoretical identity statements ('water is H_2O'). But there is a crucial difference. It is part of Kripke's reading that 'water' is a rigid designator for a kind of thing. This semantic assumption implies that the identity of water to H_2O, if true, must be necessarily true; in other words, water cannot be reduced to H_2O unless it is necessary that water has the chemical composition of H_2O. Papineau, on the other hand, makes no such semantic assumption. He leaves it open that the term 'water', as we ordinarily use it, might refer to different chemical substances in different possible worlds (Papineau, 2001, pp.285-286).

So, he does not require that the scientific reduction of water to H_2O should establish that water is necessarily H_2O. For the term 'water' might refer to a different chemical substance in another possible world. Nor does the reduction require H_2O to be the realizer of the superficial properties in every possible world, which would indeed be an unduly strong condition. For the scientific reduction of water to H_2O, Papineau thus thinks, it is only required that the very liquid that has the superficial properties in the actual world has the chemical composition of H_2O. This, of course, is not to deny that *if* 'water' is a rigid designator, then it would indeed follow from the scientific reduction of water to H_2O that water is necessarily H_2O. The point is that the semantic assumption

[6]This much seems to be what most defenders of teleosemantics would agree on, despite some possible difference in details concerning the notion of scientific reduction. See, for example, Millikan (1996, p.15 and p.110), and Neander (1996, pp.124-125).

is not forced, and more importantly, should not be considered as part of the theory which concerns what the very liquid is (ibid, p.286).

To see this point more clearly, it might be helpful to straighten out some of the semantic subtleties. Let 'L' be a rigid designator for the very liquid that we typically call 'water'. So, it is true in the actual world that for all x,

$$(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } L), \tag{α}$$

Since the terms 'L' and 'H_2O' are both rigid designators, moreover, it follows from the empirical discovery that, for all x,

$$\Box(x \text{ is } L \quad \Leftrightarrow \quad x \text{ is } H_2O). \tag{β}$$

It thus follows from (α) and (β) that it is true in the actual world that for all x,

$$(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } H_2O). \tag{γ}$$

Now if 'water' is a rigid designator, on the one hand, then (α) can be strengthened to

$$\Box(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } L), \tag{α'}$$

and this together with (β) would imply that for all x,

$$\Box(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } H_2O). \tag{γ'}$$

If, on the other hand, 'water' is a disguised description designating different chemical substances in different possible worlds, then we would have:

$$\Diamond\neg(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } L), \tag{α^*}$$

which together with (β) implies that

$$\Diamond\neg(x \text{ is water} \quad \Leftrightarrow \quad x \text{ is } H_2O); \tag{γ^*}$$

that is, water is possibly not H_2O. So, the modal status of (γ) depends on the semantics of the term 'water'. Though it is of interest in its own right whether 'water' is a rigid designator or not, Papineau thinks, the issue is orthogonal to the scientific account of what water actually is.

Now assume that we have a strong intuition that water could have a different chemical composition, say XYZ. It should be clear that this intuition by itself is consistent with the scientific reduction of water to H_2O in Papineau's sense. For it is only committed to (α), (β) and hence to (γ). In other words, it only claims that H_2O is what *actually* has the superficial properties associated with the term 'water'. It is simply not part of the theory that water is necessarily H_2O. So it is no objection to the theory that water would be XYZ in a different possible world.

Papineau thinks that essentially the same goes for teleosemantics. It starts by identifying a class of superficial—sensible or functional—properties associated with the term 'belief that P'. Again, there is no semantic assumption as to whether the term 'belief that P' is a rigid designator. Then it argues from empirical studies that the properties are realized by a selectional state covarying with P. This is not to claim that the superficial properties are realized by the same state in all possible worlds. The core doctrine of teleosemantics is simply that it is the selectional states that realize the superficial properties in the actual world (ibid, p.285). So, letting the term 'R' be a rigid designator for the *actual* realizer of the superficial properties associated with the term 'water', teleosemantics claims that for any x,

$$(x \text{ has the belief that } P \quad \Leftrightarrow \quad x \text{ contains } R); \tag{a}$$

and

$$\Box(x \text{ contains } R \quad \Leftrightarrow \quad x \text{ contains the selectional state covarying with } P); \tag{b}$$

hence to

$$(x \text{ believes that } P \quad \Leftrightarrow \quad x \text{ contains the selectional state covarying with } P); \tag{c}$$

but not to:

$$\Box(x \text{ believes that } P \quad \Leftrightarrow \quad x \text{ contains the selectional state covarying with } P). \tag{c'}$$

To be sure, *if* the term 'belief that P' is indeed a rigid designator, (a) can be strengthened to

$$\Box(x \text{ has the belief that } P \quad \Leftrightarrow \quad x \text{ contains } R), \tag{a'}$$

which together with (b) would imply (c'). But the assumption is not part of the theory.

From this understanding of teleosemantics, Papineau argues that our intuition about the Swampman scenario is consistent with teleosemantics. Assume that we have a strong intuition that the Swampman has beliefs. This is an intuition about a merely possible world. For, presumably, the Swampman does not exist.[7] So the intuition tells us what beliefs *could* be. However, teleosemantics as a scientific reduction tells us only about what the actual realizers of the superficial properties of beliefs are. It is not part of the theory that beliefs are necessarily selectional states; nor does it claim that the superficial properties of beliefs are necessarily realized by selectional states. So, it is consistent with teleosemantics that there could be ahistorical creatures with beliefs. Papineau thus concludes that teleosemanticists can coherently embrace the intuition about the Swampman scenario.

5. Explanation and modality

In their response paper titled "A Pyrrhic Victory for Teleonomy" (2002), David Braddon-Mitchell and Frank Jackson claim that the intuition cannot be so easily embraced. Suppose that teleosemanticists admit that the Swampman has beliefs. Then they would also have to maintain that the term 'belief that P' is not a rigid designator. What, according to this reconciliatory teleosemantics, would the truth conditions of attributions of belief be? It would seem, Braddon-Mitchell and Jackson argue, that the question of whether an agent has a certain belief is reduced to that of whether s/he has the superficial properties that are typically associated with the belief; in their own words, "selectional states settle what subjects believe ... only when they deliver the same answer as folk roles. But that is another way of saying that they don't settle what subjects believe ... folk roles do" (ibid, p.376). So the reconciliation comes at a "pyrrhic" cost of collapsing teleosemantics

[7] Some have argued that there actually are Swampman-like creatures. See, for example, Peters (2014).

to a version of analytical functionalism, according to which: for all x,

$\Box$(x has the belief that $P \Leftrightarrow x$ has the superficial properties associated with the term 'belief that P').[8]

Notice that implicit in this argument is the assumption that the only reasonable ground for attributing beliefs to the Swampman is that it exhibits the relevant superficial properties. Papineau in fact seems to make this assumption in his discussion (ibid, p.375 and Papineau, 2001, pp.286-287). So the objection seems to be valid.

However, it is important for the purpose of the current discussion that the assumption is not strictly forced. Recall that the Swampman is different from Davidson only in two respects, namely that it has no history whatsoever and that it is made of entirely different set of molecules; otherwise, they are just alike. So, it seems open to teleosemanticists to attribute beliefs to the Swampman on the basis not solely of the superficial properties, but also of some other properties that it shares with Davidson. Just because teleosemanticists admit that the Swampman has beliefs, therefore, it does not immediately follow that teleosemantics collapses to a version of analytical functionalism.

Now, this seems to raise an important issue about Papineau's understanding of scientific reduction. Recall that Papineau takes teleosemantics as the thesis that beliefs are selectional states *in the actual world*; so it does not by itself carry any commitments to *modal* claims about the truth conditions of belief-ascriptions. As John Post (2006, p.19) puts it, what happens in the other possible worlds is a "don't care." So teleosemanticists can say, "Beliefs are selectional states, but they could be something else." Similarly they can also say, "Beliefs are what realize the associated superficial properties, but not necessarily." And this can clearly be generalized. Then there arises a question, namely, in what sense teleosemantics offers an explanatory account of the representational nature of beliefs?

Let me elaborate this point a bit. Let Φ and Ψ be any properties. If Φ is part of the nature of Ψ then it is necessary that every Ψ is Φ. Since it is part of the nature of beliefs that they are representational, it follows that beliefs are necessarily representational. Now suppose that it is the property of being a selectional state that constitutes the representational nature of beliefs. Then

[8]This is different from the claim that (a) is necessary, because the superficial properties might not be realized by R in another possible world.

it seems reasonable to think that beliefs are necessarily selectional states. According to Papineau, however, the core tenet of teleosemantics is simply that beliefs are selectional states but not necessarily. But could it not be a coincidence that beliefs and selectional states are coextensive in the actual world? Why do we have to think that being a selectional state explains the representational nature of beliefs?

Let's go back to the H_2O theory of water. If we understand it as a scientific reduction in Papineau's sense, then the theory only states that the superficial properties that are typically associated with the term 'water' are realized by the stuff that has the chemical composition of H_2O. So, the H_2O theorist can consistently say, "Water is H_2O, but it could have been something else." So one may ask in what sense this theory is an account of the nature of water. It is perhaps a surprising coincidence that the two properties are coextensive in the actual world. Perhaps, a better example is the properties of having a kidney and having a heart. These two properties are considered coextensive in the actual world. But this does *not by itself* mean that one property is explained by the other. For this conclusion, some further story needs to be told.

In his critical defense of Papineau, Peter Schulte (2020, pp.2278-2280) proposes that for a scientific reduction of a property W (say, of being water) to another property H (say, of having the chemical composition of H_2O) to be genuinely explanatory, it should be the case that for any object x,

$$\Box(H(x) \supset W(x)). \tag{N1}$$

The idea is that an explanatory theory should tell us how it follows from an object's having the property H that it has the property W. This can be done, according to Schulte, by showing how the superficial properties of water follow from its chemical composition together with physicochemical laws.

I am sympathetic to the proposal, though I do not think that it is so important to explain the superficial properties of water by which it is typically recognized. Some properties of water may be taken as more basic and important than its superficial properties, and I am inclined to think that

if the H_2O theory is genuinely explanatory of the nature of water, it should tell us how those basic properties follow from the chemical composition.[9]

Now Schulte seems to think that this condition is sufficient for a scientific reduction of W to H to be genuinely explanatory. So, for example, he claims that the H_2O theory of water might admit that there can be many alternative realizers of water (ibid, p.2288). In other words, he denies that the reduction of W to H is genuinely explanatory only if, for any x,

$$\Box(W(x) \supset H(x)). \tag{N2}$$

This is because, according to Schulte, merely possible alternative realizers are irrelevant to the scientific and philosophical understanding of what water really is. Schulte thus agrees with Papineau that what is relevant to the scientific reduction is only the actual realizer.

As opposed to this view, I think that (N2) must also be established if H explains part of the *nature* of W. This is because W is not possible without its nature being realized, and because if part of its nature is explained by certain conditions, then the conditions must obtain for the nature to be realized. So, we should require that H constitute a necessary condition of the possibility of W. If H explains part of the nature of W, in other words, then we should have: for any x,

$$\neg\Diamond(W(x) \wedge \neg H(x)),$$

which is equivalent to (N2).

Indeed, it is not quite clear how a theory may claim to give a reductive account of the nature of W in terms of H without endorsing (N2). For the theory would then say, in effect, that it is part of the nature of water that it has the chemical composition of H_2O but it is possible that water has a different chemical composition. This seems to me to be an untenable position. For whatever constitutes the nature of water must necessarily belong to it. How is it possible that water has a different chemical composition if its chemical composition constitutes part of its nature? Here one

[9]My remark here is not intended as a criticism of Schulte's view. I only wish to point out a possible disagreement. In fact, some of the properties of water that Schulte takes to be superficial, such as high viscosity and higher electrical conductivity than most of other non-metallic liquids, can be taken as basic properties of water, but they do not seem to me to be properties by which we typically recognize water, at least outside science laboratories. Also, I do not mean to deny that some of the superficial properties can be basic.

might reply that the theory is intended to give an account of what the chemical nature of the *actual realizer L* is; so, it is no objection to the theory that it does not offer an account of the nature of *water*. Recall, however, that the crucial question concerns what water is, to put it another way, what the truth condition of the statement "*x* is water" is. So, if the theory tells us only what the chemical nature of *L* is while remaining silent to that of water, then we should say that the theory is not really about water but rather about *L*. But then there does not seem to be any substantive sense in which the theory offers a reductive account of the nature of water.

Let me illustrate my point with another example. Consider the scientific theory that sound is propagation of vibration (i.e., oscillation in a medium). Let's suppose that the theory is intended as an account of the physical nature of sound, just as teleosemantics is of the representational nature of belief and the H_2O theory is of the chemical nature of water. Thus understood, it should imply that there would be no sound at all in complete vacuum. For there is no medium to be oscillated. In other words, sound is possible only under the condition that there is a medium to be oscillated. In this sense, we may say that the existence of a medium constitutes a necessary condition for the possibility of sound.

Now suppose that someone comes up with a clever thought-experiment that purports to show that sound is possible in complete vacuum. Presumably, there is no complete vacuum in the actual world. So this thought-experiment seems to be based on a merely possible scenario. Can one really defend the theory of sound simply by saying, "Yes, there could be sound without any medium, though actually it requires a medium"? According to Papineau and Schulte, it should be enough to defend the theory, because a merely possible scenario is irrelevant to what sound really is. In my opinion, this dismissal is not only theoretically dissatisfying but also misguided. For, to repeat, whatever constitutes the nature of sound must belong to it necessarily. So, if propagation of vibration does constitute the physical nature of sound, then sound should be impossible without there being a medium to be oscillated. Hence there must be something wrong about the thought-experiment. Thus, the defenders of the theory should in principle be able to point it out, though of course they might not currently be in a position to give a satisfactory answer.

I believe that essentially the same goes for teleosemantics. Recall that teleosemantics is intended as an account of the representational nature of belief. So we have to understand conditions (T1)-(T3) above as constituting conditions under which beliefs are possible. So the theory implies that beliefs are necessarily selectional states; in symbolism, for any x,

$$\Box(B(x) \supset S(x)),$$

where B and S stand for the property of having a belief and the property of having a selectional state, respectively. But this is what ought to be given up, if one wants to make teleosemantics consistent with the intuition about the Swampman scenario. Therefore, in so far as we understand teleosemantics as a reductive account of the the representational nature of beliefs, it cannot be made consistent with the intuition about the Swampman scenario.[10]

Let me address a couple of possible objections here. I have urged that teleosemantics as an account of the representational nature of beliefs is committed to the modal claim that beliefs are necessarily selectional states. This is because it is part of the nature of beliefs that they are representational and because whatever constitutes the nature of beliefs should belong to them necessarily. It might be objected that it is not the aim of teleosemantics to find out what *constitutes* the representational nature of beliefs; rather, it aims at offering an explanation in the sense that selectional histories provide the standards from which the representational contents of the subject's beliefs can be derived.[11] According to this view, teleosemantics is not committed to the claim that the property of having a selectional history is constitutive of the representational nature of beliefs. This

[10]In their paper "Essential Properties Are Super-Explanatory: Taming Metaphysical Modality" (2020), Marion Godman, David Papineau and Antonella Mallozzi argue that kinds involve what they call *superexplanatory properties*. For example, the property of having the chemical composition of H_2O is superexplanatory of the kind water in the sense that the former explains most of the common properties of water. They further argue that the superexplanatory properties of a kind constitutes the necessary and sufficient condition in the strict sense; in other words, for any kind K, and any property (or, a cluster of properties) E, if E is superexplantory with respect to K, then

$$\Box(K(x) \equiv E(x)).$$

Notice that, according to this account, both (N1) and (N2) ought to be true if we interpret H as the property of having the chemical composition of H_2O and W as the property of being water. Now, they do not explicitly discuss teleosemantics in the paper. However it should be clear that *if* the teleosemantic thesis is to be understood as that the property of being a selectional state is superexplanatory of belief (or, more broadly, representation), then it immediately follows that beliefs are necessarily selectional states.

[11]I thank the anonymous referee for pressing me to consider this objection.

again makes unclear the sense in which teleosemantics offers a reductive explanation of beliefs. It might well be true that we can correctly ascribe beliefs to the subject on the score of its selectional histories, but that alone does not mean that beliefs are reduced to selectional states. For instance, we may successfully predict and explain what the subject believes on the basis of his/her outward behaviors. By no means does it mean, however, that beliefs are reduced to observable behaviors in any substantive sense. It would be more reasonable to think that the outward behaviors give us only a "folk" *model* of the subject's beliefs. Similarly, the weakened version of teleosemantics should be understood as providing a naturalistic *model* of belief.[12] Without denying that it might be an interesting project in its own right, I presume that teleosemantics-as-a-model would not be appealing to many teleosemanticists who are devoted to the project of naturalizing the mental.[13] So, I shall not be concerned with this view below, and focus on what problem the Swampman scenario poses for teleosemantics that aims at a reductive explanation of the representational nature of beliefs.

The other objection is that my argument implicitly assumes that the property of having a belief is not a disjunctive property. In allowing the possibility of alternative realizers, the objection goes, Papineau and Schulte in effect suggest to understand the property of having a belief as a disjunctive property. According to this disjunctivist view, we would have: for any x,

$$\Box(B(x) \equiv (S(x) \vee R_1(x) \vee ... \vee R_n(x))), \tag{Disj}$$

where R_1, ..., R_n are alternative realizers of the property of having a belief. This in fact validates only

$$\Box(S(x) \supset B(x)),$$

which is an instance of (N1), but not its converse

$$\Box(B(x) \supset S(x)),$$

[12]This route was taken by Godfrey-Smith (2006).

[13]For example, it is clear that Schulte and Papineau are interested in giving a constitutive explanation of the representational nature of beliefs. For textual evidence, see Schulte (2020, fn.20) and Papineau (2016, pp.114-116). See also Millikan (1996) and Neander (1996), who also understand teleosemantics as a theory of the real nature of beliefs.

which is an instance of (N2) and which is crucial for my argument.

As far as I know, they do not explicitly take this disjunctivist route to make teleosemantics consistent with the intuition about the Swampman scenario. So it is difficult to address this objection in a determinate way. However, there are some general considerations worth noting.

First, the disjunctivist should first show that the property of having a belief is disjunctive, if they want to give an explanatory account of the representational nature of belief. It certainly does not suffice to say, "We do not know whether it is disjunctive or not, but let's assume that it is."

Second, assuming that there are good reasons to adopt the disjunctivist view, it should be clear that a theory that yields (Disj) is superior in its explanatory power to teleosemantics. For the former tells us all possible realizers of belief, while the latter tells us only one. In this respect, taking the disjunctivist route does not refute my requirement that an explanatory theory of belief should give us the condition under which it is possible.

Third, teleosemanticists should be able to explain why it is only the property of having a selectional state, but not R_1, ..., R_n, that realizes the property of having a belief in the actual world, if they want to retain the explanatory power of teleosemantics in the actual world.

Fourth, even if it can somehow be shown that selectional states are the only realizers of belief in the actual world, teleosemantics cannot be made consistent with the intuition about the Swampman scenario. To see this, suppose that selectional states are the only realizers of belief in the actual world. So Davidson himself should have beliefs solely in virtue of having selectional states and lack all the properties R_1, ..., R_n. Since the Swampman is assumed to be similar with Davidson in every respect, except that he has no history, it should also lack the properties R_1, ..., R_n. Then it follows from (Disj) that the Swampman has no beliefs. So, taking the disjunctivist route does not make teleosemantics compatible with the intuition about the Swampman scenario.[14]

To this argument, finally, teleosemanticists may reply that they can still remain non-committal to the modal claim that beliefs are necessarily selectional states.[15] For the argument only shows that the Swampman has none of the realizers, but not that there are no alternative realizers at

[14] A version of this argument can be found in Schulte (2020, fn. 36).

[15] This is how Schulte himself replies to the argument in the same footnote. I thank the anonymous referee for pressing me to consider this reply.

all. This is certainly correct. But it is important to note that disjunctivist teleosemantics is still committed to:

$$\Box(B(x) \supset (S(x) \vee R_1(x) \vee ... \vee R_n(x))).$$

That is, the disjunctive property $(S(x) \vee R_1(x) \vee ... \vee R_n(x))$ constitutes the necessary condition of the possibility of belief. Then it should be clear that the Swampman scenario provides a putative counterexample to disjunctivist teleosemantics, just as it does to the non-disjunctivist version. For the Swampman appears to have beliefs while the theory says that it does not, whether disjunctivist or non-disjunctivist. So it is reasonable to expect, and I will later argue, that if the scenario poses any problem to non-disjunctivist teleosemantics, then it will apply *mutatis mutandis* to the disjunctivist one as well.

6. So why is it a problem?

My argument so far presents a dilemma for the defenders of teleosemantics: either they have to give up the status of teleosemantics as a reductive account of the representational nature of beliefs, or they just have to bite the bullet. In other words, teleosemantics cannot be made consistent with the intuition about the Swampman scenario if it is understood as an explanatory account.

One can still ask why exactly this is a problem for teleosemantics. And, in fact, Papineau has another line of argument to this effect.

> Well, consider this parallel argument, raised against the scientific claim that water is H_2O. "The 'water = H_2O' equation can't be right, because it doesn't tally with the way we care about water. Imagine that you were in some alien desert, dying for some water, and came to an oasis, with a delicious pool of colourless, odourless, tasteless, entirely potable liquid. As it happens, this liquid would be XYZ, rather than H_2O. But don't tell me that you would turn it down on this account as an answer to your prayer for water." ... This doesn't even look like the start of an argument against the theory that water is H_2O. (Papineau, 2001, p.283; see also his 2016, p.115.)

Here we find an argument by analogy. The Swampman scenario stands to teleosemantics just as the XYZ scenario stands to the H_2O theory of water; so, if the XYZ scenario poses no serious problem for the H_2O theory, then why think the Swampman scenario does for teleosemantics?

A similar line of argument has also been put forward by Neander (1996, p.120). She imagines scenario where a molecule-for-molecule copy of an actual cow is created by complete coincident. She claims that it seems silly to argue against contemporary evolutionary biological theories of cows that it fails to account for the Swampcow. Then why think the Swampman objection is any good against teleosemantics?

7. Kripke's requirement on the necessary aposteriori

Papineau and Neander seem to think that it is obvious that the XYZ scenario raises no problem for the H_2O theory. As opposed to their view, I believe that it raises a *conceptual* issue; and the issue seems to be not only of philosophical, but also of scientific, significance.

Let's call a proposition *aposteriori* if its truth cannot be established apriori. All the scientific claims that we have considered so far seem to fall under this category. Consider, for example, the H_2O theory of water. It does not seem possible to know solely by apriori ratiocination that water has the chemical composition of H_2O.

For an aposteriori proposition P, it always makes sense to say, "It is the case that P, but it could have been otherwise." Since we cannot apriori exclude the possibility of P's being false, it appears to be possible that it is false.

Let P be a necessary aposteriori proposition. Since P is aposteriori, it appears to be possible that P is false. This apparent possibility should not be real, however, because P is necessary. This suggests that there must be something wrong about the apparent possibility. So I propose the following requirement:

> *Kripke's Requirement*: To establish the truth of a necessary aposteriori proposition P, the apparent possibility of P's being false must be explained away.

The scientific claims under consideration, if true, should fall under the category of necessary aposteriori propositions. So they are subject to Kripke's Requirement.

Though it is rarely stated explicitly, Kripke's Requirement is known to many readers' of his *Naming and Necessity* (1980)[16]—hence the name—because it is implicit in his criticism of the mind-body identity theory as a general point about the necessary aposteriori:

> [T]he correspondence between a brain state [the stimulation of c-fibers] and a mental state [pain] seems to have a certain obvious element of contingency. We have seen that identity is not a relation which can hold contingently between objects. Therefore, if the identity thesis were correct, the element of contingency would not lie in the relation between the mental and physical states. It cannot lie, as in the case of heat and molecular motion, in the relation between the phenomenon (= heat = molecular motion) and the way it is felt or appears (sensation of [heat]), since in the case of mental phenomena there is no 'appearance' beyond the mental phenomenon itself. (p.154)

In short, there is an important disanalogy between the identity of heat to molecular motion and the identity of pain to c-fiber stimulation. Putting details aside for the moment, the disanalogy is that the apparent possibility of heat without molecular motion can be explained away, while that of pain without c-fiber stimulation cannot.

Why is this a problem for those who hold the identity of pain to c-fiber stimulation? It would not be a problem if one can just dismiss the apparent possibility as silly. As opposed to Papineau and Neander, however, Kripke thinks that the apparent possibility is something to be dealt with even in the case of the identity of heat to molecular motion. So implicit in Kripke's argument is the general requirement that the apparent possibility of P's being false must be explained away if P is necessary aposteriori.

What justifies Kripke's Requirement? This question seems to have received relatively little attention in the literature. I am inclined to think that it can be taken as a logical principle. For we have the following duality:

$$\Box P \equiv \neg\Diamond\neg P.$$

[16]See, for example, Loar (1990, pp.84-85) and Hill (1997, p.62).

So, establishing the necessity of P logically amounts to establishing the impossibility of P's being false; and in fact this is regardless of whether P is apriori or aposteriori. In the light of the duality, which I think is fundamental to our understanding of necessity and possibility, Kripke's principle should seem highly plausible.[17]

One might think that Kripke's Requirement is too demanding to be adopted as a general principle. My view is that it is preferable as a methodological principle also. Thinking about possibility and necessity is often a slippery business. So it is reasonable to require a policy of caution. Moreover, it does not seem to be a sound methodological principle that we are allowed to dismiss a putative counterexample to a theory simply on the grounds that it is unlikely to be found in the actual world; it would indeed be too restrictive to allow only actual cases in theoretical discussions.

8. How to explain apparent possibilities away

How then can one explain away, say, the apparent possibility that water does not have the chemical composition of H_2O? First, Kripke argues that 'water' rigidly designates the stuff L that we identify via its usual superficial properties (its color, feel, odor, etc.). Then he asks: given the empirical discovery that L is a chemical substance with the chemical composition of H_2O, how would we describe the XYZ scenario? Would we describe it as a situation where water does not have the chemical composition of H_2O?

We would not, argues Kripke, because 'water' rigidly refers to L and because it is part of what it is to be L in the strict sense that it has the chemical composition of H_2O—or, more generally, the chemical composition of a chemical substance is its essential property. So it follows that the situation is not where water does not have the chemical composition of H_2O. The XYZ scenario describes a situation that is *epistemically equivalent* to the actual one in the sense that we have exactly the same superficial evidence about another chemical substance as we actually have about water. So, what is false in the XYZ scenario is not the proposition that water has the chemical

[17] In the light of this argument, Kripke's Requirement can plausibly be generalized to the necessary propositions in general, whether apriori or aposteriori, though it might seem superfluous in the case of necessary apriori propositions.

composition of H_2O but its epistemic equivalent that what possesses the superficial properties of water has the chemical composition of H_2O.[18]

In the case of water, then, Kripke's strategy can be summarized as follows:

(W1) 'Water' rigidly designates the stuff L that we identify in usual ways.

(W2) L is a chemical substance that has the chemical composition of H_2O.

(W3) Every chemical substance has its chemical composition as a necessary property.

(WC) Water necessarily has the chemical composition of H_2O.

Let me note a few things about the argument. Note first that (W3) has the following logical form:

> For any chemical substance s, if s has the chemical composition C, then necessarily s has the chemical composition C.

Then it should be clear that (W1) and (W2) together imply that water is a chemical substance with the chemical composition of H_2O, which together with (W3) logically implies (WC).

Second, one might ask, why accept (W3)? My view is that it has a *definitional character* in the sense that each chemical substance is defined by its chemical composition together with some of its basic microphysical properties. So, if anything is a chemical substance, its chemical composition could not have been different from what it actually is. This is not because of what we mean by 'water', but because of our basic understanding of chemical substance and of chemical composition.[19]

Third, one might worry that the response reduces the notion of metaphysical necessity to that of analyticity. For necessity seems to come into the scene with (W3) whose truth is deemed to

[18]See Kripke (1980, pp.124-25) for his discussion about gold. I believe that there is no essential difference between the two cases of water and gold. See also pp.140-142 and pp.150-51 for his discussion of epistemically equivalent situations. Strictly speaking, Kripke is concerned with the converse problem, namely whether a chemical substance with the chemical composition of H_2O can possibly not be water. But it can be given a similar treatment.

[19]Kripke (1971, pp.16-17) claims that for some propositions P, we can know "by a priori philosophical analysis" the conditional of the form "if P then necessarily P," and gives as an example the proposition that the table is not made of ice. However, he does not further elaborate what he means by "a priori philosophical analysis." The current discussion suggests that in the case of water and H_2O, the subject of apriori philosophical analysis is scientific concepts such as chemical substance and chemical composition; roughly, the truth of (W3) can be ascertained solely on the basis of the observation that chemical substances are in part defined by their chemical compositions. In this connection, see footnote 22 below. I believe that a similar account can be given to other well-known examples, such as the necessity of identity and of origin (Kripke, 1980, p.109 and pp.113-114), although they do not necessarily involve scientific concepts. But a proper investigation into this issue must wait for another day.

be definitional. Admittedly, it is part of my claim that there is an element of analyticity in (W3) (if one prefers to speak this way). However, there is no substantive sense in which metaphysical necessity is *reduced* to analyticity. (WC) is a metaphysical necessity (if true at all), but surely (W3) by itself is not sufficient to establish (WC); (W1)-(W3) are all crucial for this conclusion. Under the current analysis, then, some but not all metaphysical necessities are analytic. This I do not find objectionable.

Finally, we can see from the argument why it is legitimate to appeal to the notion of epistemic equivalence in this case. This is because (W3) tells us that two chemical substances are distinct if they have different chemical compositions. So, given that water is a chemical substance with the chemical composition of H_2O, the XYZ scenario—where it is discovered that a certain substance superficially similar to water has a different chemical composition—should not be described as a situation where water has the chemical composition of XYZ, but as an epistemically equivalent situation where another substance superficially similar to water has the chemical composition of XYZ. For the discovery in the scenario is not about water.

To see this point a bit more clearly, let S_1 and S_2 be chemical substances. Suppose that S_1 has C_1 as its chemical composition in a possible world w_1, and that S_2 has C_2 in another possible world w_2. From this assumption alone, it does not immediately follow that S_1 and S_2 are distinct chemical substances. For they could very well be the same chemical substance whose chemical composition varies across different possible worlds. This much is purely a logical point. Pick any x and y, and suppose that x has a property P in w_1 but y does not in w_2 (where $w_1 \neq w_2$). From this it is *not* immediate that $x \neq y$, because P may be a contingent property of $x\,(=y)$. For example, take both x and y to be Socrates and Φ to be the property of being a philosopher. In order to conclude $x \neq y$, therefore, we need to assume that Φ is an essential property of $x\,(=y)$. To establish $S_1 \neq S_2$, similarly, it ought to be assumed that the chemical composition of a chemical substance is its essential property; and this is exactly (W3).

In this connection, it would be helpful to consider how the argument strategy might fail. Suppose that one tries to establish that water necessarily has the ordinary identifying qualities (such as color, feel, odor, etc.). Imagine a situation where there exists a chemical substance Q that is just

like water (in its chemical composition and etc.) except that it lacks the ordinary identifying properties of water. If one wants to claim that it is a modal illusion that water is Q in this situation, then one would have to think that the superficial properties are essential to water; or, more generally, every chemical substance has its ordinary identifying properties (if any) as essential properties. Some qualifications should perhaps be stated, since it is not entirely clear what we mean by ordinary identifying properties. Under the intended reading, however, these claims do not seem to be definitional. The fact that we ordinarily use certain properties to identify something does not by itself mean that the thing is defined by those properties. In this case, therefore, we cannot conclude that the chemical substance Q is not water. Hence it is illegitimate to appeal to the notion of epistemic equivalence.

What about the impossibility of sound without medium? The same strategy seems to work in this case as well:

(S1) 'Sound' rigidly designates what gives us auditory sensations.

(S2) What gives us auditory sensations is the propagation of vibration, i.e., mechanical oscillation.

(S3) Every mechanical oscillation is necessarily through a medium.

(SC) Sound is impossible without a medium.

Notice again that (S3) has a definitional character. Mechanical oscillation is oscillation of matter, and a medium is nothing but the oscillating matter. In a situation where there appears to be sound phenomena without a medium, then, what is false would not be the proposition that sound is impossible without a medium, but the epistemic equivalent that what appears to us to be sound is impossible without a medium.

As a final example of scientific theory, consider the kinetic theory of heat. Though the idea that heat is molecular motion is quite popular, it would be more precise, according to modern physics, that heat is energy flowing into/from an object resulting in increase/decrease of the thermal energy of the object. As an account of the underlying nature of heat, the theory should imply that heating/cooling an object is impossible without increasing/decreasing motion of molecules constituting the object:

(H1) 'Heating an object' rigidly designates what gives us the sensation of heat.

(H2) What gives us the sensation of heat is the increase of the thermal energy of the object by transferring energy into it without adding any matter.

(H3) Any such increase of thermal energy of an object necessarily increases motion of molecules constituting the object.

(HC) Heating an object is impossible without increasing motion of the constituent molecules.

Here (H3) is again definitional; the thermal energy of an object is understood as energy consisting partly of motion of its constituent molecules.

Now, let's go back to Kripke's criticism of the mind-body identity theory. He does not think that the above strategy would work in the case of the identity of pain to C-fiber stimulation. He thinks that the very phenomenal quality of pain is its essential property (ibid., pp.152-53). So, if we imagine a situation where we appear to have that phenomenal quality without C-fiber stimulation, we do have pain in that situation. What is false in the situation should then be the identity of pain to C-fiber stimulation itself, but not some epistemically equivalent proposition. The apparent possibility of having pain without C-fiber stimulation, argues Kripke, cannot explained away by appealing to the notion of epistemic equivalence.[20]

This consideration suggests that the argument scheme would not work in the case of pain. To see where it goes wrong, imagine a situation where what exactly feels like pain is accompanied, say, by Z-fiber stimulation; call the phenomenal quality *quain*. Notice again that it does not immediately follow that pain is not quain.[21] For, it might be the case that pain just is quain, but it is accompanied by C-fiber stimulation in the actual world and, say, by Z-fiber stimulation in another possible world. To exclude this case, we have to assume that it is essential to pain that it is accompanied by C-fiber stimulation. With this assumption in place, it would indeed follow that pain is impossible without C-fiber stimulation. Putting this in the argument scheme, we would have:

(P1) 'Pain' rigidly designates what we feel in pain experiences.

[20]Note again that Kripke himself discusses the converse, namely the apparent possibility of having C-fiber stimulation without having the phenomenal quality. See ibid, p.154. See also footnote 12 above.

[21]Nor does it follow that pain *is* quain. For this conclusion, we need to assume that pain is nothing over and above whatever feels like it. Kripke explicitly assumes this, but it is dispensable according to the current analysis.

(P2) What we feel in pain experiences is the phenomenal quality that is accompanied by C-fiber stimulation.

(P3) Every phenomenal quality has the accompanying neural (brain, physical, etc.) state as its necessary property.

(PC) Pain is impossible without C-fiber stimulation.

Clearly, the argument draws an exact parallel in its form to the previous ones concerning scientific theories. There is a crucial difference, however. (P3) is far from having a definitional status. For it certainly does not seem to be a straightforward contradiction that it is impossible to have a certain phenomenal character (pain) without having the corresponding neural state (C-fiber stimulation). In fact, it is a direct challenge to (P3) that it appears possible that we feel pain without C-fiber stimulation.

The current analysis again tells us why the apparent possibility cannot be explained away simply by appealing to the notion of epistemic equivalence. The appeal is legitimate only if it is first established that pain is not quain. Well, this can certainly be done if we accept (P3). But the trouble is that, unlike the other cases, (P3) is not definitional; and so it is in need of justification. Hence, a challenge for the identity theorists is to give a defense of (P3). This challenge does not of course amount to a refutation of the mind-body identity theory, as Kripke himself notes (ibid, p.155). For there might be an alternative defense of (P3), or even an entirely different way of establishing (PC). In the absence of either, however, the apparent possibility of pain without C-fiber stimulation should remain a problem for the mind-body identity theory.[22]

[22]The current analysis of Kripke's argument against the mind-body identity theory is quite different from what we may call the *orthodox interpretation* that the apparent possibility of pain without C-fiber stimulation cannot be explained away because we are thinking of pain "directly" by its phenomenal quality. This is in contrast to the H_2O theory of water, according to this line of interpretation, because we often think of water via some associated descriptive content (say, "the colorless, odorless, ... liquid that we find in lakes."). See, for example, Loar (1990), Levine (2001), Papineau (2007), Garcia-Carpintero and Josep Maciá (2006) for this line of interpretation. As opposed to the orthodox interpretation, the current analysis suggests that the argument hinges not so much on whether we think of pain directly or via some descriptive content as it does on the conceptual connection between phenomenal states and the accompanying neural (brain, physical, etc.) states. I cannot go into the exegetical details here for reasons of space.

9. Back to the two pillars of teleosemantics

Let's return to our main topic. I have argued that the teleosemantic thesis should be understood as the modal claim that beliefs are necessarily selectional states. The claim is certainly aposteriori. So, it appears to be possible that beliefs are realized by non-selectional states. Indeed, the Swampman scenario describes one such possibility. Given Kripke's Requirement, teleosemanticists cannot simply dismiss the Swampman scenario as a non-starter. They need a principled argument to the effect that it is a mistake to think that the Swampman has beliefs.

Now can teleosemanticists adopt the general strategy to explain away the apparent possibility of the Swampman's having beliefs? The Swampman seems to have states that are exactly like beliefs except that they have no selectional history; call those states *queliefs*. Recall that teleosemantics explains the representational nature of beliefs in terms of their being selectional states. According to teleosemantics, then, queliefs are not representational and hence do not count as beliefs. Now it is certainly true that *if* queliefs are not representational then they are not beliefs. For it is part of the nature of beliefs that they are representational. Just because queliefs have no selectional history, however, it does not immediately follow that queliefs are not representational. For this conclusion, one should assume that it is essential for a state to be representational that it has a selectional history. Putting this into the argument scheme, we will have the following argument:

(B1) 'Belief' rigidly designates the kind of states that we attribute to organisms in usual ways.

(B2) Those states are representational states that satisfy conditions (T1)-(T3).

(B3) Every representational state satisfies conditions (T1)-(T3) as its necessary properties.

(BC) Beliefs are impossible without selectional histories.

The trouble again is that (B3) does not seem to have a definitional status. Some might not share the intuition that there can be ahistorical creatures with beliefs, but no one would say that it is self-contradictory. This strongly suggests that (B3) is not definitional. So, we have a challenge for teleosemantics: in order to establish (BC), teleosemanticists should provide an alternative defense of (B3).[23]

[23]Now one might worry that it is illegitimate to assume (B1) because as we have seen in section 4 Papineau explicitly refuses to make this assumption. Let me get clear about the dialectic. I have already argued that teleosemanticists should somehow establish (BC). Here I am considering whether it can be done by employing Kripke's strategy which,

How then can it be defended? One natural way would be to appeal to the two pillars of teleosemantics, namely success semantics and the etiological conception of biological function. As we have already seen in §2, (B3) can be considered as a consequence of them. First, according to success semantics, a representational state has the content P in virtue of its function of producing responsive behaviors that would benefit the subject just in case P holds. So, every representational state necessarily has a function of reliably prompting responsive behaviors that are appropriate under a certain condition. For a state to have such a function, according to the etiological conception, it must necessarily be historically selected for prompting responses that would be appropriate under a certain condition. So it follows that all representational states must necessarily satisfy (T1)-(T3). Assuming the two pillars, therefore, (B3) can indeed be seen as having a definitional status.

What are these two pillars, however? They are *theoretical conceptions* of representation and of biological function, respectively. Now recall that the major intuition about the Swampman scenario goes directly against (B3). In the light of this intuition, one may reasonably ask, "why must every representational state have a certain behavioral function?" and "why must the biological function of a state be determined by the selectional history behind it?" These questions concern the adequacy of the conceptual foundations on which the very explanatory power of the theory rests. In this way, the Swampman objection raises conceptual issues that are of great importance to teleosemantics.

Earlier I claimed that the Swampman objection should pose exactly the same problem to disjunctivist teleosemantics. Recall that the disjunctivists are committed to the claim beliefs are impossible without satisfying the disjunctive property $(S(x) \vee R_1(x) \vee ... \vee R_n(x))$. To establish this, one would have to modify (B2) and (B3) above as follows:

---

as far as I know, is the only way of establishing the necessity of aposteriori propositions. So the point is that if teleosemanticists are to employ Kripke's strategy to establish (BC), then they would have to make the semantic assumption of (B1). As far as I can see, moreover, assuming (B1) does not necessarily go against the spirit of scientific reduction in Papineau's sense. According to Papineau, recall, the H_2O theorists are concerned not with our ordinary concept of water but rather with *that chemical substance* that we typically call 'water'. So (W1) should be acceptable to them if only as a theoretical stipulation. Similarly, teleosemanticists are interested in the actual states that we typically call 'belief'. So (B1) should also be seen as acceptable, again, at least as a theoretical stipulation.

(B2′) Those states are representational states that satisfy the disjunctive property $(S(x) \vee R_1(x) \vee \ldots \vee R_n(x))$.

(B3′) Every representational state satisfies the disjunctive property as its essential property.

The problem is: how one might establish (B3′)? Again, it is difficult to settle this issue determinately because we have no clue at all as to what the alternative realizers $R_1, \ldots, R_n$ would be. At any rate, it does seem to be a very difficult task to come up with a proper disjunctive property that would indeed make (B3′) definitional.

The same point can be made with Neander's Swampcow as well. Why would we have to say the Swampcow is not a cow? She claims that according to contemporary evolutionary biology, it would not count as a cow because it lacks the right kind of causal connection. If this is true, then contemporary evolutionary biology is indeed making a conceptual commitment as to what a biological species is. So the Swampcow scenario indeed raises a conceptual issue: "why must conspecifics be linked by descent?" (Neander, 1996, pp.119-120). Can we really dismiss this question as irrelevant or unimportant to contemporary biology? If not, why dismiss the Swampman as irrelevant to teleosemantics?[24]

In sum, the standard strategy is not by itself sufficient to solve the Swampman problem. This is where the analogy between the H_2O theory of water and teleosemantics breaks down. Here it is of little help to try to fix this shortcoming by appealing to the two pillars of teleosemantics. For the intuition about the Swampman scenario goes directly against the theoretical analyses that they offer of the notions of representation and of biological function. This of course does not mean that the Swampman objection is a refutation of teleosemantics. There might be alternative ways of defending (B3) or entirely different strategies to establish (BC). I cannot discuss all possible moves here, but I hope that the discussion so far makes it clear that the Swampman scenario raises important conceptual issues for teleosemantics and that the issues have more to do with their fundamental commitments to success semantics and to the etiological conception of biological functions than with the notion of belief.

[24]Note that I am not claiming that our intuitions about the Swamp scenarios are by themselves sufficient to overrule contemporary biology and teleosemantics. The point is rather that they raise problems on the conceptual foundations of these theories and that those problems should not be dismissed as irrelevant to the theories.

10. Conclusion

By way of conclusion, let me briefly note how the discussion so far relates to the main problem of this dissertation. Recall that the main problem is: what makes necessary truths necessary and how do we recognize them as such? By applying Kripke's Principle, I first argued that what makes a proposition *P* necessary is that which excludes all the ways in which *P* might be falsified as unreal. What is it, then, that excludes the ways in which *P* might be falsified as unreal? Close examinations of the standard cases of theoretical identity statements seem to give us the beginning of an answer to this question. For they reveal that behind each standard case of theoretical identity statements there lies an analytic general principle—such as (W3), (S3) and (H3)—implying that the statement could not have been false if it is true at all. Given that the statement is in fact true, therefore, it follows from the general principle that the statement could not have been false. And this analysis also gives an easy explanation of how we can obtain knowledge of the necessity of the statement. Assuming that we have knowledge that the statement is in fact true, we can immediately infer from the general principle that the statement is necessary. There thus emerges a hint of a metaphysical and epistemological account of necessity.

CHAPTER 4

The Kripkean Explanation of Aposteriori Necessity: In the case of identity statements about chemical substances

1. Introduction

In the previous chapter, I have briefly hinted at a Kripkean account of how necessary aposteriori statements are possible. In such a case, there lies a general principle telling us that it is necessary if true at all; so, from empirical knowledge of its truth we may obtain empirical knowledge of its necessity. Though straightforward in its broad compass, this account faces two obvious questions in its application: in each case of necessary aposteriori statements, what is the underlying analytic principle, and how is it to be established? In the previous chapter, I have also given a brief treatment of these questions with respect to theoretical identity statements concerning chemical substances, such as 'water is H_2O'. There I contended, without much argument, that the principle underlying the necessity of theoretical identity statements concerning chemical substances is:

(NC) *if a chemical substance has a certain chemical composition, then it could not have had any other chemical composition*;[1]

and that it is analytic in that it can be derived from the theoretical concept of chemical substance. This chapter aims to substantiate these contentions with a sufficient level of formal rigor.

The plan for this chapter is as follows. It appears to be the standard treatment of the matter that the necessity of theoretical identity statements about chemical substances is a consequence of the principle that any identity statements between two rigid designators are necessary if true at all, which can be derived as a theorem of logic from the notion of rigid designation.[2] As opposed to this, I shall first argue that (NC) is required to establish the necessity of the theoretical identity statements about chemical substances. This is because terms like 'H_2O' designate the

[1]This is a slightly different, and yet equivalent, formulation of (W3) from the previous chapter.

[2]See, for example, Miller (2007, p.323), LaPorte (2013, pp.9-10 and §3.2 of 2022), Nimtz (2017, p.953)

corresponding chemical substances by describing their chemical compositions. Hence their rigidity can be established only if chemical substances have their chemical compositions as necessary properties—and this is exactly what (NC) states.

Then I turn to the issue of how exactly (NC) is to be established. There has been a considerable amount of controversy about this in the literature. As Burgess (2013, p.77) notes, Kripke himself has claimed that principles like (NC) can in general be established apriori by philosophical analysis, suggesting that "they are ultimately analytic, resting on rules of language." Though the suggestion may intuitively seem plausible, there has been a famous objection by Nathan Salmon (1979; 1981) that (NC) involves a substantive essentialist commitment and hence cannot be derived solely from conceptual or semantic grounds. His argument has been influential and many have followed him in this regard (Forbes, 1985; Bird, 2018). As opposed to Salmon and others, I shall argue that (NC) is derivable from the theoretical concept of chemical substance. After briefly explaining that Salmon's objection, though by no means groundless, is not decisive, I shall devote the core of this chapter to a novel derivation of (NC) from what can plausibly be taken as the theoretical notion of chemical substance. The derivation will be formalized in the appendix; and logical principles required for the derivation will also be stated and defended.

It should be noted that the claim that (NC) is derivable from the theoretical concept of chemical substance is not itself original. In fact, it has been maintained by some philosophers even in recent decades; most notably, Soames (2002, pp.273-274) and Hale (2013, p.280). Their cases for the claim, however, seem to rest largely on intuitions in want of a detailed argument on how exactly (NC) is derived.[3] This is unfortunate especially given the influence of Salmon's objection. This

[3]For example, see Hale (2013, p.280):

> To be a pure substance, in the chemical sense of the term, just is to be matter having a certain chemical composition—i.e. to be wholly composed of atoms of a certain element, or of molecules of a certain compound. There is, accordingly, no mystery about how we may know, a priori, the general principle needed for the application of Kripke's inferential model to substances. Our knowledge that gold is a certain element—the element with atomic number 79, and that water is a certain compound—the compound composed of molecules in which two hydrogen atoms covalently bond with an oxygen atom—is of course a posteriori, but we know, courtesy of the very meaning of the term, that if a substance has a certain chemical composition, it necessarily does so.

Here Hale writes as if it is obvious that (NC) follows from the concept of chemical substance. But how exactly is it supposed to follow? With this said, the influence of Hale's discussion on my own account will become clearer later on.

chapter will fill this lacuna in the literature by providing a novel derivation of (NC) from the theoretical concept of chemical substance with a sufficient level of formal rigor. In the course of argument, a new light will be shed upon how the logical form of theoretical identity statements concerning chemical substances should be understood. Then I will conclude with a brief discussion of how the present analysis extends to other necessary statements.

2. On why (NC) is needed

Let me begin with a brief review of Kripke's explanation of aposteriori necessity. Consider an identity statement '$a = b$' where both 'a' and 'b' are rigid designators. In this case, it can be established as a theorem of logic that

$$(a = b) \supset \Box(a = b); \tag{NI}$$

that is, any identity statement between two rigid designators is necessary if true.[4] For example, assuming that both 'Hesperus' and 'Phosphorus' are rigid, from the empirical discovery that Hesperus is Phosphorus we may conclude that the identity is necessary.

So far, so good. Complications come in with Kripke's claim that (NI) also applies to theoretical identity statements because they are typically between two rigid designators (1980, pp.140). Since 'water' and 'H_2O' are both rigid, according to his conception, we may infer from (NI) that the identity statement 'water is H_2O' is necessary if true. I do not want to dispute the rigidity of the term 'water'. What about the term 'H_2O', however? It is naturally understood as part of a notation system designed to pick out a chemical substance by specifying its chemical composition. In this respect, it is not a proper name in the ordinary sense; it behaves semantically like a definite description in that it is supposed to pick out a chemical substance which is composed of hydrogen and oxygen atoms in the ratio of 2:1.[5] In order to establish the rigidity of 'H_2O', therefore, it needs to be assumed that that chemical substance could not have had any other chemical composition. For, otherwise, there would be a possible world where that substance exists without being composed

[4]The proof should be familiar to the present readership. See Hale (2013, pp.260-263) for a detailed discussion. It should also be noted that the proof can be, and has in fact been, challenged. See Priest (2021, pp.1882-1884).

[5]For this line of analysis, see also Barnett (2000, p.109), Burgess (2013, pp.68-69; 2021, p.401) and Soames (2007, p.36).

of hydrogen and oxygen atoms in the ratio of 2:1; and the term 'H_2O' would not refer to it in that world. Of course, there is nothing special about the term 'H_2O'; similar considerations apply to chemical formulas in general. Generally, therefore, (NC) is required to establish the necessity of identity statements concerning chemical substances.

One might think that chemical formulas should be treated as proper names. Nathan Salmon (2003, p.488), for example, argues: "If the term is ... a general-term version of a proper name whose reference is fixed through a scientific convention concerning chemical-compound terms, then it is rigid *de jure* and the necessity of 'Water is H_2O' depends only on philosophical semantics (rigidity of 'H_2O') taken in conjunction with the empirical fact that water is the compound of two parts hydrogen, one part oxygen." That is, we identify a certain kind of matter as the chemical substance that is composed of hydrogen and oxygen atoms in the ratio of 2:1 and introduce the term 'H_2O' to rigidly designate that kind of stuff; in doing so, we do not intend to make the term synonymous with the description used to identify the reference. On this view, the necessity of the identity statement 'water is H_2O' is a consequence of (NI) together with the empirical discovery that the statement is true. To spell out Salmon's account in the form of argument:

(P1) The term 'water' rigidly designates a kind of stuff which science has discovered to be the chemical substance that is composed of hydrogen and oxygen atoms in the ratio of 2:1.

(P2) The term 'H_2O' is introduced to rigidly designate the chemical substance that is composed of hydrogen and oxygen atoms in the ratio of 2:1.

(C1) So, the terms 'water' and 'H_2O' are both rigid designators for the same thing.

(C2) By (NI), therefore, the identity statement 'water is H_2O' is necessary.

Hence the necessity of the identity statement can be established with no appeal to (NC).

Though this argument is certainly valid, Salmon seems to overlook a crucial point. The reference of a term may be fixed by a description in that the term is supposed to rigidly designate an entity which satisfies the description, but the fact that that entity satisfies the description may very well be contingent. For example, we may fix the reference of the term 'blue' as the color of the sky; but it is plainly a contingent fact that blue is the color of the sky. Similarly, we may introduce the term 'H_2O' to rigidly designate the chemical substance that is composed of hydrogen and oxygen

atoms in the ratio of 2:1, and yet it may well be a contingent fact that that chemical substance is so composed. This means that it is consistent with Salmon's key premise (P2) to assume that

(P3) it is contingent that H_2O is composed of hydrogen and oxygen atoms in the ratio of 2:1,

which, together with (C1), implies that

(C3) it is contingent that water is composed of hydrogen and oxygen atoms in the ratio of 2:1.

On Salmon's account, therefore, one may consistently hold both that water is necessarily H_2O and that it is contingent that water is composed of hydrogen and oxygen atoms in the ratio of 2:1. This means that the identity statement 'water is H_2O' is understood in such a way that the fact that it is necessary has no implication on what water is necessarily like in its chemical composition. Without this implication, however, it is difficult to see why in the first place anyone would insist on the necessity of the identity statement. Hence Salmon's account establishes the necessity of the identity statement only in name but not in substance.

How, then, can this defect be remedied?; that is, how in Salmon's account can the metaphysical import of the necessity of the identity statement be secured? It has to be assumed that the reference of the term 'H_2O' is fixed by its *necessary* property, i.e., that the chemical substance which is in fact composed of hydrogen and oxygen atoms in the ratio of 2:1 is necessarily so composed. But this amounts to assuming (NC) in the most general case. Even on Salmon's account, therefore, (NC) is indispensable to establish the necessity of identity statements concerning chemical substances with their intended implication.

3. The derivation of (NC) from the concept of chemical substance

How, then, can (NC) be established apriori? The answer, I shall now argue, is that it derives from the theoretical concept of chemical substance. Before turning to the derivation, let me first clear up the ground.

In the modern theoretical sense of the term, as Hale (2013, p.280) notes, a chemical substance is by definition a kind of matter with a unique chemical composition. For instance, H_2O is a chemical substance which is composed of hydrogen and oxygen atoms in the ratio of 2:1. Though this definition is fairly standard and straightforward, a couple of points are worth noting.

First, there are some complications as to how exactly the notion of chemical composition should be understood. What Hale seems to have in mind is elemental composition, i.e., the identity and proportion of constituent elements. Some might argue that this notion of chemical composition is objectionable because of isomerism. Isomers have the exactly same constituent elements in exactly the same proportions, and yet they are typically considered distinct substances because they exhibit different chemical and physical properties due to the difference in their molecular structures. Hence it might be argued that the notion of chemical substance ought to be defined in terms of molecular structure; and in fact, this appears to be more standard in actual practice even though it also has various problems.[6] Though this and other related issues are of great importance in their own right, they make little difference to the present discussion. For all we require of the notion of chemical composition is that it provides an adequate criterion of individuation for chemical substances. So, without further analysis, we understand it simply as whatever macro- and microscopic characteristics of matter that do the required job.

Second, the uniqueness condition also deserves a comment. It may happen that what is considered a single kind of matter turns out to be two (or more) distinct chemical substances. Jade, for example, was discovered to be two distinct kinds of chemical substances: nephrite ($Ca_2(Mg,Fe)_5Si_8O_{22}(OH)$) and jadeite ($NaAlSi_2O_6$). For this reason, jade itself is not considered a chemical substance; rather, the two chemical substances are said to be two different kinds of jade. This way, chemical substances are individuated in terms of their chemical compositions—hence the uniqueness condition.

Now, it may intuitively seem plausible that (NC) is a conceptual truth derivable from the theoretical definition of chemical substance. For, if a chemical substance is by definition a matter with a unique chemical composition, how could it possibly have another chemical composition? However, it has been put by Nathan Salmon that this line of thinking is fundamentally confused about the logic of quantified modality. He claims that (NC) has a substantive essentialist import in the sense that it states a necessary condition for the *cross-world* identification of chemical substances (Salmon, 1979, p.715). To see this, it may help to restate (NC) in the possible world terminology thus: if a chemical substance S_1 has a certain chemical composition c in the actual world, then

[6]See Needham (2000) and (2002) for discussion.

in any possible world w and for any chemical substance S_2, $S_1 = S_2$ only if S_2 has c as its chemical composition in w. In other words, chemical substances are identified in part by their *actual* chemical compositions *across different possible worlds*.

The problem here, Salmon thinks, is that such a cross-world criterion of identification is not derivable solely from conceptual grounds. For they only afford a criterion of *intra-world* identification. By way of illustration, let us consider an analogous case of sets. It may be admitted that it is part of the concept of set that sets are individuated by their members. So, we have the axiom of extensionality:

$$(x)(y)((x = y) \equiv (z)(z \in x \equiv z \in y)),$$

where it is understood that variables range over sets. Since this is definitive of sets, it is necessary that they satisfy the axiom; so, we have:

$$\Box(x)(y)((x = y) \equiv (z)(z \in x \equiv z \in y)). \tag{In}$$

But this only states that in any possible world w, two sets are identical just in case they have exactly the same member in that possible world w. In this sense, it only concerns when two sets are identical *within a given possible world*. But this does not imply the desired conclusion that a set could not have had different members; in symbol:

$$(x)(z)(z \in x \supset \Box(y)(x = y \supset z \in y)). \tag{Cr}$$

That is, if a set x has z as its member in the actual world, then it has z in any possible world.[7]

Here it may help to see the matter from the perspective of quantified modal logic. Notice that (In) involves only *de dicto* modality in the sense that it does not quantify into a modal context;

[7]It should be noted that Salmon himself does not discuss the case of sets in his 1979 paper. His main concern is with Kripke's argument for the so-called necessity of material origin, which states that if a material object has its material origin in a certain hunk of matter then it could not have had another material origin (a cross-world identification criterion for material objects). His analysis is basically that this cannot be derived from the principle that in any possible world w, two material objects are identical only if they have the same material origin (the intra-world criterion of identification for material objects) without additional substantive modal assumptions. It should be clear that this intra-world identification criterion for material objects is completely analogous in its logical form to the one for sets given above, namely (In). So, I discuss the case of sets for the sake of simplicity. See pp.705-707 p.711 of his (1979) where he claims that the same consideration applies to the case of chemical substances as well. For Salmon's discussion of chemical substance, see pp.166-169 and pp.183-189 of his (1981).

in other words, no quantifier has a wider scope than the modal operator. On the other hand, (Cr) has quantifiers that have a wider scope than the modal operator; and in this sense, (Cr) is said to involve *de re* modality. It is a well-known fact about quantified modal logic that de re modality is *not* reducible to de dicto modality in any standard systems of modal logic.[8] The trouble here is that the definition of a concept can only give rise to de dicto modalities. To see this, suppose that a predicate P is introduced to a language with the definition:

$$P(x) \equiv_{def} \phi(x),$$

for some formula $\phi(x)$. Then we shall have as a definitional axiom

$$(x)(P(x) \equiv \phi(x)),$$

which implies:

$$\Box(x)(P(x) \equiv \phi(x)).$$

But this involves only de dicto modality. So, we cannot conclude from this that the things that are in fact P are necessarily ϕ. For the claim would have the form:

$$(x)(Px \supset \Box\phi(x)),$$

which clearly involves de re modality. Hence no de re modality is derivable simply from conceptual or semantic grounds. So, in particular, there simply is no route from (In) to (Cr).

Like considerations apply to the case of chemical substances, according to Salmon. As noted above, chemical substances are by definition individuated by their chemical compositions. Unlike sets, of course, chemical substances have unique chemical compositions. But this makes little difference to the matter at hand. The definition of the concept of chemical substance only implies that in each possible world w, a chemical substance has a unique chemical composition that distinguishes it from others *within* w. In this sense, it only affords an intra-world criterion of

[8]See Hughes and Cresswell (1996, pp.250-254) for a detailed discussion.

identification for chemical substances. But this is strictly weaker than (NC), which gives a cross-world criterion of identification: chemical substances can be identified by their actual chemical compositions *across different possible worlds*.

Here, however, Salmon seems to overlook a crucial difference. Sets are most naturally understood as *objects*, such as individual tables, people, etc. On the other hand, it is kinds of matter—such as gold, water, etc.—that are said to be chemical substances. Each chemical substance should thus be considered a *property* and be distinguished from its instances, i.e., concrete physical quantities of matter.[9] So, strictly speaking, chemical substances are abstract entities; and as such, they are not composed of any physical matter.

What then do we mean when we say, for example, of water that it is composed of hydrogen and oxygen atoms in the ratio of 2:1? We mean that every molecule of water consists of two hydrogen and one oxygen atoms, and so that every quantity of (pure) water is composed of hydrogen and oxygen atoms in the ratio of 2:1. Generally, then, the notion of a kind K of matter having a certain chemical composition c can be understood thus:

(CC) *a kind K of matter has c as its chemical composition if and only if every quantity of K is composed c-wise.*

Hence the theoretical definition of chemical substance states, more precisely, that *a chemical substance is a kind of matter every instance of which is composed c-wise, for some unique chemical composition c.* For a cup of liquid to be an instance of H_2O, for example, is for it to be composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1. This suggests that the chemical substance H_2O can be identified with the property of being composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1. Indeed, the identification intuitively seems to be natural considering that a chemical substance is typically identified by how its molecules are composed in terms of the constituent elements.

To see what this conception of chemical substance implies, let us return to the identity statement 'water is H_2O'. On the current conception, the statement can be understood as expressing a property identity; specifically, it states that water—i.e., the property of being water as applied

[9]See Burgess (2013, p.67-68), Salmon (1981, p.177-178) and Soames (2006, p.714) for a similar analysis.

to quantities of matter—is identical to the property of being composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1. Since two identical properties are necessarily coextensive, it is immediate from the identity that

(1) *it is necessary that for any quantity x of matter, x is water if and only if x is composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1.*

Observe that (1) states a necessary fact about water as a property. This can be seen clearly if we restate it thus:

(2) *water is such that it is necessary that for any quantity x of matter, x is water if and only if x is composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1,*

where it is understood that the first occurrence of the term 'water' stands for the property of being water. By (CC), then, (2) may be restated thus:

(3) *water is necessarily composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1.*[10]

(3) thus follows from the identity of water to H_2O, or more expressly, from the identity of the property of being water to the property of being composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1.

Here it will help to add a bit of formal precision to this argument. Let us write $W(x)$ to mean that x is water and $H_2O(x)$ to mean that x is composed of H_2O molecules. We shall also write $\lambda x W(x)$ and $\lambda x H_2O(x)$ as the corresponding general term 'water' and 'H_2O'. Then we write $\lambda x W(x) \approx \lambda x H_2O(x)$ to mean that water is identical to H_2O as properties. Given this identity, it follows from the principle of necessary coextensiveness of identical properties that

$$\Box(x)(W(x) \equiv H_2O(x)), \tag{1*}$$

which is a symbolized version of (1). Notice that (1*)—and hence (1)—involves only de dicto modality for there is no quantification into a modal context. As noted above, however, it still states a necessary fact about the property of being water, namely that it is necessarily coextensive with

[10]Strictly, the equivalence between (2) and (3) requires an additional assumption that there is an actual instance of H_2O. See the derivation of (NC) below.

the property of being H_2O. This can be expressed using the second-order machinery of property identity thus: letting X be a second-order variable,

$$(X)((\lambda x X(x) \approx \lambda x W(x)) \supset \Box(x)(X(x) \equiv H_2O(x))), \tag{2*}$$

which is a symbolized version of (2). Notice that (2*) involves a second-order quantification into a modal context. With the help of the machinery of second-order logic, therefore, there is a way to obtain de re modality from de dicto modality.[11]

Again, there is nothing special about water. Any kind of matter counts as a chemical substance if it can be identified in the same way as indicated above, and conversely. So, we may define the notion of a chemical substance as a second-order property thus:

(DC) *a kind S of matter is* a chemical substance *if and only if there is a chemical composition c such that S is identical to the property of being composed uniquely c-wise.*

The second-order formulation is not objectionable; for, to repeat, we say of kinds of matter, which are themselves properties, that they are chemical substances.

Turning finally to the derivation of (NC), we first show as a lemma:

(E) for any kind S of matter and any chemical composition c, S is the chemical substance having c as its chemical composition if and only if S is identical to the property of being composed uniquely c-wise.

Suppose (i) that a kind S of matter is a chemical substance, and also (ii) that it has a certain chemical composition c. We show that S is identical to the property of being composed uniquely c-wise. Note that by (DC), it follows from (i) that there is a certain chemical composition c' such that S is identical to the property of being composed uniquely c'-wise. Assume, for *reductio*, that $c \neq c'$. Let x be any sample of S. Since S is identical to the property of being composed uniquely c'-wise, on the one hand, it follows that x is composed uniquely c'-wise. This implies, by the *reductio*

[11] Here I am using the term 'de re modality' in the broad sense that a statement involves de re modality just in case it quantifies into a modal context, regardless of whether the quantification is first-order or second-order. However, it is perhaps helpful to make a distinction between the two cases. Note that (Cr) contains a first-order quantification into a modal context; so, it states what is necessary of *individual sets*. In contrast, (2*) contains a second-order quantification into a modal context. So, it states what is necessary of *water as a kind of matter*, but it has no implication on what is necessary of particular instances of water. So, there is an important difference between first-order and second-order de re modalities. For want of a better term, I should perhaps call the latter *de qualitate* modality.

assumption, that x is not composed c-wise. By (ii), on the other hand, S is assumed to have c as its chemical composition. This means, by (CC), that every instance of S is composed c-wise; so, in particular, x is composed c-wise. Contradiction. By *reductio*, therefore, $c = c'$. Hence we may conclude that S is identical to the property of being composed uniquely c-wise. The other direction is trivial, so we omit it.

To finally derive (NC) from (E), suppose that S is the chemical substance having a certain chemical composition c. By (E), then, S is identical to the property of being composed uniquely c-wise. Consider any possible world w where there exist instances of S, and pick an arbitrary sample x of S in w. Notice that in w, x is composed uniquely c-wise. This is because S is identical to the property of being composed uniquely c-wise and because two identical properties are necessarily coextensive. So, for any other chemical composition c', x is not composed c'-wise. By (CC), then, it follows that S has only c as its chemical composition in w. Since w is an arbitrary possible world where S has instances, it follows that in any possible world where S instantiates, S has only c as its chemical composition; that is, S could not have had any chemical composition other than c. This completes the derivation of (NC) from (DC) and (CC). Since both (DC) and (CC) are pieces of conceptual analysis, the derivation shows that (NC) is also a conceptual truth. (See Appendix below for the formalization of the derivation of (NC).)

A couple of comments on the derivation. First, the derivation of (E) makes use of an extra assumption that S has an actual instance. I do not find this objectionable. But if one does, then we may simply restrict (NC) to chemical substances with actual instances; nothing significant would thereby be lost. (See, however, my discussion of the first objection in the next chapter and the footnote appended.) Second, the derivation can be formalized in the system K—hence, in any normal system—of second-order modal logic with a property-term forming operator. This means that the derivation depends on the minimal assumption about metaphysical modality that its logic is normal.

4. Possible objections

Let me discuss a few possible objections. In deriving (NC) from (E), first, we have restricted the set of possible worlds to those where *S* has instances. This restriction is reasonable for most practical purposes, because in thinking about what chemical composition a kind of matter could have, we would naturally consider those worlds where it has instances. One might nevertheless wonder what chemical composition *S* has in possible worlds where *S* has no instances. In such a world, it is vacuously true that for any chemical composition *c*, every quantity of *S* is composed *c*-wise. This means by (CC) that in possible worlds where *S* has no instances, *S* may vacuously have any chemical composition. This consequence may seem objectionable, if one thinks that a chemical substance should be considered to have, in some non-vacuous sense, a definite chemical composition even in those worlds where it has no instances. It is difficult to address this worry in a determinate way because it is not clear what exactly such non-vacuous sense might be. But one idea may be something like this: even in a possible world *w* where *S* does not instantiate, the following counterfactual statement may still be true: *S could have had only c as its chemical composition had it instantiated*—in such a case, *S* should be said to have *c* as its unique chemical composition in *w*. Though reasonable it may seem, the suggestion poses little problem to the present account. For it can be easily accommodated within the present account, assuming transitivity of the accessibility relation between possible worlds (or, equivalently, axiom 4: $\Diamond\Diamond P \supset \Diamond P$), which is relatively unproblematic. To see this, let w' be any world which is accessible to *w* and in which *S* has instances. By the assumption of transitivity, then, w' is also accessible to the actual world. Since it is assumed that there are instances of *S* in w', the reasoning from the previous section applies; so, *S* has no chemical composition other than *c* in w', i.e., in any world which is accessible to *w* and in which *S* instantiates. Hence the counterfactual statement holds in any world where *S* has no instances; in this sense, *S* may still be said to have *c* as its unique chemical composition even in worlds where it has no instances.[12]

[12]So, according to my original account, if a substance *S* has never had and will never have any instance in the actual world, then it is vacuously true in the actual world that *S* has any chemical composition. As I noted above, it is not entirely clear to me that this consequence is in itself objectionable. However, my response to the first objection above shows that, if one wants, there is a way to assign a determinate chemical composition to such a substance in the counterfactual sense specified above. So, the restriction on (NC) can be lifted with the help of axiom 4.

Let me turn to the second objection. The derivation depends on the principle that two identical properties are necessarily coextensive. I take this to be a logical principle; it can be proved in any standard system of second-order modal logic with Leibniz's law for second-order identity, which can be stated thus: *if the property of being F is identical to the property of being G then the corresponding predicates 'F' and 'G' are intersubstitutable salva veritate*. Since F is necessarily coextensive with itself (in symbol, $\Box\forall x(Fx \equiv Fx)$), it follows from Leibniz's law that if the property of being F is identical to the property of being G, then it is necessary that the two properties are coextensive (in symbol, $\Box\forall x(Fx \equiv Gx)$). Some might nevertheless object that this principle is subject to counterexamples of the following sort. As a matter of historical fact, Isaac Newton was unmarried for his entire life time. So, the property of being a bachelor may be identified with the gendered marital status of Newton. But it is certainly a contingent fact that Newton is unmarried. Then consider a possible world where Newton is married to someone. In that world, the property of being a bachelor is not coextensive with the gendered marital status of Newton. Hence, the objection goes, the principle is invalid.

To answer this objection, let me make a distinction. We may identify a property by describing the property itself as a unique satisfier of a certain second-order property. In the case just mentioned, for example, the second-order property of being a gendered marital status of Newton is used to describe the property of being a bachelor itself. But it does not serve to specify the characteristics of individual bachelors. It is plainly absurd to say of individual bachelors that they *are* a gendered marital status of Newton, though, of course, they *have* the same gendered marital status as Newton. This way of identifying a property we may call the *descriptive* identification. On the other hand, we may identify a property by specifying a set of characteristic features of its instances. An example of this is the identification of the property of being a bachelor with the property of being an unmarried man. In making this identification, we certainly do not say of the property of being a bachelor that it is unmarried or that it is a man; rather, we intend to state what it is for a person to be a bachelor, i.e., to specify the characteristics of its instances. In this sense, we may call this way of property identification *specificative*.

Notice that descriptive identifications admit of a Russellian analysis of description in terms of specificative identification. The identity statement 'the property of being a bachelor is identical to the gendered marital status of Newton', for example, can be analyzed as expressing that there is a unique property of being an F such that Newton is an F in terms of his marital status and such that the property of being a bachelor is identical, in the specificative sense, to the property of being an F. With this analysis we can easily see the sense in which the property of being a bachelor is necessarily coextensive with the gendered marital status of Newton: on the wide scope reading of the description, it means that there is a unique property of being an F such that, as a matter of contingent fact, Newton is an F in terms of his marital status and such that the property of being an F is necessarily coextensive with the property of being a bachelor. But this does not contradict the narrow scope reading, on which the objection is based, that it is necessary that whatever gendered marital status Newton has is coextensive with the property of being a bachelor. Cases of this sort, therefore, do not constitute a counterexample to the principle of necessary coextensiveness of identical properties.[13]

In this connection, it may help to briefly reconsider how the identity statement 'water is H_2O' may be understood. I have suggested two readings of it. It may be read, perhaps most naturally, as the descriptive identification of water with the chemical substance composed of hydrogen and oxygen atoms in the ratio of 2:1. But it can also be understood as the specificative identification of water with the property of being composed of hydrogen and oxygen atoms in the ratio of 2:1. Though the latter may not be as intuitive as the former, it is important to note that, given (CC) and (DC), the two readings are equivalent; and this is basically what (E) shows.

Finally, one might think that even though the derivation of (NC) is itself sound, it is a mistake to conclude that (NC) is a *conceptual* truth. For the conclusion seems to imply that the necessity of theoretical identity statements concerning chemical substances is reduced to a form of analyticity.

[13]This response to the objection should be reminiscent of the standard defense of (NI) against counterexamples involving non-rigid definition descriptions of objects. This might lead one to think that, for terms for properties, the notion of specificative identification coincides with the notion of rigid designation. It appears to be true that a term rigidly designates a property if the term identifies the property in the specificative sense. The converse, however, may fail. For instance, suppose that there is a person, call him *Mr. RedLover*, whose favorite color is red in every possible world. Then the description "Mr. RedLover's favorite color" should pick out the property of being red in every possible world. This, however, does not mean that the description identifies the property of being red in the specificative sense.

But it is obvious that the necessity of the identity of water to H_2O, for example, is a substantive metaphysical claim, which can never be established solely on conceptual or semantic grounds. What this means, the objection continues, is that (DC) should not be taken as a nominal definition, i.e., one that defines the meaning of the theoretical term 'chemical substance'; rather, it ought to be understood as a *real* definition, i.e., one that defines *what it is to be a chemical substance* itself. Hence (NC) should be rather understood as a consequence of the nature of chemical substance.[14]

In response, note that there is no substantive sense in which the proposed Kripkean explanation of aposteriori necessity *reduces* the necessity of theoretical identity statements concerning chemical substances to a form of analyticity. To see this, let us consider how the proposed explanation is supposed to work in specific instances. In the case of water, for example, the explanation has three elements. One is the rigidity of the ordinary natural kind term 'water'. Another element is what I have argued to be a conceptual truth, namely (NC) that every chemical substance has its chemical composition as a necessary property. Now, these two are certainly semantic elements in some suitably broad sense. However, according to the present analysis, the two semantic elements are not by themselves sufficient to establish the necessity of the statement 'water is H_2O'. For this, we also need the third element, namely the scientific discovery of the nature of water, namely that water is in fact a chemical substance composed of hydrogen and oxygen atoms in the ratio of 2:1. Hence it is not quite correct to say that the proposed analysis purports to derive substantive metaphysical claims *solely* from conceptual or semantic grounds. Quite the contrary: according to the present analysis, the necessity of identity statements concerning chemical substance is explained (in part) by the nature of each chemical substance, i.e., *what each chemical substance is*.

5. Conclusion

The Kripkean explanation of aposteriori necessity is that in each case of necessary aposteriori statements, there is an apriori general principle telling us that it is necessary if true. Applying this to the case of identity statements concerning chemical substances, we face two problems: what is the

[14]This so-called essentialist idea has become increasingly popular in the recent literature on metaphysical necessity under the influence of Kit Fine (1994), Bob Hale (2002, 2013) and Edward Jonathan Lowe (2008, 2012). It should be noted that Lowe thinks that that the identity statements concerning natural kinds are in general not necessary, though he is certainly in the essentialist camp. See his (2007) for detailed discussion.

principle underlying the necessity of those statements, and how can it be established apriori? I have first argued that the underlying principle is (NC). For terms for chemical substances are understood as designating chemical substances by describing their chemical compositions. Then I have argued that (NC) is a conceptual truth derivable from the theoretical concept of chemical substance. A chemical substance is by definition a kind of matter with a unique chemical composition. This must be interpreted with caution. Chemical substances themselves are properties. So, when we say of a quantity of matter that it is (an instance of) H_2O, we mean that the quantity of matter is composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1. Hence the definition states that each chemical substance can be identified—in the specificative sense—with the property of being composed uniquely in a certain way; and this is how (DC) defines the notion of chemical substance. Then we saw that (NC) is derivable from (DC) taken together with (CC). The key to the derivation was (E), which shows the equivalence between the descriptive and specificative identifications of chemical substances.

Before closing, let me briefly discuss how the present analysis extends to other cases. It should be clear that it extends to other theoretical identity statements. Consider, for example, the identity between heat and molecular motion. As noted in the previous chapter, the general principle underlying the necessity of this identity is that it is impossible to increase the thermal energy of an object without increasing motion of the constituent molecules (assuming that the mass of the object remains the same). And this can be established from the theoretical definition of thermal energy: the thermal energy of an object is identical, in the specificative sense, to energy consisting partly of motion of its constituent molecules.

Appendix

In this appendix, I shall formalize the derivation of (NC). Logical principles required for the derivation will also be stated and defended.

Let $\phi(x)$, $\psi(x)$, $\cdots$ be formulas with one open variable x. We also let $\overline{\phi}$, $\overline{\psi}$, $\cdots$ be the properties denoted by predicates $\phi(x)$, $\psi(x)$, $\cdots$. We use $\lambda x\phi(x)$, $\lambda x\psi(x)$, $\cdots$ as terms for $\overline{\phi}$, $\overline{\psi}$, $\cdots$. For any $\phi(x)$ and $\psi(x)$, we write $\lambda x\phi(x) \approx \lambda x\psi(x)$ to say that $\overline{\phi}$ is identical to $\overline{\psi}$ as properties in the

specificative sense. Given a unary predicate $P(x)$ and a formula $\phi(x)$, we shall write $P(x) \equiv_{def} \phi(x)$ to mean that $P(x)$ is (introduced as) a shorthand for $\phi(x)$.

We now formulate two basic principles:

(P1) For any unary predicate $P(x)$ and a formula $\phi(x)$, if $P(x) \equiv_{def} \phi(x)$, then $\lambda x P(x) \approx \lambda x \phi(x)$.

(P2) For any formulas $\phi(x)$ and $\psi(x)$, if $\lambda x \phi(x) \approx \lambda x \psi(x)$ then $\xi \supset \xi[\phi/\psi]$, where $\xi[\phi/\psi]$ is a formula obtained by substituting ϕ in zero or more (but not necessarily all) occurrences for ψ in ξ.

Comments and qualifications are in order. First, (P1) and (P2) together imply that

$$\textit{if} \quad P(x) \equiv_{def} \phi(x), \quad \textit{then} \quad (X)((\lambda x X x \approx \lambda x P(x)) \supset \Box(x)(P(x) \equiv \phi(x))), \tag{D-N}$$

To see this, suppose that $P(x) \equiv_{def} \phi(x)$. By (P1), then, we have $\lambda x P(x) \approx \lambda x \phi(x)$. By (P2), then, we may substitute P for ϕ in $\Box(x)(\phi(x) \equiv \phi(x))$, which is a logical truth, thereby obtaining $\Box(x)(P(x) \equiv \phi(x))$, which is equivalent to $(X)((\lambda x X x \approx \lambda x P(x) \supset \Box(x)(P(x) \equiv \phi(x)))$. This shows that de re—or, more precisely, de qualitate—modality can be obtained by the definition of a concept.

Second, there is nothing substantively metaphysical about principles (P1) and (P2). (P1) can be justified on the purely semantic grounds that the explicit definition of the form $P(x) \equiv_{def} \phi(x)$ stipulates that the defined and defining expressions have the same meaning, and that two expressions with the same meaning have the same denotation. (P2) is a second-order version of Leibniz's Law and hence can reasonably be taken as a logical principle. Here it might be objected that (P2) is not valid. Suppose that Mary believes that this cup of coffee is mostly water. Since water is H_2O, it follows from (P2) that Mary believes that this cup of coffee is mostly H_2O. But it might not be correct to say that she has that belief. For, perhaps, she does not know that water is H_2O. So here we seem to have a counterexample to (P2). It is certainly true that (P2) ought to be properly qualified when our language has expressive resources for attitude reports. There might be other contexts in which (P2) turns out to be problematic. We do not have to settle this issue here decisively. For the purpose of this chapter, it suffices to note that (P2) is valid when we interpret the modal operator $\Box$

as standing for metaphysical necessity. This point is perhaps obvious to anyone who is sufficiently familiar with the standard semantics of quantified modal logic. For, if $\lambda x\phi(x) \approx \lambda x\psi(x)$, then $\phi(x)$ and $\psi(x)$ should have the same interpretation in every possible world. Hence $\phi(x)$ and $\psi(x)$ ought to be satisfied by the same objects in every possible world. Hence substitution of $\phi(x)$ for $\psi(x)$ in any formula should preserve the truth-value.[15] Notice here that substitution may occur within lambda terms. The lambda operator, as I use it, is a property-term forming operator: its application to a predicate (e.g., '... is water') gives a term ('water') for the corresponding property. So, intuitively, if the property of being water is identical to the property of being H_2O, then 'H_2O' should be substitutable for 'water' *salva designatione*, say, in 'the property of being as dense as water'.[16]

To state the definition of the concept of chemical composition, we write $\Phi(x,y)$ to mean that a particular quantity x of matter is chemically composed y-wise. Then we define the notion of chemical substance—in symbol, $\Psi(\lambda xX(x))$—as follows:

$$\Psi(\lambda xX(x)) \equiv_{def} (\exists y)(\lambda xX(x) \approx \lambda x(\Phi(x,y) \wedge (z)(\Phi(x,z) \supset z = y))). \qquad \text{(DC*)}$$

For a property $\overline{X}$ to be a chemical substance is for $\overline{X}$ to be the property of being matter having y as its unique chemical composition, for some chemical composition y. A moment's reflection should be sufficient for one to see that (DC*) is a straightforward formalization of (DC).

[15]The argument assumes that an interpretation of an open formula is given by its satisfaction condition at each possible world.

[16]A bit more formally, we need to show the validity of the following principle:

$$\textit{if } \lambda x\phi(x) \approx \lambda x\psi(x) \text{ then } \lambda x\xi(x) \approx \lambda x\xi[\phi/\psi](x),$$

where $\lambda x\xi[\phi/\psi](x)$ is the result of substituting ϕ in zero or more (but not necessarily all) occurrences of ψ in $\lambda x\xi(x)$. The principle is valid if we give an open formula $\phi(x)$ and the corresponding lambda term $\lambda x\phi(x)$ the same interpretation, namely the satisfaction condition at each possible world. To see this, suppose that $\lambda x\phi(x) \approx \lambda x\psi(x)$. Then $\phi(x)$ and $\psi(x)$ have exactly the same satisfaction condition at each possible world, which in turn implies that $\xi(x)$ and $\xi[\phi/\psi]$ have the same satisfaction condition at each possible world. Hence the corresponding lambda terms $\lambda x\xi(x)$ and $\lambda x\xi[\phi/\psi](x)$ will receive the same interpretation. Therefore, $\lambda x\xi(x) \approx \lambda x\xi[\phi/\psi](x)$. It might be objected that this proof relies on the implicit understanding of properties as intensions, which is suspicious. Though I am somewhat sympathetic to this objection, I also believe that any reasonable calculus of properties should validate the above principle. For our purpose, it is sufficient to note that the principle is valid in a natural extension of the standard semantics for second-order modal logic.

Now we derive the lemma (E), which can be stated in symbols thus:

$$(\Psi(\lambda x S(x)) \wedge (x)(X(x) \supset \Phi(x,c))) \;\equiv\; (\lambda x S(x) \approx \lambda x(\Phi(x,c) \wedge (z)(\Phi(x,z) \supset z = c))). \quad \text{(E*)}$$

We verify the left-to-right direction; the other direction is immediate by (DC*). Suppose that a kind $\overline{S}$ of matter is a chemical substance having c as its chemical composition; that is,

$$\Psi(\lambda x S(x)) \;\wedge\; (x)(S(x) \supset \Phi(x,c)). \quad \text{(I)}$$

By (DC*) and (P1), then, it follows from the left conjunct that:

$$(\exists y)(\lambda x S(x) \;\approx\; \lambda x(\Phi(x,y) \wedge (z)(\Phi(x,z) \supset z = y))).$$

By existential instantiation, we have: for some c',

$$\lambda x S(x) \;\approx\; \lambda x(\Phi(x,c') \wedge (z)(\Phi(x,z) \supset z = c')). \quad \text{(II)}$$

Assume, for reductio, that $c \neq c'$. Let a be any instance of $\overline{S}$. From the right conjunct of (I), on the one hand, it follows that $\Phi(a,c)$. Since a is an instance of $\overline{S}$, on the other hand, (II) implies by (P2) that $\Phi(a,c') \wedge (z)(\Phi(a,z) \supset z = c')$. Since we have assumed that $c \neq c'$, it follows from the second conjunct that $\neg\Phi(a,c)$; contradiction. Hence $c = c'$. Then by substituting c for c' in (II), we obtain:

$$\lambda x S(x) \;\approx\; \lambda x(\Phi(x,c) \wedge (z)(\Phi(x,z) \supset z = c)), \quad \text{(III)}$$

which states that $\overline{S}$ is identical in the specificative sense to the property of having c as its unique chemical composition. This completes the proof of (E).

Finally, we derive (NC). Recall that (I) states that a kind $\overline{S}$ of matter is a chemical substance having c as its chemical composition. To derive (NC), therefore, we suppose (I) and show that the assumption that there can possibly be an instance of S that is composed in another way leads to a contradiction. So, for *reductio*, assume that:

$$\Diamond(\exists x)(\exists z)(S(x) \;\wedge\; \Phi(x,z) \;\wedge\; z \neq c).$$

Now, (I) implies by (E*) that

$$\lambda x S(x) \approx \lambda x(\Phi(x,c) \wedge (z)(\Phi(x,z) \supset z = c)).$$

This implies by (P2) that

$$\Box(x)(S(x) \equiv (\Phi(x,c) \wedge (z)(\Phi(x,z) \supset z = c))),$$

which logically implies

$$\Box(x)(z)(S(x) \supset (\Phi(x,c) \supset z = c)),$$

which is logically equivalent to

$$\neg\Diamond(\exists x)(\exists z)(S(x) \wedge \Phi(x,z) \wedge z \neq c),$$

contradicting the reductio assumption. Q.E.D.

CHAPTER 5

Critique of Essentialism—Preliminaries

1. Introduction

In the previous two chapters, I have argued that behind each standard case of theoretical identity statements there lies a general principle implying that the statement could not have been false if it is true at all, and that the principle is analytic in the sense it can be derived from the definitions of relevant concepts. Is this appeal to analyticity indispensable for giving an account necessity, however? In this and the next two chapters, I shall mainly be concerned with an approach that seeks to give an account of necessity without appeal to analyticity, and argue that it faces a stiff challenge.

Inspired by Kripke's discussion and Fine's pioneering paper (1994), many recent philosophers have been attracted to what I call the (*orthodox*) *essentialist approach to necessity*. Instead of appealing to analyticity, this approach seeks to give an account of necessity as having its metaphysical and epistemological source in essence. More specifically, it maintains that necessary truths are those that hold in virtue of the essences of things, and so that knowledge of necessity derives from knowledge of essence. The essentialist approach faces two prior questions: what is essence and how do we recognize it? In other words, it should be accompanied by a prior metaphysical and epistemological account of essence. Below I shall argue that this demand creates a serious problem for the essentialist approach.

In the recent essentialist literature, there appear to be two major metaphysical accounts of essence. One is the explanation-based account, according to which the essence of an entity consists in the property that explains why the entity is the way it is. This account has recently been endorsed and developed by Godman et al. (2020), Kment (2014), and Mallozzi (2021a). The other is the definition-based account which understands the notion of essence on the model of definition. This account has been favored most notably by Fine (1994), Hale (2002, 2013) and Lowe (2012). In

the next two chapters, I argue that both accounts are up against a stiff challenge. I first argue that the explanation-based account is explanatorily inadequate in the sense that it fails to explain why the essence of an entity should be considered as its necessary property. Then I argue that the definition-based account of essence is epistemologically inadequate in the sense that it fails to give a reasonable epistemological account of essence. The upshot will thus be that the essentialist approach faces a version of what Christopher Peacocke (1999) calls the *integration challenge*: no metaphysics of essence can plausibly be reconciled with a reasonable epistemology of essence while at the same time satisfying the original explanatory goal of the essentialist approach.

The aim of this chapter is to briefly clear up the ground before launching into criticism. I will begin with a review of the basic ideas of the essentialist approach. Then I shall present three constraints that any prior account of essence must satisfy for it to be adequate for the purpose of the essentialist approach. On the basis of the constraints, I shall formulate the integration challenge for the essentialist approach.

2. The essentialist approach to necessity

In his well-known paper "Essence and Necessity" (1994), Fine claims that there is a problem with the *modal conception* of essence. According to this conception, the essence of an entity is the sum of its necessary properties, i.e., those properties that the entity could not fail to have if it exists at all. Fine argues that this conception has counterintuitive consequences. For example, consider Socrates and $\{\text{Socrates}\}$, i.e., the set whose sole member is Socrates. It is a standard assumption in modal set theory that it is necessary that Socrates belongs to $\{\text{Socrates}\}$ if he exists. According to the modal conception, then, it is part of the essence of Socrates that he belongs to $\{\text{Socrates}\}$ as its sole member. This consequence is objectionable, however, because it does not seem to be part of the essence of Socrates that he is a member of a certain set (ibid, pp.4-5). Indeed, this line of argument can be generalized to any entity and to any necessary truth. Consider, for instance, Socrates and the mathematical truth that $2+2=4$. Certainly, it is necessary that Socrates is such that $2+2=4$ if he exists. According to the modal conception, then, it is in the essence of Socrates that $2+2=4$. This way, the modal conception implies that for any entity X and for any necessary

truth P, it is part of the essence of X that P holds. But in the case just mentioned, what does Socrates have to do with the mathematical truth that $2+2=4$? Hence something must have gone wrong, Fine concludes, with the modal conception.

Why is the modal conception subject to counterexamples like the ones above? Fine says:

> [A]ny essentialist attribution will give rise to a necessary truth; if certain objects are essentially related then it is necessarily true that the objects are so related (or necessarily true given that the objects exist). However, the resulting necessary truth is not necessary simpliciter. For it is true in virtue of the identity of the objects in question; the necessity has its source in those objects which are the subject of the underlying essentialist claim. $\cdots$ The concept of metaphysical necessity, on the other hand, is insensitive to source: all objects are treated equally as possible grounds of necessary truth; they are all grist to the necessitarian mill. (ibid, pp.8-9)

That is, the attribution of an essential property to an entity indicates the source of the corresponding necessary truth. So, for example, when we say that that it is in the essence of Socrates that he belongs to $\{$Socrates$\}$, this implies that it holds by virtue of what Socrates is that he belongs to $\{$Socrates$\}$. But this is a mistake, according to Fine, because the truth presumably holds by virtue of what $\{$Socrates$\}$ is, but not of what Socrates is. In contrast, the attribution of a necessary property is blind to source. It is necessary of Socrates that he belongs to $\{$Socrates$\}$ as its sole member, and this is regardless of in virtue of what this truth holds. In this sense, there is no backward road from necessity to essence. The modal conception is thus fundamentally misguided in attempting to explain the notion of essence in terms of necessity. The correct picture, Fine contends, is the other way around: necessity ought to be explained in terms of essence.

Let me make a couple of remarks on this critique of the modal conception of essence. First, one might find an implicit argument here that the notion of essence is *hyper-intensional* in the following sense: that a certain property Φ is necessarily equivalent to the essence Ψ of an entity does not imply that Φ constitutes in the essence of the entity. Consider, for example, the property of being a set whose sole member is Socrates and the property of being a set whose sole member is Socrates and such that $2+2=4$. These two properties are necessarily equivalent; in any possible

world, whatever has the first property also has the second and vice versa. Moreover, it may be admitted that the first property constitutes the essence of $\{\text{Socrates}\}$. From these, however, it should not follow, according to Fine's critique, that the second property also constitutes the essence of $\{\text{Socrates}\}$. For that would imply that it is part of the essence of $\{\text{Socrates}\}$ that $2+2=4$.

Second, it is important to note that Fine is not denying that any essential property of an entity is also a necessary property. Indeed, as noted by Mackie (2020, p.249), it is in fact crucial for the goal of explaining necessity as having its metaphysical source in essence to establish what we may call the *bridge principle*: for any properties Φ and Ψ,

(BPP) *if* Ψ *is part of the essence of* Φ*, then necessarily, every* Φ *is* Ψ.

Here the formulation is given in terms of properties. But, of course, we can also formulate an analogous principle for individual entities also: for any individual x and a property Φ,

(BPO) *if* Φ *is part of the essence of x, then necessarily x is* Φ.

A few points are worth noting about these bridge principles. First, it should be clear that they encapsulate the core essentialist idea that necessity has its source in essence. Second, notice the difference in the logical forms of (BPP) and (BPO). The consequent of (BPP) involves only *de dicto* necessity in the sense that it contains no open individual variables in the modal context. That is, its consequent concerns only the conditions for a thing to count as Φ in the strict sense, but not what the individual things that are actually Φ must necessarily be. On the other hand, (BPO) involves *de re* necessity in that its consequent contains an open individual variable in the modal context; it states of a particular individual x that x must necessarily be Φ. Finally, observe that these principles offer a straightforward account of how we may obtain knowledge of necessary truths. Suppose, for example, that we know that being composed of H_2O molecules is part of the essence of water. Then we may conclude from (BPP) that it is necessary that every quantity of water is composed of H_2O molecules thus:

(1) Being composed of H_2O molecules is part of the essence of water.

(2) If being composed of H_2O molecules is part of the essence of water, then necessarily every instance of water is composed of H_2O molecules.

(3) Necessarily, therefore, every instance of water is composed of H_2O molecules.

Here (2) is clearly an instance of (BPP). Similarly, (BPO) allows us to infer a necessary truth about an object from a truth about its essence. Assuming that it is part of the essence of Socrates that he is a human being, for instance, we may establish that Socrates is necessarily a human being as follows:

(1′) Being a human being constitutes part of the essence of Socrates.

(2′) If being a human being is part of the essence of Socrates, then necessarily Socrates is a human being.

(3′) Necessarily, therefore, Socrates is a human being.

This way, according to the essentialist account, knowledge of necessity can be obtained from knowledge of essence.

3. The questions about essence

Once its main theses are spelled out along these lines, it becomes obvious that the essentialist approach ought to address a couple of prior questions about essence: *what is essence, and how do we recognize it?* In other words, the essentialist account should provide a prior account of what it is for a certain property to be part of the essence of an entity and also of how such facts about essence may be known. It would otherwise only create more problems than it solves.

Now, there are a few obvious constraints that any acceptable prior account of essence should satisfy in addressing these questions about essence. First, it ought to be *metaphysically adequate* in the sense that it explicates the notion of essence, at least in its most basic form, without indispensable appeal to metaphysical modality. In other words, any reasonable prior account of essence should provide a way of understanding what it is for a property to constitute the essence of an entity independently of what is necessary of the entity. For if the modal conception of essence is wrong for the reason Fine gives—no backward road from necessity to essence—then surely an alternative account should be put forth in its place that seeks to explicate the notion of essence in

non-modal terms. Indeed, any explication of the notion of essence that makes indispensable appeal to metaphysical modality would make the overall account entirely trifling and of no explanatory value, for the ultimate goal is to explain metaphysical modality in terms of essence.

Second, a prior account of essence ought to be *explanatorily adequate* in the sense that it offers an account of in what sense and how necessity arises from essence. This means that any reasonable prior account of essence should at least account for the truth of the bridge principles. This is of course hardly required under the modal conception of essence, according to which both (BPP) and (BPO) count as conceptual truths. However, both principles should be regarded as substantive theses under the essentialist approach because the notion of essence is supposed to be understood in non-model terms.

Finally, it should also be *epistemologically adequate* in the sense that the epistemological account of how we may come to know facts about essence should not make appeal to prior knowledge of necessary facts. Recall that according to the essentialist account, our knowledge of necessity is supposed to derive from knowledge of essence. So, for example, our knowledge of (1) should not be grounded on our prior knowledge of what is necessary of water. For the latter is supposed to follow from the former, but not the other way around.

Let me make a couple of comments on these constraints. Note that there is nothing in the first two constraints that requires that the essentialist approach should provide a *reductive* account of metaphysical modality in terms of essence understood as a non-modal notion. Admittedly, one may reasonably think that the goal of the essentialist approach is not to offer such a reduction, but rather to exhibit "the class of necessities as structured in a certain way, by identifying those necessities as basic or fundamental, and the rest as dependent, inheriting their necessity, ultimately, from necessities in the base class" (Hale, 2013, pp.157-158). But even so, as Leech (2021, p.890-891) points out, it should still be explained what kind of truths those basic truths are and why they should be considered metaphysically necessary in the first place. And this is exactly what the metaphysical and explanatory adequacy constraints require. Hence these constraints ought to be met by any prior account of essence, regardless of whether the essentialist approach has the ambition of reducing the modal to a non-modal.

Second, one might think that we need an additional epistemological constraint that any prior account of essence should also explain how we obtain knowledge of the bridge principles.[1] We could of course add this extra constraint. However, I am inclined to think that no such additional constraint is strictly required. For, as Hale (2013, p.269) notes, if a compelling case can be made for the truth of bridge principles on the basis of a non-modal notion of essence, then it can *eo ipso* serve as the basis of an explanation of how they may in principle be known. So, we shall require no additional epistemological constraint.

4. The integration challenge

In the next two chapters, I shall examine whether the three constraints are satisfied by the explanation-based and definition-based accounts of essence, respectively. I shall argue that the explanation-based account fails at the explanatory adequacy constraint (chapter 6), and that the definition-based account fails at the epistemological adequacy constraint (chapter 7). Hence neither accounts addresses the prior questions about essence in a way that is adequate for the purpose of giving a metaphysical and epistemological account of necessity in terms of essence.

Here it may help to see the matter in terms of what Christopher Peacocke (1999, p.1) dubbed the *integration challenge*:

> In a number of diverse areas of philosophy, we face a common problem. The problem is one of reconciliation. We have to reconcile a plausible account of what is involved in the truth of statements of a given kind with a credible account of how we can know those statements, when we do know them. ... I call the general task of providing, for a given area, a simultaneously acceptable metaphysics and epistemology, and showing them to be so, the Integration Challenge for that area.

So, in particular, the integration challenge for the truths about essence is to provide a unitary metaphysical and epistemological account of them.[2] We understand "unitary" in the weakest possible sense that a metaphysics of essence should not preclude a credible epistemology of essence, and

[1] For example, see Horvath (2014) for a criticism of Lowe's epistemological account of essence that his account is incomplete because it does not give an account of how the bridge principles can be known.

[2] As Peacocke himself explicitly acknowledges, the integration challenge is a generalization of the problem that Paul Benacerraf (1973) raises with respect to mathematical truths. See Roca-Royes (2021) for a survey of the problem.

vice versa. Recall here that the ultimate goal of the essentialist approach is to give a metaphysical and epistemological account of necessity in terms of essence. So, we may formulate the integration challenge for the essentialist approach thus: to reconcile a plausible metaphysics of essence with a reasonable epistemology in a way that is adequate for the purpose of providing a metaphysical and epistemological account of necessity in terms of essence.

The integration challenge, I think, creates a serious problem for the essentialist approach; and this will be the upshot of the discussion in the next two chapters. Chapter 6 argues that the explanation-based account of essence is not adequate for explaining necessity as having its metaphysical source in essence. And chapter 7 argues that the definition-based account does not offer any credible epistemological account of essence. It thus seems that neither account can meets the integration challenge; it fails either at giving an explanation of necessity as having its metaphysical source in essence or at providing a credible epistemology of essence. The conclusion will thus be that the essentialist approach is up against a stiff challenge.

CHAPTER 6

Critique of Essentialism—On the explanation-based account of essence

1. Introduction

In this chapter, I shall be mainly concerned with the explanation-based account of essence. For definiteness, I shall concentrate on a version developed by Marion Godman, Antonella Mallozzi and David Papineau (henceforth, simply, GMP) in their recent paper "Essential Properties are Super-Explanatory: Taming Metaphysical Modality" (2020). According to their account, the essence of a kind consists in the *super-explanatory property*—a single property that is causally responsible for a multitude of commonalities shared by the instances of the kind. And they argue that this super-explanatory account of essence offers a principled account of aposteriori necessities concerning kinds as consequences of truths about super-explanatory properties. Below I am going to argue that their account fails to satisfy the explanatory adequacy constraint. I shall examine two main arguments of GMP that the super-explanatory property of a kind is metaphysically necessary and argue that they both are fallacious. Along the way, a general problem will emerge that applies to any account that tries to explicate the notion of essence in terms of an explanatory relation.

2. The super-explanatory account of essence

Let me begin by giving a brief review of GMP's super-explanatory account of essence. They begin by observing that a certain category K of things exhibit multiple commonalities that admit of "informative synthetic generalizations of the form: *All K are G*" (Godman et al., 2020, p.317). For example, all samples of (pure) water exhibit many shared features, such as transparency, the same electric conductivity and the same liquid density, etc. Similarly, all horses are alike in their appearance, diet, reproductive behavior, etc. In cases like these, we may say that a category of things share a set of *correlated* properties in the sense that they are co-instantiated more often than we would expect given their separate probabilities of occurrence (ibid, p.319). According to GMP,

Kinds are any categories of thing whose instances share many correlated properties. The correlated properties of a Kind may be called its *characteristic features* (ibid, p.318).

Here one might wonder whether GMP's account really corresponds to the ordinary, intuitive notion of a kind. But such an issue is of little significance for our purpose. So, we may ignore it and simply use the term 'Kind' (with capital 'K') for GMP's technical notion. It should be noted that not every property picks out a Kind in this technical sense of the term. Consider, for example, the property of being a quadrilateral. It does not seem to admit of any "informative synthetic" generalizations of the form: *All quadrilaterals are G*. So, according to GMP, the property of being a quadrilateral do not pick out a Kind (ibid, p.317). Here let me briefly note that the property of being a quadrilateral still admits of some generalizations, such as "all quadrilaterals are four-sided polygons." This might seem to be utterly trivial. It may nevertheless be said to provide an "explanation" of many shared features of quadrilaterals; for example, it can be used to prove that the interior angles of any quadrilateral add up to 360 degrees. Proofs like this are not of course *causal* explanations; but they still provide a kind of explanation of the properties of quadrilaterals. The relevance of this example to the current discussion will become clearer later.

The characteristic features of a Kind are sometimes discovered to have a common cause. In such a case, the Kind is said to be *genuine* and the common cause to be its *super-explanatory* property (ibid, p.6). For example, water is a genuine Kind because its aforementioned characteristic features are supposed to be explained by the fact that every molecule of water is constituted by two hydrogen and one oxygen atoms. This chemical composition of water may thus be said to be its super-explanatory property. When a Kind does not admit of such a unified causal explanation, it is said to be *ersatz*; in other words, ersatz Kinds are ones that have no super-explanatory property. In what follows, we shall be concerned mainly with genuine kinds.

It should be noted that super-explanation is intended primarily as a *metaphysical* relation, but not an epistemic one. When a certain property E is said to be super-explanatory of a Kind K, it simply means that it is by virtue of having the property E that the instances of K have the characteristic properties, regardless of whether we have knowledge of such causal connections.

Now, GMP's main proposal is that the essence of a genuine Kind is constituted by its super-explanatory property, i.e., the property that is causally responsible for the characteristic features of the Kind. According to GMP, in other words, the super-explanatory properties are the source of metaphysical necessities concerning genuine Kinds. This idea is given a straightforward formulation in the following principle:

(BPK) *For any Kind K, any property E and any individual x, if E is super-explanatory with respect to K, then necessarily (x is K if and only if x is E).* (ibid, p.328)

It should be clear that this Bridge Principle for Kinds (BPK) is just another formulation of (BPP) in terms of super-explanatory property and kind. So, as suggested in the previous chapter, this principle offers a straightforward account of metaphysical necessities concerning Kinds. To repeat the previous example of water: Suppose that the characteristic features of water are explained by its chemical composition. Then it follows from (BPK) that it is necessary that every quantity of water is composed of H_2O molecules as follows:

> Being composed of H_2O molecules is super-explanatory of water.
> If being composed of H_2O molecules is super-explanatory of water, then necessarily every instance of water is composed of H_2O molecules.
>
> ---
>
> Necessarily, therefore, every instance of water is composed of H_2O molecules;

where the second premise is an instance of (BPK). In this way, according to GMP, metaphysical necessities concerning Kinds are consequences of facts about their super-explanatory properties taken together with the relevant instances of (BPK).

Now, the crucial question concerns whether the super-explanatory account of essence is adequate for the essentialist approach. That is, does it satisfy the three conditions for adequacy as explicated in the previous chapter? Below I shall try to show that the super-explanatory account of essence is not explanatorily adequate. A detailed examination of two main arguments of GMP that the super-explanatory property of a Kind is metaphysically necessary will reveal that they rest

on fallacious grounds. But it will be both instructive, and helpful for my purposes, to begin with some of the issues relating to the metaphysical adequacy condition.[1]

3. Metaphysical adequacy

The metaphysical adequacy condition states that a prior account of essence ought to explicate the notion of essence without indispensable appeal to metaphysical modality. One might think that this condition already creates a serious problem for the super-explanatory account of essence. To see what the problem is, notice first that what take the center stage in GMP's account are the causal relations that hold in the actual world (ibid, pp.329-330). For the super-explanatory property of a Kind K is basically one that is causally responsible for the characteristic features shared by the instances of K in the actual world. But aren't causal facts modal facts? For example, consider one billiard ball's hitting another. Under suitable background conditions, it is *necessary* that the second ball moves. It may thus seem that causal facts already involve metaphysical modality.

However, there seems to be a natural way to get around this problem. The modal force that is relevant to cases like above is that of *natural necessities*—the ones that hold in those possible worlds that have the same laws of nature as the actual world—but not metaphysical necessities proper—the ones that hold in every possible world. Here, of course, natural necessities are characterized by the notion of metaphysically possible world. But such a characterization is not strictly required. We can simply take the notion of natural necessity as primitive and give an account of causal truths as ones pertaining to a kind of explanatory relation that hold by virtue of the laws of nature.[2]

It may help to consider the matter in terms of counterfactual conditional. It is widely agreed that causal truths are closely tied to counterfactual conditionals in the sense that counterfactual conditionals often track causal relations, and conversely that causal truths can reliably be tracked by a set of appropriate counterfactual conditionals. The problem for GMP is that it is quite standard

[1] I shall not be concerned with the issue of whether GMP's account satisfies the epistemological adequacy constraint. I am inclined to think that it does. In this connection, see my brief remark in section 5 below.

[2] See Fine (2002) for the distinction between natural and metaphysical necessity and some of the difficulties in trying to reduce one in terms of the other. The influence of Fine's discussion on my view will become obvious in my critique of GMP's account in the next section.

to analyze counterfactual conditionals in terms of metaphysical modality; e.g., the Lewis-Stalnaker style of possible world semantics (Lewis, 1973; Stalnaker, 1968).

As Williamson (2007) argues, however, we can equally well reverse the order of analysis and characterize metaphysical necessity and possibility in terms of counterfactual conditionals. To see how this goes, let $\Box$ stand for the necessity operator, $\Diamond$ stand for the possibility operator, $\neg$ for the negation operator, $\bot$ for an absurdity and $\Box\!\!\rightarrow$ for the counterfactual operator (so that '$P \Box\!\!\rightarrow Q$' abbreviates 'if it were the case that P then it would be the case that Q'). Then we can have the following equivalence:

$$\Box P \equiv (\neg P \Box\!\!\rightarrow \bot).$$

That is, P is necessary just in case any counterfactual supposition to the contrary leads to an absurdity. By the duality between necessity and possibility, then, it also follows:

$$\Diamond P \equiv \neg(P \Box\!\!\rightarrow \bot).$$

That is, P is possible just in case it is not that P would obtain only with an absurdity.

Under this characterization, metaphysical necessity and possibility are understood as special cases of counterfactual truth. But counterfactual truth itself is taken as a primitive notion understood independently of metaphysical modality. Hence causal truth need not be regarded as involving metaphysical modality. For it can be comprehended in terms of its aforementioned relation to counterfactual truth. Given the tight connection between the causal and the counterfactual, indeed, it might even seem natural to give an account of metaphysical modality in terms of causal truth, rather than the other way around.[3]

[3]Here some might object that this way of meeting the metaphysical adequacy condition is hardly satisfactory. For counterfactual truth is itself a modal notion in some suitably broad sense of the term. Recall, however, that the metaphysical adequacy condition does not require that the notion of essence be reducible to a non-modal notion. Instead, it only requires that the notion of essence be explicated in terms of what can plausibly be understood independently the notion of metaphysical modality. And the point here is that it is not unreasonable to take counterfactual truth as primitive and understand metaphysical necessity and possibility just as special cases thereof.

4. Explanatory adequacy—the first argument

Let me now turn to the question of whether the super-explanatory account of essence satisfies the explanatory adequacy condition. In what sense and how necessities concerning Kinds arise from the super-explanatory properties? On GMP's view, recall, the core essentialist idea that necessity has its source in essence is given a straightforward formulation in (BPK). This principle surely offers an easy explanation of necessities concerning Kinds as consequences of their super-explanatory properties. But the question is: why accept (BPK) in the first place? Why, for any Kind, should its super-explanatory property be metaphysically necessary? In other words, what goes wrong if we counterfactually suppose that a Kind lacks its super-explanatory property? In what follows, I will examine two main arguments of GMP that the super-explanatory property of a kind is metaphysically necessary and argue that they both are fallacious.

The first argument of GMP for the necessity of super-explanatory property can be found in the following passage:

> One important aspect of metaphysical modality is the way it structures counterfactual thinking. Many everyday concerns call for us to consider what would have happened under some counterfactual supposition. When we counterfactually suppose that some item lacks some property, we naturally hold most of its other actual properties fixed. But when we counterfactually suppose that some Kind lacks a super-explanatory property, we are prevented from holding most of its other properties fixed, given that counterfactually supposing away a cause typically requires us to suppose away its effects, too. So when we suppose away the super-explanatory core of some Kind, all bets are off. We are hypothesizing away all the many correlated properties that distinguish that Kind from others. The natural reaction is that we have thereby supposed away the Kind altogether. (Godman et al. 2020, pp.12-13)

By way of illustration, consider a counterfactual scenario where water is not composed of hydrogen and oxygen atoms in the ratio of 2:1. GMP suggest that one has to think that in the imagined scenario, water also lacks its characteristic features, such as its density, electrical conductivity, etc.

For, when supposing a cause away, one should suppose away its effects also. In the counterfactual situation, therefore, water ought to be considered as lacking all the features that distinguish it from other stuffs. But then it is natural to say that water does not exist at all in the counterfactual situation. So, the imagined scenario is not where water is not composed of hydrogen and oxygen atoms in the ratio of 2:1. It is rather a situation where there is another chemical substance with a different chemical composition. Hence one cannot be said to reason counterfactually about water without holding its chemical composition fixed. In any counterfactual situation where water exists, therefore, water must be composed of H_2O; that is, it is necessary that water is composed of hydrogen and oxygen atoms in the ratio of 2:1.

Let us be more precise about the structure of the argument. It should already be obvious that the key premises of this argument are:

(P1) that counterfactually supposing away a cause requires to suppose away its effects too; and

(P2) that a Kind is distinguished from others by its characteristic features.

Let K be a Kind, C be its characteristic features and E be its super-explanatory property. Since E is super-explanatory of K, it is immediate that E is the common cause of C. Then it follows by (P1), according to GMP, that in any counterfactual situation where K lacks E, K also lacks C. This in turn implies by (P2) that nothing would count as K in any counterfactual situation where K lacks E. Hence it is incoherent to suppose counterfactually that K exists without E.

Let us then examine (P1) and (P2). I shall begin with (P2). (P2) may at first sight appear plausible. It should be noted, however, that there is a considerable ambiguity in (P2). So far as I can see, it has at least four readings. To formulate them in the possible world terminology, we have: for any Kind K_1 and K_2,

(R1) $K_1 = K_2$ only if the set C_1 of characteristic features that K_1 has in the actual world is identical to the set C_2 of characteristic features that K_2 has in the actual world.

(Actual world reading)

(R2) $K_1 = K_2$ only if, for any possible world w, the set C_1 of characteristic features that K_1 has in w is identical to the set C_2 of characteristic features that K_2 has w.

(Intra-world reading)

(R3) $K_1 = K_2$ only if, for any possible world w, the set C_1 of characteristic features that K_1 has in the actual world is identical to the set C_2 of characteristic features that K_2 has w.

(Cross-world reading)

(R4) $K_1 = K_2$ only if, for any possible worlds w_1 and w_2, the set C_1 of characteristic features that K_1 has in w_1 is identical to the set C_2 of characteristic features that K_2 has w_2.

(Generalized cross-world reading)

Notice that each implies the previous under the natural assumption that the actual world is a possible world; for instance, (R1) can be obtained from (R2) by letting w be the actual world.

Now, the first two readings are unobjectionable. Note that what (R2) says is simply that any two Kinds are identical only if they have the same characteristic features *within each possible world* (and in the case of (R1), *within the actual world*). In fact, it can be proved in any standard modal logic using Leibniz's Law (the indiscernability of identicals) and hence be regarded a logical truth. To see how the proof goes, let's write $\Phi(K,C)$ to mean K has C as its characteristic features. Leibniz's law states that for any two entities x and y, if x is identical to y then if x has a certain property P then y also has P; in symbol:

$$(x)(y)(x = y \supset (Px \supset Py)). \tag{1}$$

As a substitution instance of (1), then, we have:

$$K_1 = K_2 \supset (\Box(C)(\Phi(K_1,C) \equiv \Phi(K_1,C)) \supset \Box(C)(\Phi(K_1,C) \equiv \Phi(K_2,C))), \tag{2}$$

Now, since

$$(C)(\Phi(K_1,C) \equiv \Phi(K_1,C)) \tag{3}$$

is a logical truth, by the necessitation rule, we have:

$$\Box(C)(\Phi(K_1,C) \equiv \Phi(K_1,C)). \tag{4}$$

From (2) and (4), then, it follows:

$$K_1 = K_2 \supset \Box(C)(\Phi(K_1,C) \equiv \Phi(K_2,C))), \tag{5}$$

which states the desired conclusion that two identical Kinds share exactly the same characteristic features *within any possible world*.

Here the problem is of course that GMP's argument above is not valid under the first two readings. For its conclusion concerns whether something would count as K *in a counterfactual situation* where it lacks C, namely the *actual* characteristic features of K. For this conclusion, (P2) should be given a cross-world reading, i.e., either (R3) or (R4). Since (R4) implies (R3), let us restrict our attention to (R3). What it basically states is that Kinds are individuated by their *actual* characteristic features *across different possible worlds*. But this appears to be highly dubious. For why do we have to think, for example, water must necessarily be transparent? In fact, GMP themselves believe that the characteristic features of a Kind are generally contingent. They say:

> [N]ote that it is a consequence of our view that modal Kind essences would stay fixed even in possible worlds where the relevant laws of nature were different. Even if variation in laws of nature meant that H_2O were no longer odorless, colorless, tasteless and so on, it would still be *water*.
>
> (Godman et al., 2020, p.330; emphasis original)

That is, the identity of a Kind is determined in the strict sense by its super-explanatory property, but not by its characteristic features. Hence the initial plausibility of (P2) immediately vanishes once its ambiguity is removed and understood in a proper way.

Let me now turn to (P1). It may be admitted that there is a tight connection between causal and counterfactual dependences in the sense specified above. To be a bit more precise, it seems plausible to think that *if A is a cause of B then if A had not occurred then (ceteris paribus) B would not have occurred either*; in symbol:

$$(A \text{ causes } B) \supset (\neg A \mathrel{\Box\!\!\rightarrow} \neg B).$$

In this sense, as GMP urges, (P1) may be said to be true enough for most purposes. So, one may reasonably agree with GMP that given that E is super-explanatory of K, we have the following counterfactual conditional:

(I) If K had not had E, then K would not have had C also.

So far, so good. But the problem comes in at the very next step. For what GMP infer from this counterfactual conditional is this:

(II) In any counterfactual situation, if K does not have E then K does not have C also.

In order to infer (II) from (I), we need to assume in the most general case the following principle:

$$(P \mathrel{\Box\!\!\rightarrow} Q) \supset \Box(P \supset Q).$$

But this principle is clearly not valid. For a counterexample, consider the Oswald shooting. Presumably, had Oswald failed to shoot Kennedy on November 22nd, 1963, Kennedy would not have died that day. From this, however, it does not follow that it is necessary that if Oswald failed to shoot Kennedy on November 22nd, 1963, Kennedy did not die on that day. It thus seems that the argument is resting on a fallacious ground.

The main problem with (P1) is two-fold. First, the fact that A is a cause of B does not in general imply that A is the only possible cause of B. For it is often legitimate to posit an alternative cause in counterfactual reasoning. Imagine a counterfactual scenario where Oswald had failed to shoot Kennedy on November 22nd, 1963. We may further suppose counterfactually that Oswald had hired a highly effective back-up in case of his failure. On this scenario, Kennedy would still have died on that very date.

Second, even when A is the only possible cause of B, it does not follow that B is impossible without A. Here we should carefully distinguish two senses of possibility. As noted above, causal facts are dependent upon the laws of nature. In some cases, it may be implied by the relevant laws of nature that there is no way to bring about a certain course B of event other than by another event A. When A is said to be the only possible cause of B in this sense, the "possible" here is that of natural possibility. However, metaphysical possibility in the proper sense of the term is not restricted to natural possibility. We may coherently imagine a counterfactual scenario where the laws of nature are completely different from the actual ones. So, even if A is the only *naturally possible* cause of B, it does not follow that B is *metaphysically impossible* without A.

It may help to see the problem with (P1) from the perspective of the standard formal semantics for counterfactual and modal operators. In the Lewis-Stalnaker style semantics for counterfactual

conditionals, $(P \mathrel{\Box\!\!\rightarrow} Q)$ is true if and only if all P-worlds that are *most similar* to the actual world are also Q-worlds (according to a given similarity metric). In contrast, according to the standard semantics for the modal operator, $\Box(P \supset Q)$ is true at w just in case *all* P-worlds—whether similar or not—are also Q-worlds. This is already sufficient to see that the inference from $(P \mathrel{\Box\!\!\rightarrow} Q)$ to $\Box(P \supset Q)$ is not valid. For, in general, some P-worlds may not count as most similar to w.

When A causes B, this holds by virtue of some suitable background conditions and the laws of nature. Let us call them the *basis of causal explanation*. When we reason about a counterfactual situation where A fails to obtain, it may indeed be highly natural to hold the basis fixed and to consider what would happen in that situation; and in this sense, we may be said to consider only $\neg A$-worlds that are "most similar" to the actual world. And given that A causes B, it may seem also natural to think that B should also fail to hold in such worlds. In this regard, the principle $(A \text{ causes } B) \supset (\neg A \mathrel{\Box\!\!\rightarrow} \neg B)$ seems to be true enough for most practical and theoretical purposes.[4] However, this does not in general imply that B fails in *all* possible worlds where A fails. For, again, there may be $\neg A$-worlds where some (or all) of the basis of causal explanation fail to hold; such worlds would not count as sufficiently similar to the actual world and hence be excluded in evaluating the truth of the counterfactual conditional. Hence the invalidity of the inference from $(\neg A \mathrel{\Box\!\!\rightarrow} \neg B)$ to $\Box(\neg A \supset \neg B)$.

In fact, this line of argument can be generalized to any explanatory relation. For an explanatory relation, whether causal or non-causal, holds by virtue of a set of background conditions and some general laws, which we may call the *explanatory basis*. Unless the explanatory basis is shown to be metaphysically necessary, that X is explanatory of Y does not imply that Y is impossible without X. What it at best shows is that X entails Y in every world where the explanatory basis holds. Hence any account that tries to explicate the notion of essence in terms of an explanatory relation faces the challenge of showing the necessity of the explanatory basis itself. But there is an obvious problem here. It would be entirely trifling to show that an explanatory basis is metaphysically necessary on the basis of that very explanatory basis. To show the necessity of any given explanatory basis, we need another explanatory basis which is arguably more fundamental than the first. This results in

[4] In making this remark, I do not intend to endorse the simple counterfactual analysis of causation. The point is rather that I am willing to grant this much to GMP if only for the sake of argument.

a chain of explanation. The chain may come to an end; I see no compelling reason to suppose that the regress will be infinite. The problem, as I see it, is rather that the original explanatory relation, in terms of which we try to give an account of essence, is never by itself adequate for the purpose of explaining necessity.

5. Explanatory adequacy—the second argument

Now let me turn to the second argument of GMP that the super-explanatory of a Kind is metaphysically necessary:

> It has been put to us that the modal consequences of [(BPK)] follow directly from the necessity of identity, and therefore that super-explanatoriness is not needed to explain them. The thought is that such identities as *water*$=H_2O$, and *tigerhood*=*being descended from the original tigers*, and so on, together with the necessity of identity, are themselves enough to ensure the necessary coextensiveness specified in the consequent of [(BPK)], without any help from the super-explanatoriness referred to in the antecedents.
>
> We agree that the modal consequences of [(BPK)] follow from Kind identities like *water* $= H_2O$, and *tigerhood* $=$ *being descended from the original tigers*, together with the necessity of identity. But we think that such Kind identities are themselves a consequence of [(BPK)]. We identify water with H_2O specifically because this chemical composition is super-explanatory of the many shared properties of water. Because of this super-explanatoriness, we regard water and H_2O as necessarily coextensive, and this then leads us to identify them.

Here the objection to GMP's account is that the identity of water to H_2O is sufficient to derive their necessary coextensiveness; and hence that there is no need to appeal to super-explanatoriness. In replying to this objection, GMP argues that the identity itself is a consequence of (BPK) and hence that their necessary coextensiveness should ultimately be explained by super-explanatoriness. The argument runs as follows. Being composed of H_2O molecules is super-explanatory of water.

By (BPK), then, it follows that water is necessarily coextensive with H_2O. From this necessary coextensiveness it follows that water is identical to H_2O.

There seem to be two problems with this argument. First, recall that our main question concerns why (BPK) is true. To this question, this argument is no answer. For it makes use of the very principle in question. Second, it can be agreed that that two identical properties are necessarily coextensive. So, the identity of water to H_2O implies that they are necessarily coextensive; in symbol:

$$(\text{water} = H_2O) \supset (x)\Box(x \text{ is water } \equiv x \text{ is } H_2O),$$

which I take to be a logical truth. Note, however, that the last step of GMP's argument appeals to the converse principle, namely that any two properties, if necessarily coextensive, are identical. But this coarse-grained view about property identity seems to be at odd with their overall essentialist approach to necessity. To see this, let us go back to one of Fine's counterexamples to the modal conception of essence: the property of being a set whose sole member is Socrates and the property of being a set whose sole member is Socrates and such that $2+2=4$. While the two properties are necessarily equivalent, it is only the former property that may reasonably be said to constitute the essence of $\{\text{Socrates}\}$. One might think that this line of argument naturally goes hand-in-hand with the hyper-intensional conception of properties. For it seems to suggest that the two properties should be considered distinct despite the fact that they are necessarily coextensive. And basically the same consideration applies to GMP's super-explanatory account of essence as well. Consider the property of being composed of H_2O molecules and the property of being composed of H_2O molecules and being such that $2+2=4$. Here again, the two properties are necessarily coextensive, but it is only the former that can be said to constitute the essence of water. For, certainly, the mathematical fact that $2+2=4$ does not seem to be causally responsible for the characteristic features of water. It may thus seem that the coarse-grained view about property identity does not fit well with GMP's account.[5]

[5]Here I do not intend to claim that the examples like the ones above show that the essentialist approach is *logically incompatible* with the coarse-grained view. For, as Rosen (2015, pp.204-205) notes, one may hold the coarse-grained view and still coherently claim that the essentialist attribution of properties create hyper-intensional contexts. But he is also in agreement that the coarse-grained view is in some tension with the essentialist approach (ibid, pp.203-204).

How might GMP address these worries? They might say that their view is simply that the identity between water and H_2O is grounded in the fact that the latter is super-explanatory of the former. Indeed, something along this line is all that GMP needs to answer the objection without making use of (BPK). For, once this much is granted, then they can just appeal to the necessary coextensiveness of identical properties to secure the truth of (BPK).

> If property E is super-explanatory of property K then property E is identical to property K.
> If property E is identical to property K, then $(x)\Box(Ex \equiv Kx)$.
>
> ---
>
> Therefore, if property E is super-explanatory of property K then $(x)\Box(Ex \equiv Kx)$.

If this line of argument is what they have in mind, then it seems to me that there is a confusion between epistemic and metaphysical considerations here. The fact that the characteristic features of water are all causally explained by its chemical composition provides a compelling abductive ground for identifying water to H_2O. In this sense, it is certainly true that we identify water to H_2O on the ground of super-explanatoriness. But this is clearly an epistemic consideration. In the metaphysical order of the matter, I am inclined to think that the opposite is true. That is, when a certain property E is super-explanatory of a Kind K, this holds (in part) because the two are identical as properties. For example, it is *because* water is identical to H_2O—i.e., the chemical substance composed of hydrogen and oxygen in the ratio of 2:1—that many properties of water are explained by its chemical composition. In cases like this, some set of properties of an entity is explained by a more fundamental property of *that very entity*. It is indeed very hard to see how this explanation is supposed to work without the identity relation. As opposed to GMP, the correct metaphysical picture seems to me that super-explanatoriness is grounded in the identity relation.

This picture also affords a simple account of why GMP's account seem to fit "well with a range of common intuitions about which properties are metaphysically necessary to Kinds" (Godman et al., 2020, p.327). On this picture, super-explanatory facts hold by virtue of underlying identities between relevant Kinds and properties. These identities, when conjoined by the principle of necessary coextensiveness of identical properties, imply necessary truths about Kinds.

Small wonder, then, that super-explanatoriness provides a reliable epistemic guidance in discovering identities between properties and aligns well with common intuitions about which properties are metaphysically necessary to Kinds.

Moreover, this alternative picture can easily generalize to many other properties to which the super-explanatory account is not applicable. To see this point, let us consider the case of quadrilaterals again. We may say that the property of being a quadrilateral is identical to the property of being a four-sided polygon. For to be a quadrilateral is to be a four-sided polygon. And on the basis of this identity, we can prove—and in this sense explain—many properties of quadrilaterals. Given this, it is hard to see why, unlike other properties, the identities of Kinds should be grounded in a special form of causal explanation.

6. Summary

To sum up the whole discussion: GMP's super-explanatory account of Kind essence states that the essence of a Kind consists in the super-explanatory property, i.e., the common cause to its characteristic features. I have argued that this account fails to satisfy the explanatory adequacy condition. To show this, I have examined two main arguments of GMP that metaphysical necessities concerning Kinds are consequences of truths about super-explanatory properties and argued that they both rest on fallacious grounds. GMP's first argument is that given that the characteristic features of a Kind K are all causally explained by its super-explanatory property, in any counterfactual situation where it lacks its super-explanatory property, virtually none of its characteristic features would be left to distinguish it from others; so, K could not possibly exist without its super-explanatory property. I have argued that this argument is problematic on two scores. First, it relies on the dubious premise that Kinds can be identified by their characteristic features across different possible worlds. Second, the argument involves a fallacious inference from a causal truth of the form "A causes B" to a conclusion of the form "In any counterfactual situation where A does not occur, B does not occur either" (in symbol, $\Box(\neg A \supset \neg B)$). It may be admitted that a causal truth of this form implies a counterfactual truth of the form "Had A not occurred then B would not have occurred either" (in symbol, $(\neg A \mathrel{\Box\!\!\rightarrow} \neg B)$). This counterfactual statement, however, should

not be confused with the desired conclusion. For the truth-value of the counterfactual statement $(\neg A \mathrel{\Box\!\!\rightarrow} \neg B)$ is supposed to be evaluated in reference only to those possible worlds that are sufficiently similar to the actual world, while the desired conclusion $\Box(\neg A \supset \neg B)$ concerns all possible worlds, whether similar or not. In the case of causal truths, in particular, their truths-values are evaluated in reference to those possible worlds that share the same laws of nature as the actual world. We have also noted that like remarks apply more generally to other types of explanatory relations, for they also depend on a set of relevant background conditions and general principles. The second argument of GMP is that the identity of a Kind is determined by its super-explanatory property. I have argued that there is a confusion between metaphysical and epistemological considerations. Truths about super-explanation provide compelling epistemic grounds for identifying Kinds with their super-explanatory properties. But this does not mean that the identities are grounded upon super-explanatoriness in the metaphysical sense. The correct picture seems to be the other way around: it is because a Kind is identical to a certain property that the latter can uniformly explain the characteristic features of the former.

CHAPTER 7

Critique of Essentialism—On the definition-based account of essence

1. Introduction

The definition-based account maintains that the essence of an entity is what defines the identity of the entity. This account has recently been developed extensively in the works, most notably, of Fine (1994), Hale (2002, 2013) and Lowe (2012). It is the aim of this chapter to examine whether the definition-based account satisfies the metaphysical, explanatory and epistemological adequacy constraints. I shall argue that it satisfies the first two but not the third. In general, an entity admits of a number of different but suitably fundamental identifications. However, not all such identifications may count as definitive of the entity because of the hyper-intensional character of essence. So, there naturally arises what I call *the question of essentialist knowledge*: how do we know which of the identifications is definitive of the entity? I shall examine some of the well-known definition-based epistemologies of essence and argue that they all fail to address this question. The upshot will thus be that the question of essentialist knowledge creates a serious epistemological problem for the definition-based account. In the course of argument, I shall also discuss a recent objection that the definition-based account is not explanatorily adequate; and it will be seen that the objection, though not entirely groundless, is not decisive.

2. The Definition-based account of essence

The definition-based account explicates the notion of essence on the model of definition. It is often thought that there are two kinds of definition: real and nominal definitions. A case of *real* definition is when we define an entity by stating what it is to be that very entity. A *nominal* definition, on the other hand, defines the meaning of a word for an entity. On the definition-based account, the relevant notion of definition is the real one; that is, the essence of an entity is specified

by its definition in the real sense (Fine, 1994; Lowe, 2012; Hale, 2013); in other words, the essence of an entity is definitive of what it is to be that very entity.

Now, what exactly is it to give the real definition of an entity? Certainly, not every identification of an entity may count as its real definition. Consider, for example, the identification of Aristotle as the teacher of Alexander the Great and the identification of water as the transparent liquid that covers around 70% of the surface of the Earth. Though these are uniquely true respectively of water and of Aristotle, they can hardly be said to define the identities of the corresponding entities in the required sense because they are plainly contingent. Nor is it sufficient for an identification of an entity to count as a real definition that it picks out the same entity in every possible world. For the real definition of an entity is supposed to specify the essence of the entity, which as noted in chapter 5 is hyper-intensional. Assuming that mathematical truths are necessary, for example, $\{\text{Socrates}\}$ may be identified in every possible world as a set whose sole member is Socrates and such that $2+2=4$. Certainly, however, it should not be part of the essence of $\{\text{Socrates}\}$ that $2+2=4$. Hence this identification should not count as definitive of $\{\text{Socrates}\}$. Now, recall here that the definition-based account of essence should be metaphysically adequate in the sense that it explicates the notion of essence without appeal to metaphysical modality. Hence it ought to provide a non-modal criterion for an identification to count as a real definition that would rule out ones like above as non-definitional.

How can this be done? One way is to require that the real definition of an entity should state what the entity *fundamentally* is (Lowe, 2011, p.4), or what is *constitutive* of the identity of the entity (Fine, 1994, p.3 and pp.8-9; see also his 2015, pp. 306-308). It is difficult to say in more precise terms what this qualification amounts to. However, its intended interpretation can be made reasonably clear in terms of what form an adequate definition should have. E. J. Lowe (2011, p.4) says:

> ... [E]ssence is revealed by its so-called *real definition* ... For instance, *man* as a kind of substance is famously defined, in the Aristotelian tradition, as *a rational animal*. This sort of definition is said to be *per genus et differentiam*: that is, a species of substance is defined in terms of the higher *genus* to which it belongs and

> the *specific difference* which renders it distinct from any other species of that same genus. (Emphasis original)

So, according to Lowe, a real definition should identify an entity by specifying what kind of entity it is at a sufficiently basic level and how it is distinguished from other entities of the same category. Water, for example, may be defined as a chemical substance—the kind of entity it is—and in terms of its chemical formula H_2O—how it is distinguished from other entities of the same kind; and similarly, Aristotle may be defined as a human being—the kind of entity it is—with such-and-such biological origin—how it is distinguished from other entities of the same kind.

This conception of real definition, though admittedly not completely precise, seems to provide an intuitive and practical understanding of the notion of real definition that would rule out identifications like the ones above as non-definitional. The identification of Aristotle as the teacher of Alexander the Great, for example, should not count as his real definition because being a teacher does not seem to represent a basic category to which Aristotle belongs. Water may be identified as a liquid at a sufficiently basic level, but it does not seem to be part of how different kinds of liquid are supposed to be individuated how many percentages of the Earth surface they cover. Similarly, it should not count as definitive of what it is to be $\{\text{Socrates}\}$ that it is a set whose sole member is Socrates and which is such that $2+2=4$. For it is not in general part of how sets are supposed to be individuated that $2+2=4$. Hence these identifications should not count as definitional.

It is important to note here that in spelling out the notion of real definition along this line, no appeal to metaphysical modality has been made. It thus seems to be possible to articulate, in non-modal terms, a notion of real definition that provides a reasonably clear criterion for distinguishing definitional identifications from non-definitional ones. Here it may still be complained that the notion of real definition has not yet been given a proper analysis. However, I think that the explication given above is sufficient for an intuitive and practical grasp of the notion. I am also inclined to think that it is not really fair to press one too far on this, considering that the notion of real definition may reasonably be considered primitive in the definition-based account. In what follows, therefore, I shall assume, if only for the sake of argument, that the definition-based account satisfies the metaphysical adequacy constraint.

3. The explanatory adequacy constraint

Let us then turn to the explanatory adequacy constraint. The question now is: does the definition-based account provide an adequate account of the truth of the bridge principles? In other words, why according to the definition-based account should essential properties of an entity also be necessary properties? There seems to be an obvious worry here. The real definition of an entity is supposed to state what the entity *is* in some fundamental sense. But then why should the essence of an entity, as specified by its real definition, be considered *necessary*? This worry has recently been expressed by many philosophers, such as Casullo (2020), Leech (2021), Mackie (2021), Noonan (2018) and Romero (2019). For definiteness, I shall below concentrate on Mackie's critique.

Mackie argues that there are cases where it is coherent to think that the real definition of an entity does not represent its necessary property. In some cases, in other words, it is coherent to hold of an entity X both

(i) that it is definitional of X that P, and
(ii) that it is not necessary of X that P,

where 'P' is to be replaced by an appropriate statement. But this cannot be, if the definition-based account of essence licenses the truth of the bridge principles. Mackie thus concludes that the definition-based account of essence is not explanatorily adequate.

Now, Mackie discusses two main cases for her claim. The first case concerns Lockean real essences of natural kinds. The Lockean real essence of a natural kind consists in its internal constitution by which its discoverable qualities are supposed to be explained (ibid, p.255). For example, the Lockean real essence of a chemical substance—e.g., water—may be said to consist in its chemical composition—e.g., being H_2O. For the latter is supposed to explain many basic properties of the former, such as its density, electric conductivity, transparency, etc.

Given the explanatory power, Mackie thinks, the Lockean real essence of a natural kind may reasonably be regarded as definitive of the kind—i.e., representative of what the kind fundamentally is. But this need not be taken to imply, according to Mackie, that the Lockean real essence of

a natural kind constitutes its necessary property. Consider the case of water by way of illustration. As noted above, one may reasonably hold:

(Wi) that it is definitive of (and hence essential to) water that its samples are composed of H_2O molecules

in the light of the fact that the characteristic features of water are explained by its molecular structure. As we have seen in chapter 6, however, the explanation depends crucially on the physical and chemical laws that hold in the actual world. But it seems that we can legitimately suppose a possible world whose laws of nature are completely different from ours. In such a possible world, it might be the case that the characteristic features of water are given an alternative explanation, say, in terms of a different molecular structure, XYZ. In that possible world, one may think, water should be identified with XYZ but not with H_2O. It thus seems to be coherent to hold, in conjunction with (Wi),

(Wii) that it is not necessary of water that its samples are composed of H_2O molecules,

thereby contradicting (BPP). Here it should be noted that Mackie is *not* claiming that the case of water is a counterexample to (BPP) by arguing directly for the truth of (Wi) and (Wii). Her main point is rather that it is coherent to maintain both (Wi) and (Wii), and hence to think that what is definitive of a property need not represent what is necessary of it. It thus seems that the definition-based account fails to provide an account of the truth of (BPP).

Here one might object that Mackie's argument for the compatibility of (Wi) and (Wii) is mistaken. The key assumption in the argument is that in a possible world where the characteristic features of water are had by another substance, water should be identified with that substance; and so, it is possible that water has another chemical composition. But this just amounts to assuming that it is not fundamental to water that it is composed of H_2O molecules, in which case (Wi)—that it is definitive of water that it is H_2O—should be rejected. Hence it is simply inconsistent to hold (Wi) and (Wii) simultaneously.[1]

The problem with the objection, Mackie responds (2021, p.257), is that it makes appeal to what is necessary of an entity to determine what is definitive of the entity. In other words, the

[1]See, for example, Lowe (2011, pp.14-15).

objection seems to assume that whatever is not necessary is not definitive—in particular, that if it is not necessary that the samples of water are composed of H_2O molecules then it is not definitive of water that its samples are so composed. However, Mackie argues, no such appeal to necessity can legitimately be made in determining what is definitive of an entity due to the metaphysical adequacy constraint. So, this case seems to raise a challenge for the definition-based account: to debunk the apparent consistency of (Wi) and (Wii) without building necessity into the notion of real definition.

Now, it seems to me that there is a simple solution of the problem. In fact, I have already discuss this in chapter 4. Just to briefly repeat the solution, the idea is to take the real definition of a property as stating what the entity is in the sense of *specificative identification*. So, when it is said to be definitive of water that its samples are composed of H_2O molecules, this should be understood as expressing the specificative identification of the property of being water with the property of being composed of H_2O molecules. As we have seen, then, the identification logically implies that:

(Wiii) it is necessary of water that its samples are composed of H_2O molecules.

Hence it is indeed *logically* inconsistent to hold both (Wi) and (Wii).

More generally, we may state the proposed solution to Mackie's problem as follows. In its real definition, a property Φ should be identified specificatively, or at least in a way that implies a specificative identification of the form "the property of being Φ is identical to the property of being Ψ." Then it can logically be shown that the definition implies that the two properties are necessarily coextensive. This way, the definition-based account can offer a straightforward account of the truth of (BPP).

Let us then turn to Mackie's second case. This case concerns what she calls *substance sortals*. The defining feature of a substance sortal is that it is permanent in the sense that if an individual falls under a sortal S then it falls under S at any time of its existence. For example, consider the sortal *human being*. It may appear that no human beings could possibly cease to be a human being without going out of existences; like considerations seem to apply to sortals such as *horse*, *cow*,

etc.[2] Now, it seems reasonable to think that a substance sortal often represents a property that is fundamental to the entities that fall under it. For example, it appear to be fundamental to individual human beings that they are a human being; and similarly for horses and cows. Mackie thus argues that substances sortals may often be regarded as definitive of the entities that fall under them.

Now, does this imply that substance sortals also represent necessary properties of the entities that fall under them? Mackie argues that it does not:

> It is important to note, however, that it does not follow, merely from the definition of a substance sortal, that substance sortals are necessary sortals. The notion of a substance sortal is indeed a modal notion. But the modality involved is not the modality *de re* that is involved in the concept of a necessary sortal, but a version of *de dicto* modality. The definition of a substance sortal implies that if a thing falls under a substance sortal S at some time in its existence in a possible world w, it falls under S throughout its existence in world w. The definition does not imply that if a thing falls under S in world w, it falls under S in every possible world in which it exists—that it could not have existed without falling under S. Moreover, this point—that there is a logical gap between being a "necessarily permanent" property and being a necessary property—is an uncontroversial one, and explicitly acknowledged by many writers on this topic. (ibid, pp.259-260)

To understand Mackie's objection, note first that the question of whether substances sortals represent necessary properties concerns de re modal claims of the form:

$$(x)(x \text{ is an } S \supset \Box(x \text{ is an } S)), \tag{A}$$

where 'S' stands for a substance sortal. For instance, given that the sortal *human being* is a substance sortal, does the fact that Socrates is a human being imply that he is necessarily a human being? Mackie's point is that the notion of a substance sortal implies only de dicto modal claims

[2]As noted by Priest (2021, p.1878), some might find it quite dubious that these are substance sortals, claiming that a scenario where a person turns into a frog is conceivable. Though I am somewhat sympathetic to this view (see my discussion of aposteriori necessities concerning objects in chapter 8 below), I shall simply assume that they are substance sortals for the purpose of explication of Mackie's objection.

of the form:

$$\Box(x)(x \text{ is an } S \supset x \text{ is an } S \text{ at any time of its existence}). \qquad \text{(B)}$$

So, in particular, that the sortal *human being* is a substance sortal implies only that in any possible world w, if Socrates is a human being in w then he is a human being at any time of his existence in w. From this, however, the desired conclusion that Socrates is necessarily a human being does not follow. What we can infer at best is that he is a human being at any time of his existence *in the actual world*. But this should be properly distinguished from the desired conclusion that Socrates is a human being *in every possible world*. More generally, the core of Mackie's objection is that there simply is no way of deriving de re modal claims of form (A) from de dicto ones of form (B).

Now, Mackie's objection is certainly valid *if* we restrict ourselves to the resources of the standard system of first-order modal logic. However, we can pull off a trick to get around the problem by moving up to *higher-order* logic. To see how this goes, let us consider the case of Socrates. The idea is to define *the property of being Socrates*, instead of Socrates himself, by using the machinery of the specificative identification of a property. For any unary predicate Φ, let $\lambda x(\Phi x)$ mean the property term "the property of being Φ." Then we define the property of being Socrates as follows:

$$(x = \text{Socrates}) \equiv_{\text{def}} (x \text{ is a human being having } O \text{ as its origin});$$

so, we have:

$$\lambda x(x = \text{Socrates}) \approx \lambda x(x \text{ is a human being having } O \text{ as its origin}).$$

By the the principle of necessary coextensiveness of identical properties, it follows that:

$$\Box(x)(x = \text{Socrates} \equiv x \text{ is a human being having } O \text{ as its origin}),$$

which implies by the necessity of identity that:

$$(x)(x = \text{Socrates} \supset \Box(x \text{ is a human being having } O \text{ as its origin}));$$

that is, Socrates is necessarily a human being having O as its origin.[3]

To generalize this trick, let us define the notion of a fundamental sortal thus: a substance sortal S is *fundamental* if and only if, for any object x that falls under S, the property of being x is identical to the property of being an S satisfying a certain condition associated with S that individuates x from the other S's. Symbolically: a sortal S is fundamental just in case, for any object x that falls under S,

$$\lambda y(y = x) \approx \lambda y(y \text{ is an } S \text{ such that } \Phi(y)),$$

[3]Here we should note that the last step of this argument involves a subtle issue about existence. To see this, notice that necessity should be interpreted weakly here in the sense that a statement about certain objects counts as necessary if it is true in every possible world in which the objects exist. So, for example, the conclusion of the argument should be interpreted as stating that Socrates is a human being having O as its origin in every possible world in which he exists. The standard way of formalizing this weak interpretation of necessity is to use free logic with the existence predicate $\mathfrak{E}(x)$ and to reformulate formulas of the form $\Box\Phi(a)$ into $\Box(\mathfrak{E}(a) \supset \Phi(a))$; see chapter 16 of Hughes and Cresswell (1996) for further technical details. Then we can formalize the above argument as follows. Let s be a constant standing for Socrates and $H(x)$ be a predicate meaning that x is a human being having O as its origin. From the definition of the property of being Socrates given above, we have:

$$\lambda x(x = s) \;\approx\; \lambda x H(x).$$

Then by the principle of necessary coextensiveness of identical properties, we have as our premise that

$$\Box(x)(x = s \;\equiv\; H(x)).$$

Now, in free logic, we have as an axiom:

$$(x)\Phi(x) \supset (\mathfrak{E}(y) \supset \Phi[y/x]),$$

where y is substitutable for x in Φ. Note that this is to restrict the classical rule of universal instantiation to the existing objects. By substituting $x = s \supset H(x)$ for $\Phi(x)$ and x for y, we obtain:

$$(x)(x = s \supset H(x)) \;\supset\; (\mathfrak{E}(x) \supset (x = s \supset H(x))).$$

By the necessitation rule, then, it follows:

$$\Box[(x)(x = s \supset H(x)) \;\supset\; (\mathfrak{E}(x) \supset (x = s \supset H(x)))].$$

By distributing $\Box$, we obtain:

$$\Box(x)(x = s \supset H(x)) \;\supset\; \Box(\mathfrak{E}(x) \supset (x = s \supset H(x))).$$

This together with the premise implies:

$$\Box(\mathfrak{E}(x) \supset (x = s \supset H(x))).$$

From this it follows by the basic rules of propositional modal logic:

$$\Box((\mathfrak{E}(x) \supset x = s) \;\supset\; (\mathfrak{E}(x) \supset H(x))).$$

By distributing $\Box$, again, we have:

$$\Box(\mathfrak{E}(x) \supset x = s) \supset \Box(\mathfrak{E}(x) \supset H(x)).$$

Since we have as a theorem:

$$x = s \;\supset\; \Box(\mathfrak{E}(x) \supset x = s),$$

it follows:

$$x = s \supset\; \Box\,(\mathfrak{E}(x) \supset H(x)).$$

Since no special assumption is made of x, finally, we obtain our desired conclusion:

$$(x)(x = s \supset\; \Box\,(\mathfrak{E}(x) \supset H(x)));$$

that is, Socrates is necessarily such that if he exists then he is a human being having O as his origin.

where $\Phi(y)$ stands for an individuation condition associated with S. For another example of a fundamental sortal, one might consider the notion of a set. Since sets are individuated by their members, for instance, we may define the property of being $\{\text{Socrates}\}$ thus:

$$\lambda x(x = \{\text{Socrates}\}) \approx \lambda x(x \text{ is a set whose sole member is Socrates}),$$

from which it can be inferred that

$$(x)(x = \{\text{Socrates}\} \supset \Box(x \text{ is a set whose sole member is Socrates})).$$

This way, it can easily be shown that if an object falls under a fundamental sortal then it does so necessarily.[4]

A couple of points are worth noting about the notion of a fundamental sortal. Notice first that it is required of any second-order identification of an object x of the form

$$\lambda y(y = x) \approx \lambda y(\Psi(y))$$

that $\Psi(y)$ be true uniquely of x. For, otherwise, it would imply that x is identical two or more distinct things, contradicting the logic of identity. This requirement is guaranteed to be met by any second-order identification of an object in terms of a fundamental sortal. To see this, let S be a fundamental sortal and x be an object that falls under S. Then it holds by the definition of a fundamental sortal that

$$\lambda y(y = x) \approx \lambda y(y \text{ is an } S \text{ such that } \Phi(y)).$$

Recall here that $\Phi(y)$ is supposed to individuate x among the objects that fall under S; and obviously, x is distinguished from all objects that do not fall under S by its being an S. Hence the condition on the right-hand side is true uniquely of x.

[4]Note that I do not mean to assert that the sortals *set* and *human being* are fundamental in the sense just defined. Here I am simply arguing on behalf of the definition-based account that there is a way to meet Mackie's challenge. With this said, however, I should also note that some of the ideas given here will be adapted later to provide my own account of necessity; see my discussion of the notions of human being and of set in chapter 8 below.

In this connection, second, it might be thought that the second-order identification of an object of the form above amounts in effect to a version of identification of an object by a definite description, i.e.,

$$x = \iota y(\Psi(y)),$$

where ι is the definite description forming operator. The two notions, however, should not be confused. For the identification of an object by a definite description lacks the required modal implications. The fact that $\Psi(y)$ is uniquely true of x implies neither that $\Psi(y)$ is true of x in every possible world. But, as we have seen above, this modal implication can be secured by the second-order identification $\lambda y(y = x) \approx \lambda y(\Psi(y))$; and it was precisely for this reason that the notion of a fundamental sortal is given in terms of second-order identification.

Here let me briefly address one possible objection to this way of meeting Mackie's objection. It might be objected that there are no compelling theoretical or intuitive grounds to suppose that every object admits of a second-order definition. The second-order definition of Socrates, for example, may indeed seem to be quite artificial. Recall, however, that it is one of the core essentialist commitments that every entity is what it is in virtue of its essence. Given this commitment, it is not an unreasonable to think that the essence of an entity consists in *what it is to be that very entity*. And the notion of a second-order definition of an object may just be considered as a more precise account of this essentialist idea.

There thus seems to be a way to meet Mackie's challenge in this case also. It is to think that every object admits of a second-order definition in terms of a fundamental sortal to which it belongs and an associated condition of individuation. The definition, as we have seen, logically implies that it is necessary that the object belongs to the sortal and that it satisfies the condition of individuation. So, the definition-based account can maintain that there just is no logical gap between what is definitive, and what is necessary, of an object. And hence it can give an account of the truth of (BPO); that is, necessary truths about an object can be based on its essence as specified by its real definition.

Before we leave Mackie's objection to the definition-based account, one feature of the proposed solution is worthy of note. In both cases of (BPP) and (BPO), it is a matter of logical form that

the real definition of an entity—either a property or an object—implies necessary truths about it. In other words, what matter is whether the definition provides a specificative identification of the entity in an appropriate form. And this has nothing to do with whether this identification is in terms of some "fundamental" properties in some non-formal and substantive sense. For example, that water is identified in the specificative fashion to H_2O is by itself a sufficient ground for necessary truths about water. And this is regardless of whether water is identifiable in terms of a more basic property, say, the behaviors of protons and electrons. Similarly, one might reasonably think that some objects fall under two distinct fundamental sortals. So, for instance, Socrates may be identified by the sortals *human being* and *person*. Of course, philosophical arguments can perhaps be given concerning which of the two sortals are more fundamental to the identity of Socrates and even concerning whether one is reducible to the other. Regardless, however, both identifications may serve as an adequate ground for necessary truths about Socrates. For, as long as they are formulated appropriately using the machinery of specificative identification, they both logically imply necessary truths about Socrates. The relevance of this point to the present discussion will become clearer later on.

To summarize the discussion of this section, the worry expressed by Mackie and other philosophers was that the essence of an entity, as specified by its real definition, need not be considered as its *necessary* property, for the definition is supposed to state only what the entity *is*; and hence the definition-based account leaves an explanatory gap between essence and necessity. In this section, I have argued on behalf of the definition-based account that this objection can be met. The idea is to require real definitions to have an appropriate logical form. The real definition of a property Φ, on the one hand, ought to identify the property in the specificative way, i.e., by specifying the condition for something to have that very property in the form "$\lambda x(\Phi(x)) \approx \lambda x(\Psi(x))$." Then it can easily be shown that the definition logically implies that the properties Φ and Ψ are necessarily equivalent. On the other hand, an object x can be identified by specifying what it is to be that very object in the form of "$\lambda y(y = x) \approx \lambda x(\Phi(x))$." From the definition of x, then, it logically follows that the object x necessarily has the property Φ. In these ways, the definition-based account of essence can give a straightforward account of the truth of the bridge principles (BPP) and (BPO).

Hence when real definitions are conceived to have an appropriate logical form, there is indeed no explanatory gap between what is definitive, and what is necessary, of an entity.

4. The epistemological adequacy constraint

We shall now turn to the question of whether the definition-based account of essence satisfies the epistemological adequacy constraints. That is, can the definition-based account offer a plausible account of how we may come to have knowledge of the essence of an entity with no appeal to prior knowledge of necessary truths about it?

One might think that this question presents no special difficulty for the definition-based account. For, as we have seen, the real definition of an entity is supposed to state what the entity is; and in many cases, we seem to have knowledge of it. This line of argument has been given by Lowe in his recent paper "What is the source of our knowledge of modal truths?" (2012). He argues:

> Given that all metaphysical modality is grounded in essence, we can have knowledge of metaphysical modality, provided we can have knowledge of essence. Can we? Most assuredly we can. ... Knowing an entity's essence is simply knowing what that entity is. And at least in the case of *some* entities, we must be able to know *what they are*, because otherwise it would be hard to see how we could know anything at all about them. ... In order to *think* comprehendingly about something, I surely need to know *what it is* that I am thinking about. ... [W]ays of thinking of things can be more or less adequate reflections of their essence. ... But I see no reason to suppose that we may *never* fully grasp the whole essence of any kind of entity. (p.944; emphasis original)

In short, Lowe's argument runs as follows. In order to think about an entity at all, we have to identify the entity in one way or another. At least in some cases, the way we identify an entity may represent the essence of the entity. In such a case, we may be said to have knowledge of the essence of the entity. As an example of one such case, Lowe gives what he takes to be the real definition of a circle:

> Consider a familiar geometrical figure, such as a circle. And suppose that someone asks us what a circle is. This can be understood as a request for a real definition of this kind of geometrical figure ... And here, plausibly, is the real definition that is required ... :
>
> (C1) A circle is the locus of a point moving continuously in a plane at a fixed distance from a given point.
>
> The given point in question is, of course, the circle's *centre*. (ibid, p.935; emphasis original)

To analyze Lowe's argument, let us first make a distinction between two kinds of knowledge of essence. Let X be any entity and P be its real definition specifying its essence. Then we say that one has the *non-essentialist* knowledge of the essence of X just in case one knows the *truth* of P. On the other hand, one is said to have the *essentialist* knowledge of the essence of X just in case one knows that P represents the essence of X (i.e., that P is the real definition of X, or the likes). Assuming that (C1) in fact captures the essence of circle, for example, one is said to have the non-essentialist knowledge of the essence of circle if one knows simply that (C1) is true, and to have the essentialist one if one knows that (C1) represents the essence of circle.

It should be clear that Lowe's argument is fallacious if what he means by 'knowledge of essence' is the essentialist one. This is because essentialist knowledge is strictly stronger than non-essentialist knowledge; that is, non-essentialist knowledge does not imply essentialist knowledge. By way of illustration, let us consider the case of circle. Lowe certainly seems to be right that we can think about circles only if we know, in one way or another, what it is to be a circle. We might even grant, for the sake of argument, that we typically think of a circle in accordance with (C1) and even that it represents the essence of circle. What all this amounts to, however, is that we have the *non-essentialist* knowledge of the essence of circle. But this does not imply the desired conclusion that we have the essentialist knowledge, i.e., the knowledge *that (C1) represents the essence of circle*. For one may coherently doubt whether (C1) represents the essence of circle even if he knows that it is true.

To see this point more clearly, imagine a philosophically minded geometer—call him Jones—who is committed to discover the essence of circle. After some hard work, he discovers that circles admit of at least three different but suitably fundamental identifications. In addition to (C1), he has:

(C2) A circle is an ellipse whose focal points coincide.

(C3) A circle is a conic section obtained by cutting it by a plane parallel to the base.

Suppose further that he discovers that all three identifications are provably, and hence necessarily, equivalent. As a committed essentialist, however, Jones does not think the three identifications are all equally good as a real definition of what it is to be a circle. For, recall, the notion of essence is *hyper-intensional*: just because two identifications of an entity are necessarily equivalent, it does not follow that they are equally good as its definition. In fact, Lowe (2012, p.936) himself seems to be in favor of the view that only (C1), but neither (C2) nor (C3), should be taken as representing the essence of circle. This may be right, but what is important for our purpose is to observe that Jones may have doubts as to which of the three identifications represents the essence of circle despite his knowledge that they are all suitably fundamental identifications of circle. In other words, Jones may have only the non-essentialist knowledge of the essence of circle, but not the essentialist one (assuming of course that one of (C1)-(C3) in fact captures the essence of circle).

For another example, let us consider the case of Socrates. Socrates may be identified either as a human being or as a person. It may be the case that one of the identifications captures the essence of Socrates, and it may also be the case that we usually identify him accordingly. Obviously, however, this does not settle the question of which of the two identifications is *definitive* of Socrates, i.e., offers the specification of the essence of Socrates.

More generally, it seems that an entity may admit of more than one suitably fundamental identification which are necessarily equivalent to its definition but which nevertheless should not count as representing its essence. So, even when one think of an entity in a way that corresponds to its essence, one may still have doubts as to whether it represents the essence of the entity. Hence our having the non-essentialist knowledge of the essence of an entity does not imply our having the essentialist knowledge of the essence of an entity.

Let us then suppose that the kind of knowledge Lowe has in mind is the non-essentialist one. Then his claim may appear to have some plausibility, because it would just be that we sometimes identify an entity in a way that aligns with its essence even though we may not know that it does. In fact, this might indeed be what he has in mind when he says, "knowing an entity's essence is simply knowing what that entity is." In another place, he also says:

> All that grasping an essence amounts to, on my view, is *understanding a real definition*, that is understanding a special kind of *proposition*. To know *what a circle is*, for instance, I need to understand that *a circle is the locus of a point moving continuously in a plane at a fixed distance from a given point*. Provided that I understand what a *point* and a *plane* are, and what *motion* and *distance* are, I can understand what a circle is, by grasping this real definition. (ibid, p.946; emphasis original)

In other words, knowing the essence of circle amounts simply to knowing the *truth* of (C1).

However, there seems to be problems with this view also. Recall that like other essentialists, Lowe ultimately seeks to explain knowledge of necessity in terms of knowledge of essence. So, if what Lowe means by 'knowledge of essence' is of the non-essentialist kind, then the resultant view would be, for example, that knowledge of necessary truths about circle derives from knowledge simply of the *truth* of (C1). But this view seems not only to diverge significantly, but also undermines, the core essentialist thesis that knowledge of necessity derives from knowledge of essence.

To see first how it diverges from the essentialist explanation of knowledge of necessity, recall from chapter 5 that according to the essentialist approach, knowledge of necessity is supposed to be explained by *essentialist* knowledge. Knowledge that water is necessarily H_2O, for example, is supposed to be explained by knowledge that water is *essentially* H_2O in conjunction with the corresponding instance of (BPP) as follows:

(1) Being composed of H_2O molecules is part of the essence of water.

(2) If being composed of H_2O molecules is part of the essence of water, then necessarily every instance of water is composed of H_2O molecules.

(3) Necessarily, therefore, every instance of water is composed of H_2O molecules.

Clearly, inferences like this clearly requires the essentialist knowledge of the essence of water. Similarly, (BPO) only allows us to infer a necessary truth about an object from the essentialist truth about its essence. In this respect, the view that knowledge of necessity derives from non-essentialist knowledge of essence diverges significantly from the usual essentialist explanation of knowledge of necessity.

Moreover, if knowledge of necessity only requires non-essentialist knowledge of essence, in what sense our knowledge of necessity depends on knowledge of essence? Let us go back to the case of circle. We have already noted that (C1)-(C3) are all provably equivalent. So, if knowledge of necessary truths about circle is obtained simply from knowledge of the *truth* of (C1)—which we assume to represent the essence of circle—it can also be obtained from knowledge of the truth of (C2) or (C3). Then it seems that knowledge of necessary truths about circle can be explained with no appeal to knowledge of the essence of circle at all, whether it be essentialist or non-essentialist. But then what explanatory value does the notion of essence have in giving an account of knowledge of necessity? None, it would seem.

The discussion of Lowe's account so far reveals a problem that applies to any epistemological account of essence. The problem is this. It is one of the basic commitments of the essentialist approach that the notion of essence is hyper-intensional. This means that not all identifications of an entity may count as representative of the essence of the entity, even though they are all suitably fundamental and necessarily equivalent to one another. This hyper-intensional character of essence seems to preclude an epistemological account of essence that is based simply on non-essentialist knowledge of essence. For, in general, an entity may admit of a number of suitably fundamental and necessarily equivalent identifications. So, if knowledge of necessary truths about an entity derives from the non-essentialist knowledge of the essence of the entity, then it can also derive from any other identifications that are necessarily equivalent to the definitive one. But then it would seem that that the notion of essence is of zero explanatory value in accounting for knowledge of necessity. So, according to the essentialist account, knowledge of necessity should in principle be explained on the basis of essentialist knowledge. And this creates what I take to be the main

problem for the essentialist epistemology of essence. It may be agreed with Lowe that at least in some cases, we do have knowledge of what an entity is in some suitably fundamental sense. Furthermore, such knowledge need not be thought of as dependent upon prior knowledge of what is necessarily true of the entity (unless the entity in question is of some peculiar modal character). But again, an entity may in general have more than one suitably fundamental identification. So, there arises the question of *how we know which identification is representative of the essence of the entity*. This question about essentialist knowledge is what I think is the main epistemological problem for the essentialist approach. It will not do to say, as some might think Lowe is saying, that we just do know which is representative of the essence of an entity. What we need is an account of *how* in principle we may obtain such knowledge, though we may not always be in a position to apply it successfully in specific instances. The problem with Lowe's account is that it simply offers no answer to this question.

Let us then turn to another epistemological account of essence. In his book *Necessary Beings* (2013), Hale seeks to explain our knowledge of essence as grounded (partly) in knowledge of meaning. His approach is shaped by the division of apriori and aposteriori knowledge of essence, so let us discuss each in turn.

When and how is apriori knowledge of essence is possible? According to Hale:

> [T]here are ... some noteworthy similarities between [real and nominal definitions]. In particular, it often happens that what are plausibly taken to be the correct definition of a thing and the correct definition of a word for the thing can be stated using the very same words. ... I think that the cases in which the definition of a thing and the definition of a word for the thing coincide match up with those in which it is plausible to clam that we have apriori knowledge of essence, and that the coincidence helps to explain how it is that we can have such knowledge. (ibid, p.254)

In short, we have apriori knowledge of essence when the real definition of an entity coincides with the nominal definition of a word for the entity. For instance, (C1) may be taken both as the real definition of circle and as the nominal definition specifying the meaning of the word 'circle'. In

this way, the real definition of an entity may coincide with the nominal definition of a word for the entity. In such a case, we can have apriori knowledge of the essence of an entity from the meaning of a word for the entity.

What about in the case of aposteriori knowledge of essence, such as that water is essentially H_2O, that Socrates is essentially a human being, etc. In such a case, according to Hale, there is a general principle telling us that *if an entity X has the property* Φ *then it is in the essence of X that it has* Φ. Then he gives a few examples of such general principles, including:

> **Kind membership**: *any object is essentially an object of a certain general kind*;
> **The origins of organisms**: *any organism is essentially of that very origin that it actually has*;
> **The chemical compositions of chemical substances**: *the chemical composition of a chemical substance is its essential property*.

Notice that these principles imply *conditional* truths about essence. For example, **Kind membership** implies that if Socrates is a human being then he is essentially a human being. So, any empirical knowledge that Socrates is in fact a human being yields empirical knowledge of his essence, namely that he is essentially a human being. Similarly, we may infer from **The chemical compositions of chemical substances** that if water is a chemical substance composed of hydrogen and oxygen atoms in the ratio of 2:1 then water is essentially so composed. Hence the empirical discovery that water is in fact composed of hydrogen and oxygen atoms in the ratio of 2:1 yields empirical knowledge that water is essentially so composed. So, in general, aposteriori knowledge of essence can be obtained from knowledge of a general principle which implies conditional truths of the form "If X is Φ then X is essentially Φ" in conjunction with empirical knowledge that X is in fact Φ.

How, then, can we have knowledge of the general principles about essence themselves? Hale thinks that they can in general be known apriori (ibid, p.269). He believe that the first two principles give above can be shown to be plausible by philosophical argumentation though he himself seems to have some reservations as to how exactly they should be formulated and established (see p.273 and pp.278-279). In contrast, he think that **the chemical compositions of chemical substances** can be established directly from the meaning of the theoretical term 'chemical substance'. He says:

> To be a pure substance, in the chemical sense of the term, just is to be matter having a certain chemical composition. ... There is, accordingly, no mystery about how we may know, a priori, the general principle needed for the application of Kripke's inferential model to substances. Our knowledge that gold is a certain element ... is of course a posteriori, but we know, courtesy of the very meaning of the term, that if a substance has a certain chemical composition, it necessarily does so. (ibid., p.280)

That is, we know that the essence of chemical substance from the meaning of the theoretical term 'chemical substance', just as we know the essence of circle from the meaning of the term 'circle'. Hale claims that similar considerations apply to theoretical identifications in general, such as the identification of light to the stream of photons and of heat to molecular motion (p.281). So, at least for an important class of aposteriori knowledge about essence, according to Hale, the underlying general principles about essence can be known from the meanings of the relevant terms.

This gives a complete account of how we obtain aposteriori knowledge of essence at least for an important class of cases. Taking the case of water for illustration, the account runs as follows. We know from the meaning of the term 'chemical substance' that a chemical substance has its chemical composition as an essential property. Hence any empirical knowledge that water is a chemical substance composed of hydrogen and oxygen atoms in the ratio of 2:1 automatically yields empirical knowledge of the essence of water, namely that water is essentially so composed.

It should be clear from the exposition so far that in Hale's account, there is an important sense in which knowledge of essence is grounded in knowledge of meaning. This dependence seems to raise an immediate problem. For the kind of knowledge of essence that can be grounded in knowledge of meaning seems to be the non-essentialist one. For example, anyone who takes (C1) as definitive of the meaning of the term 'circle' would certainly know that (C1) is *true*. So, if (C1) is in fact representative of the essence of circle, he would certainly have the *non-essentialist* knowledge of the essence of circle. Once again, however, this does not mean that he has the *essentialist* knowledge that (C1) captures the essence of circle.

In fact, Hale himself discusses the problem in his recent paper "Essence and definition by abstraction" (2018), where he says:

> The explanation tells us how, when the definition of a word and real definition of the thing for which it stands coincide, we can gain knowledge of the nature of the thing from our grasp of the meaning of the word. But how is this happy coincidence—between the definition of the word and the essence of the thing—itself to be explained? The meanings of words are, in some suitably broad sense, a matter of convention. We fix them, either expressly by explicit or implicit definition, or in less formal ways, by our ongoing linguistic practices. So, how does it come about that in cases of the kind I have been discussing, we fix them so that they match up so well with the essence or nature of the things the words stand for? (p.11)

That is, why think that the meaning of a word matches up with the essence of the entity for which the word stands? By way of illustration, let us return to Jones the philosophically minded geometer. Jones embarks on his study of circle taking (C1) as its nominal definition. Then he discovers that a circle is a special case of an ellipse whose two focal points coincide and also that an ellipse is a special case of conic section. Our geometer thus realizes that (C2) and (C3) could equally well be taken as a definition of circle. He may thus wonder which of the three captures the essence of circle, i.e., defines what it is to be a circle in the real sense.

In response to this problem, Hale claims:

> *To a first approximation*, we might say that the right picture *reverses* the direction of fit: it is not that we somehow miraculously fix meaning so that it conforms to an independently pre-determinate essence—instead, in these cases, *we fix essence by fixing meaning*. (ibid, p.2012; emphasis original)

That is, we can be in general confident that our knowledge of the meaning of a word corresponds to the essence of the entity for which the word stands because we fix the essence by fixing the meaning. This response even as "a first approximation" seems to me to be plainly wrong, or even preposterous. How could the essence of an entity be fixed by fixing the meaning of a word for

the entity? Would the essence of circle have been different had we adopted (C2) as the nominal definition of 'circle'? For another example, consider the case of chemical substances. How could it possibly be the case that the essence of a chemical substance is fixed by fixing the meaning of the word 'chemical substance'?

What truth, then, is there in this picture? Hale further elaborates his view as follows:

> We fix the meaning of the word—'square' say—by stipulating that to be a square, a plane figure must be composed of four sides of equal length meeting at right angles. But that *there exists* such a property—the property of being a figure so composed—is entirely independent of any such definition being given ... What is right about it—or at least very nearly right—lies in the suggestion that there is little or no room for error. (ibid, p.2014; emphasis original)

The idea seems to be that by fixing the meaning of the term 'circle', we can safely latch onto a property—the property of being a circle—though this property exists objectively and independently of the linguistic stipulation. This much is reasonable enough, but it should be obvious that this does not address the worry. For the problem does not concern how it is that there *exists* a property that corresponds to our linguistic stipulation. It rather concerns how it is that we stipulate the meaning of the word 'circle' so as to capture the *essence* of circle, which is determined objectively and independently of any linguistic stipulations. To this question, Hale's remark gives no answer.

Indeed, it is hard to see how the question about essentialist knowledge can ever be settled by knowledge of meaning. For illustration, let us go back to our philosophically minded geometer. He started his study of circle by taking (C1) as the nominal definition of circle and realized that it is provably equivalent to both (C2) and (C3). He has no doubt that circles exist, and yet still wonders which of the definition is fully adequate for defining what it is to be a circle in the real sense. Similar remarks apply to the case of chemical substance. A chemist may begin his study of chemical substances with the stipulation that a chemical substance is to be a kind of matter with a unique chemical composition. Yet he might think that chemical substances should ultimately be characterized in terms of the behaviors of protons and electrons constituting their molecules.

In general, there seems to be no good reason to assume that the meaning of a word for an entity generally aligns with the essence of the entity.

It thus seems that Hale's epistemology of essence offers a misguided picture of the matter. What it amounts to is the view that the question of what the essence of an entity is can be settled, at least in a large number of cases, by what the meaning of a word for the entity is. But this seems to be plainly wrong. It is one thing that we can latch onto an entity by fixing the meaning of a word for it, but it is another thing that the meaning so fixed *represents the essence* of the entity. Considering that the essence of an entity is given objectively and independently of any conventions we might have, it seems that no linguistic stipulation can ever settle the question of what the essence of an entity is. Hence knowledge of meaning is not by itself an adequate ground for essentialist knowledge.

So much for Hale's meaning-based epistemology of essence. Let me briefly repudiate yet another epistemological account of essence. Anand Vaidya in his "Understanding and Essence" (2010) presents an account of how the use of imagination may help us obtain knowledge of essence. *Variation in imagination* (VIM, in short) is a process of imaginatively constructing a series S_1, S_2, ... of counterfactual scenarios where the representative properties of an entity X are replaced, one by one, with other properties (ibid, pp.819-820). A property Φ is said to be a *variant* property of X if there is a counterfactual scenario S_i in the series where X survives the change of Φ with another property. Otherwise, Φ is said to be an *invariant* property of X; in other words, Φ is invariant if X should be considered as non-existent in every counterfactual scenario in the series where Φ is replaced by another property. The invariant property of X, Vaidya claims, can be judged to be essential to X (p.820).

One obvious problem with Vaidya's account is that it fails to address the question of essentialist knowledge. Recall that the question concerns how to discern essential properties of an entity. Vaidya seems to think that we would be justified in judging that a property is essential to an entity if it can be seen through VIM that the property should be held invariant in any counterfactual scenarios. Obviously, however, what the invariant at best shows is that the property is *necessary*. To see this, suppose that the subject arrives via VIM at knowledge that Socrates could not exist

without being a human being. Then by simple inference, one may also obtain knowledge that Socrates could not exist without being a human being such that $2+2=4$. So, the property of being such that $2+2=4$ is also an invariant property of Socrates. On Vaidya's account, then, it should count as part of the essence of Socrates that $2+2=4$. But this consequence just means that VIM fails to discern essential properties from merely necessary ones. Hence Vaidya's account seems to fall short of delivering essentialist knowledge.

Another objection to Vaidya's account ought to be mentioned also. Vaidya (ibid, pp.819-822) himself points out that there is a sense in which the subject cannot be said to *gain* knowledge of essence via VIM. For the subject of VIM can obtain knowledge about the essence of an entity only if it makes a conscious judgment of whether the replacement of a property with another in a counterfactual scenario would imply the non-existence of the entity. But this seems to imply that the subject should already have implicit knowledge of what the essence of the entity is. Hence VIM can at best be regarded as a process of making explicit what the subject already implicitly knows.

Here one might think that this objection poses no serious problem for Vaidya's account. By way of analogy, consider any non-trivial theorem of the standard ZFC set theory, say the Cantor-Schröder-Bernstein Theorem. The usual proof of this axiom involves an elaborate construction. No matter how elaborate, however, the proof can be carried out with the axioms of ZFC. So, anyone who knows the truth of the axioms should be in a position to prove, and thereby to know the truth of, the theorem. But, intuitively, this does not mean that anyone who knows the truth of the axioms should already be in possession of the implicit knowledge of the theorem, nor that the proof plays no role in gaining the knowledge of the theorem. When viewed this way, it should be clear that the above objection to Vaidya's account is just an instance of the long-standing and more general problem of epistemic closure. So, any solution to the problem of epistemic closure should address the worry. In this regard, the objection creates no special problem for Vaidya's account. [5]

[5]It should be noted that this is not Vaidya's own answer to the objection. Vaidya admits that we cannot obtain essentialist *knowledge* of essence through VIM. Vaidya nevertheless thinks that VIM can produce what he calls *objectual understanding* of the essence of an entity. Objectual understanding, according to Vaidya (ibid, pp.823-824), is distinguished from knowledge in that it is compatible with epistemic luck. So, when the subject of VIM accidentally reaches a correct judgment that an entity has a certain property as its invariant property, it can still be said to obtain objectual understanding—though not knowledge—of the essence of the entity. It is not entirely clear what Vaidya means

However, the real issue here does not concern whether VIM can help the subject gain "new" knowledge of essence, whatever it might exactly mean. What is really problematic for Vaidya's account is that the kind of knowledge that the subject ought to have for VIM to work is *modal* one. Recall that VIM involves reasoning about which of its properties an entity could survive losing. In other words, the subject of VIM should be able to make a correct judgment as to whether an entity would exist in various counterfactual situations. It should be clear that the kind of knowledge that the subject ought to have for such a judgment is modal one. Hence the real problem for Vaidya's account is that VIM relies on prior knowledge of necessity, violating the epistemological adequacy constraint.

Indeed, this problem of dependence upon prior modal knowledge seems to apply generally to any epistemology of essence based on imagination. As noted by Mallozzi (2021, pp.1334-1335), imaginative reasoning about counterfactual scenarios "can be developed virtually in any way, and so become theoretically and practically irrelevant" unless it is properly constrained. Whatever the constraints may exactly be, the subject should be able to reason correctly about what would be the case in counterfactual scenarios. It thus seems that the subject ought already to have at least implicit modal knowledge if one is to obtain knowledge of essence via imaginative reasoning about counterfactual scenarios. In other words, it appears that any epistemology of essence based on imagination should be in violation of the epistemological adequacy constraint.

To summarize: We began by distinguishing two kinds of knowledge of essence, namely essentialist and non-essentialist knowledge. We have the essentialist knowledge of the essence of an entity when we know that the entity *essentially* has a certain property. When we have the non-essentialist knowledge, on the other hand, we know that the property belongs to the entity as a matter of fact without necessarily realizing that the property constitutes (part of) the essence of the

---

by "accidental." Whatever this might exactly mean, it would be absurd to think that the subject can reach objectual understanding of the essence of an entity by complete coincidence. At the very least, some reasons must be given for holding a certain property fixed throughout the process. Though the reasons might not be conducive to knowledge, they ought to be conducive to objectual understanding. But then the same problem just repeats at the level of objectual understanding: it seems the subject must already have implicit objectual understanding of the essence of an entity if VIM is to be of any use. The notion of objectual understanding, therefore, seems to face the exact same objection. See Michels (2020, pp.343-345) for a similar criticism of Vaidya's account.

entity. There appears to be no special problem about how we may obtain the latter kind of knowledge. For, as Lowe seems to suggests, we should be able to identify an entity in one way or another if we are to think about it at all; and it might very well happen that we identify the entity in terms of its essential property. What the defenders of the essentialist approach needs to explain, however, is how we can obtain essentialist knowledge of essence. This was in fact obvious from the usual essentialist explanation of knowledge of necessity. How, then, is essentialist knowledge possible? Here there is a major epistemological difficulty for the essentialist approach. An entity in general admits of a number of different but suitably fundamental identifications, which can sometimes be shown to be necessarily equivalent. Even in such cases, however, they cannot be said to be all equally good definitions of the entity. Hence the defenders of the essentialist approach ought to provide a principled account of how we know which of the identifications is definitive, i.e., represents the essence, of the entity. *Pace* Lowe, therefore, they cannot simply say that there just is no problem concerning how we have knowledge of essence. Then we have examined Hale's meaning-based epistemology of essence, according to which knowledge of essence is grounded partly upon knowledge of meaning. We have seen that Hale's account also fails to address the question of essentialist knowledge. The nominal definition of a term for an entity may coincide with the real definition of the entity. Even in such cases, however, knowledge of the nominal definition of a term for an entity is never by itself sufficient for the essentialist knowledge of the essence of the entity. For one may not know that the two definitions coincide. Finally, Vaidya's imagination-based epistemology of essence maintains that we can come to have essentialist knowledge of essence by imaginative reasoning about counterfactual situations. We have noted two problems with this account. First, Vaidya claims that a property may be considered essential to an entity if it remains invariant in all counterfactual situations; but this is to confuse an essential property with a merely necessary property. Second, Vaidya's account seems to be in violation of the epistemological adequacy condition. This is because VIM process can yield knowledge of essence only if the subject is already in possession of *modal* knowledge as to what would happen in counterfactual scenarios.

5. Summary

Let me briefly restate the main points. The definition-based account of essence maintains that the essence of an entity is specified by its real definition, and so that we come to know what the essence of an entity is by knowing its real definition. One of the main objections to the definition-based account is that it fails to give an account of the truth of the bridge principles. I have first argued that this objection is not decisive. For when the real definition of an entity is conceived to have an appropriate logical form, it can logically be shown to imply necessary facts about the entity. Hence the alleged explanatory gap between essence and necessity can be bridged. Then I have argued that the main problem for the definition-based account is the question of essentialist knowledge: given that an entity admits of a number of different but suitably fundamental identifications, how do we know which is definitive of the entity? I have examined some of the extant epistemological accounts of essence and argued that they all leave this question unanswered. Of course, the discussion so far is by no means a refutation of the definition-based account of essence. For it only shows that the attempts so far have failed. But I hope to have shown at least that the question of essentialist knowledge presents a stiff challenge for the definition-based account.

CHAPTER 8

Toward An Analytic Essentialist Account of Necessity

What makes necessary truths not merely true but necessarily so, and how do we recognize them as such? In this dissertation, I have attempted to address these metaphysical and epistemological questions about necessity from both logical and philosophical perspectives. In giving a philosophical treatment of these questions, I have paid special attention to theoretical identity statements, such as "Water is H_2O". In this final chapter, I aim to do three things. First, I shall bring the findings of the previous chapters into a brief sketch of a novel account of necessity, which I call the *analytic essentialist* account. Second, I shall discuss the analytic essentialist account extends to other kinds of necessary truths. Then finally, I shall conclude with a brief discussion of how it is different from the orthodox essentialist approach.

In chapters 1 and 2, I began by proposing Kripke's Principle as a formal, logical answer to the metaphysical question about necessity: what makes a proposition necessarily true is that which excludes all the ways in which the proposition might be falsified from the realm of real possibilities. This led us to the substantive, philosophical question: in each case of necessary statements, what exactly is it that excludes all the ways in which the statement might be falsified as unreal possibilities?

To address this question, I have examined various cases of theoretical identity statements. Some theoretical identity statements are widely considered to be necessary, while others are not. What exactly is the difference between the two cases? In chapters 3, I have argued that behind each of those theoretical identity statements that are widely agreed to be necessary, there lies a general principle implying that the statement is necessary if it is true at all. For example, it is a general principle about chemical substances that if a chemical substance has a certain chemical composition then it could not have had another chemical composition. Given that water is in fact a chemical substance that is composed of hydrogen and oxygen atoms in the ratio of 2:1, therefore, the general principle

implies that any counterfactual supposition to the contrary is not a real possibility; in other words, that water is necessarily H_2O. In general, the necessity of a theoretical identity statement about an entity is established as a consequence of two things: what the entity in fact is and a general principle implying that it could not have been otherwise. How then can such general principles be established? In chapter 4, I have argued that they are in general analytic in the sense that they are derivable from the definitions of relevant concepts. In particular, I have presented a derivation of the above principle about chemical substances from the concept of a chemical substance as a kind of matter with a unique chemical composition.

These considerations give us a straightforward account of the epistemological question about necessity, which is in accordance with Kripke's as expounded in his paper "Identity and Necessity" (1971):

> [I]f P is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis, some conditional of the form "if P, then necessarily P." If the table is not made of ice, it is necessarily not made of ice. On the other hand, then, we know by empirical investigation that P, the antecedent of the conditional, is true—that this table is not made of ice. We can conclude by modus ponens:
>
> $$\frac{P \qquad P \supset \Box P}{\Box P.}$$
>
> The conclusion—'$\Box P$'—is that it is necessary that the table not be made of ice, and this conclusion is known a posteriori, since one of the premises on which it is based is a posteriori. So, the notion of essential properties can be maintained only by distinguishing between the notions of a priori and necessary truth, and I do maintain it. (p.153)

More generally: we can sometimes know from philosophical analysis that if an entity is a certain way then it could not have been otherwise. So, if we also discover, either empirically or apriori, that the entity is in fact that way, then we may infer that the entity could not have been otherwise. This way, we can come to have knowledge of necessity. In this brief remark, Kripke does not explain what he means by 'philosophical analysis', leaving unclear how exactly the major premise

of the inference, namely $P \supset \Box P$, may be known. The present analysis fills this gap by showing that it can be established by analysis of relevant concepts.

We thus have a complete metaphysical and epistemological account of necessity. Its main thesis can be stated as follows:

- The necessity of a statement about an entity is established as a consequence of a general principle implying that if the entity is a certain way then it could not have been otherwise and the fact that the entity is indeed that way.
- The general principle is analytic in the sense that it can be established by analysis of relevant concepts.
- Hence the necessity of the statement can be known by investigating what the entity is actually like and by conceptual analysis.

I call this new account of necessity *analytic essentialism*. On this account, necessity is neither a sheer creation of our linguistic conventions nor a feature arising solely from the natures of the entities in the world. It is a joint product of the language and the world. Accordingly, knowledge of necessity is obtained through the joint effort of philosophical analysis and scientific investigation.

The analytic essentialist account has been motivated and developed primarily by analysis of theoretical identity statements. One might thus wonder whether and how the present account extends to other kinds of necessary truths, especially those concerning objects. So, let me briefly discuss how the present account extends to these cases.

Consider first how aposteriori necessities about objects can be established; e.g., "Socrates has such-and-such a biological origin." Its necessity can be established as follows: letting o be the actual biological origin of Socrates,

(O1) 'Socrates' rigidly designates an object x that we identify in such-and-so ways.

(O2) x is a human being that has o as its biological origin.

(O3) If a human being has a certain biological origin then it could not have had another biological origin.

(OC) Socrates necessarily has o as its biological origin.

Notice that this argument is completely analogous in its form to the ones we gave for the necessity of theoretical identity statements in chapter 3. (O2) states an aposteriori truth about what Socrates is. (O3) states a general principle about human beings; and just like previously, it can be shown to be analytic in the sense that it derives from the concept of a human being, which it may be defined as thus:

(DH) an object *x* is a *human being* if and only if there is a human biological origin *o* such that the property of being *x* is identical, in the specificative sense, to the property of having *o* as its unique biological origin.[1]

Notice that (DH) is analogous in its logical form to that of chemical substance given in chapter 4 except that it defines a first-order concept applicable to objects. With some slight modifications in the argument, therefore, it can easily be shown that this definition implies (O3). In this way, aposteriori necessities about objects can be explained in quite the same way as the necessity of theoretical identity statements.

Here I should express my reservations about this and some other purported examples of aposteriori necessities concerning objects. Is it really necessary that Socrates has the very biological origin that he actually has? Why, for another example, should the material origin of an artifact be considered its necessary property? My intuitions concerning these cases are much shakier than those concerning theoretical identity statements. But I do not wish to prejudge the issue, and perhaps my personal views about these particular cases are less important than the general analysis of the matter. The important point here is that if these cases turn out to be genuine examples of necessary truths then they can easily be accounted for by the present account.

Notice that basically the same account applies to apriori necessities concerning objects; e.g., "{Socrates} has Socrates as its sole member." The necessity of this truth can be established thus:

(M1) {Socrates} is a set that has Socrates as its sole member.

[1]Let me hasten to add that (DH) is given only by way of example; and I do not pretend that it offers a completely adequate definition of the concept of a human being though I certainly think that it aligns fairly well with the modern understanding of biological species in terms of biological lineage. However, one might think that biological species ought to be defined partly in terms of their internal structures; see, for example, Devitt (2021). This and other related issues are certainly important in their own right, but they need not concern us here.

(M2) If a set x has certain entities and nothing else as its members then it is necessary that x has exactly those entities as its members.

(MC) $\{$Socrates$\}$ is necessarily a set that has Socrates as its sole member.

As before, (M2) may plausibly be considered an analytic principle deriving from the concept of set, which we may define as follow:

(DS) an object x is a *set* if and only if there are certain entities such that the property of being x is identical, in the specificative sense, to the property of having exactly those entities and nothing else as its members.

Obviously, again, this definition is analogous in its form to (DC) and (DH) above. So, it should not be difficult to see that this definition implies (M2); and hence (M2) is apriori. Since (M1) is certainly apriori, it follows that (MC) is also apriori.[2] [3]

It should be noted here that analytic essentialism does not seek to reduce metaphysical necessity into a form of analyticity. In fact, quite the opposite is the case. To see this, consider the identity of water to H_2O again. This is a substantive scientific discovery and cannot be established simply by analysis of the meanings of the terms involved. What is established by philosophical analysis in this particular case is only a general truth about chemical substances that if a chemical substance has a certain chemical composition then it could not have had another chemical composition. From this alone, however, the desired conclusion that water is necessarily H_2O simply does not follow. To reach this conclusion, we must also have the premise that water is in fact H_2O. Like considerations apply also to the other cases of necessary truths that we have discussed in the present dissertation. In this regard, according to analytic essentialism, there is a crucial sense in which necessary truths about an entity depend on *what the entity in fact is*.

[2]Notice that (DS) aligns well with the standard Cantorian definition of set as "a collection of entities into a whole." Some may reasonably think that (DS) is not entirely satisfactory. For one thing, there is a problem about how exactly to formalize this definition. This is mainly because it is unclear how exactly the initial existential quantification may be formalized. However, the intuitive idea behind the definition should be clear.

[3]Here one might also wonder how apriori necessities concerning properties can be accounted for. Why, for example, is it necessary that no bachelor is married? Well, that is because 'bachelor' just means unmarried male. So, necessary truths of this kind do not seem to create any problem for the present account.

Another way of seeing this point ought to be mentioned also. Recall from chapter 4 that we have shown that given (DC), the descriptive identification of water as the chemical substance composed uniquely of hydrogen and oxygen atoms in the ratio of 2:1 is equivalent to the specificative identification of water as the property that a quantity of matter has if and only if it is composed of hydrogen and oxygen atoms in the ratio of 2:1. This was basically the point of lemma (E). Notice here that basically the same applies to necessary truths concerning objects as well. Given (DH), for example, the descriptive identification of Socrates as the human being having *o* as its unique biological origin can easily be shown to be equivalent to the specificative identification of the property of being Socrates as the property that an object has just in case it has *o* as its unique biological origin; and it is from this specificative identification that we derive the desired necessary truth, namely (OC), using the principle of necessary coextensiveness of identical properties. Like considerations apply to (DS) and (MC) as well. Thus we can see again that on the present account, necessary truths about an entity depend crucially on the identity of the entity.

Analytic essentialism can easily be confused with the orthodox essentialist approach. This is because of the structural affinity in their explanations of necessity. Both, after all, seek to explain necessity by appeal to what the entities in the world are like and certain general principles. Nevertheless they differ in some significant ways. The difference lies in how each component is understood. As we have seen, the orthodox approach understands both components in terms of essence. But how exactly is the notion of essence supposed to be understood? Why should the essence of an entity be considered necessary? And how do we come to have knowledge about essence? The orthodox approach owes us a prior account of essence that addresses these questions. In chapters 6 and 7, I have reviewed two major accounts of essence and argued that neither gives us satisfactory answers to these questions. The explanation-based account of essence, on the one hand, fails to give an account of the bridge principles; so, it is inadequate for the purpose of explaining necessity as having its metaphysical source in essence. On the other hand, the definition-based account does not provide us a credible epistemology of essence; so, it is inadequate for the purpose of explaining our knowledge of necessity in terms of essence. Hence it remains as a problem for the orthodox approach to reconcile a plausible metaphysics of essence with a credible epistemology

in a way that satisfies the original goal of explaining necessity and our knowledge thereof in terms of essence—this was the integration challenge for the orthodox essentialist approach.

It should be evident that analytic essentialism does not suffer from this problem. For it makes no appeal to the notion of essence. All it requires is just the usual notion of identity. Or, to put it another way: unlike the orthodox approach, analytic essentialism can simply understand the notion of essence in terms of the usual notion of identity as follows: the essence of an entity is nothing over and above what the entity is. This simple account of essence is clearly metaphysically adequate because the notion of identity can surely be understood independently of the notion of necessity. It is also explanatorily adequate because, as we have seen, the identity of an entity when conjoined with a relevant analytic principle is sufficient to give an account of necessary truths about the entity. Finally, it does not seem to create any special epistemic problem. For, at least in many cases, we do know what an entity is in the usual sense of identity.[4] So, the simple account seems to be epistemologically adequate. In a sense, then, we may say that analytic essentialism offers an account of essence that meets the integration challenge.

[4]Here one might ask how exactly we come to have knowledge of the identity of an entity. Here I am inclined to agree with GMP that explanatory relations often provide compelling evidence for the identity of an entity. In this connection, see my discussion of GMP's super-explanatory account above in p.132-135.

Bibliography

Angelberger, A., F. L. G., Faroldi, and J. Korbmacher (2016). An exact truthmaker semantics for permission and obligation. In *Deontic Logic and Normative Systems* (Proceedings of DEON16). College Publications.

Armstrong, D. M. (2004). *Truth and truthmakers*. Cambridge: Cambridge University Press.

Ayer, A. J. (1936). Truth by convention. *Analysis 4*(2/3), 17–22.

Barnett, D. (2000). Is water necessarily identical to H_2O? *Philosophical Studies 98*, 99–112.

Belnap, N. (1977). A useful four-valued logic. In Dunn, M. and G. Epstein (eds.) *Modern Uses of Multiple-Valued Logic*. D. Reidel.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy 70*(19), pp.661–680.

Bird, A. and E. Tobin. (2018). Natural Kinds. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2018 Edition. Metaphysics Research Lab, Stanford University.

Boghossian, P. A. (1996). Analyticity reconsidered. *Noôs 30*(3), 360–391.

Braddon-Mitchell, D. and F. Jackson (1997). The teleological theory of content. *Australasian Journal of Philosophy 75*(4), 474–89.

Braddon-Mitchell, D. and F. Jackson (2002). A pyrrhic victory for teleonomy. *Australasian Journal of Philosophy 80*(3), 372–77.

Burgess, J. P. (2013). *Kripke: Puzzles and Mystery*. Cambridge: Polity.

Burgess, J. P. (2021). Kripke on modality. In O. Bueno and S. A. Shalkowski (Eds.), *The Routledge Handbook of Modality*, pp. 400–408. New York: Routledge.

Carnap, R. (1947). *Meaning and Necessity, A Study in Semantics and Modal Logic*. Chicago: The University of Chicago Press.

Casullo, A. (2020). Essence and Explanation. *Metaphysics 2*(1), pp.88–96.

Cobreros, P. et al. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic 41*(2), 347–385.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy 72*, 741–64.

Davidson, D. (1987). Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association 60*(3), 441–458.

Devitt, M. (2021). Defending intrinsic biological essentialism. *Philosophy of Science 88*(1), 67–82.

Devitt, M. (forthcoming). *Biological Essentialism.* Oxford: Oxford University Press.

Dummett, M. (1959). Wittgenstein's philosophy of mathematics. *The Philosophical Review 68*(3), pp.324–348.

Fine, K. (1994). Essence and modality: The second philosophical perspectives lecture. *Philosophical Perspectives 8*, 1–16.

Fine, K. (2002). The varieties of necessity. In T.S. Gendler and J. Hawthorne (eds.) *Conceivability and Possibility*, pp.253–82. Oxford: Clarendon Press.

Fine, K. (2014). Truth-maker semantics for intuitionistic logic. *Journal of Philosophical Logic 43*(2-3), 549–577.

Fine, K. (2015). Unified foundations for essence and ground. *Journal of the American Philosophical Association 1*(2), pp.296–311.

Fine, K. (2016). Angellic content. *Journal of Philosophical Logic 45*(2), 199–226.

Fine, K. (2017a). A theory of truthmaker content I: Conjunction, disjunction and negation. *Journal of Philosophical Logic 46*(6), 625–674.

Fine, K. (2017b). Truthmaker semantics. In Wright, C. and B. Hale (eds.) *A Companion to the Philosophy of Language*. Oxford: Blackwell.

Fine, K. (forthcoming). Introduction to "*Essays on Essences and Existence*". In Hale, B. *Essays on Essence and Existence*. Oxford: Oxford University Press.

Fitting, M. (2021). The Strict/tolerant idea and bilattices. In Arieli O. and A. Zamansky (eds.) *Arnon Avron on Semantics and Proof Theory of Non-Classical Logics*. Springer Nature.

Fitting, M. and R. L. Mendelsohn (1998). *First-Order Modal Logic*. Dordrecht: Kluwer Academic Publishers.

Forbes, G. (1985). *The Metaphysics of Modality*. Oxford: Clarendon Press.

Garcia-Carpintero, M. and J. Macià (2006). *Two-Dimensional Semantics*. Oxford: Clarendon Press.

Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In D. Papineau and G. MacDonald (Eds.), *Teleosemantics: New Philosophical Essays*, pp. 42–68. Oxford: Clarendon Press.

Godman, M., A. Mallozzi, and D. Papineau (2020). Essential properties are super-explanatory: Taming metaphysical modality. *Journal of the American Philosophical Association 6*(3), pp.316–334.

Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy 44*(5), pp.113–128.

Greenberg, M. (2005). A new map of theories of mental content: Constitutive accounts and normative theories. *Philosophical Issues 15*(1), 299–320.

Hale, B. (2002). The source of necessity. *Philosophical Perspectives 16*, 299–319.

Hale, B. (2013). *Necessary beings: An essay on ontology, modality, and the relations between them*. Oxford: Oxford University Press.

Hill, C. S. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 87*(1), 61–85.

Horvath, J. (1997). Lowe on modal knowledge. *Thought 3*(3), 208–217.

Hughes, G. E. and M. J. Creswell (1996). *A New Introduction to Modal Logic*. New York: Routledge.

Jago, M. (2020). Truthmaker semantics for relevant logic. *Journal of Philosophical Logic*, 1–22.

Kleene, S. C. (1938). On notations for ordinal numbers. *The Journal of Symbolic Logic 3*(4), 150–155.

Kleene, S. C. (1952). *Introduction to metamathematics*. New York: Elsevier.

Kim, D. (2021). Explanation and modality: on why the Swampman is still worrisome to teleosemanticists. *Synthese 199*, 2817–2839.

Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.

Kripke, S. A. (1959). A completeness theorem in modal logic. *The Journal of Symbolic Logic 24*(1), 1–14.

Kripke, S. A. (1963). Semantical analysis of modal logic I. Normal propositional calculi. *Zeitschrift fur mathematische Logik und Grundlagen der Mathematik 9*, 67–96.

Kripke, S. A. (1971). Identity and necessity. In M. K. Munitz (Ed.), *Identity and Individuation*, pp. 135–164. New York: New York University Press.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.

Kripke, S. A. (2011). *Philosophical Troubles. Collected Papers Vol I*. New York: Oxford University Press.

LaPorte, J. (2013). *Rigid Designation and Theoretical Identities.* Oxford: Oxford University Press.

LaPorte, J. (2022). Rigid Designators. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy,* Spring 2022 Edition. https://plato.stanford.edu/archives/spr2022/entries/rigid-designators/.

Leech, J. (2021). From Essence to Necessity via Identity. *Mind 130*(519), pp.887–908.

Lepore, E. and B. C. Smith (2006). *The Oxford Handbook to the Philosophy of Language*. Oxford: Clarendon Press.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Loar, B. (1990). Phenomenal states. *Philosophical Perspectives 4*, 81–108.

Lowe, E. J. (2007). A problem for a posteriori essentialism concerning natural kinds. *Analysis 67*(4), 286–292.

Lowe, E. J. (2008). Essentialism, metaphysical realism, and the errors of conceptualism. *Philosophia Scientiae 12*(1), 9–33.

Lowe, E. J. (2011). Locke on Real Essence and Water as a Natural Kind: A Qualified Defence . *Proceedings of the Aristotelian Society Supplementary Volume 85*, pp.1–19.

Lowe, E. J. (2012). What is the source of our knowledge of modal truths? *Mind 121*(484), 919–950.

Luper, S. (2020). Epistemic closure. In Zalta, E. N. (ed.). *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), URL=¡https://plato.stanford.edu/archives/sum2020/entries/closure-epistemic/¿.

Macdonald, G. and D. Papineau (2006). *Teleosemantics: New Philosophical Essays*. Oxford: Clarendon Press.

Mackie, P. (2020). Can metaphysical modality be based on essence? In Dumitri, M. (ed.). *Metaphysics, Meaning, and Modality: Themes from the Work of Kit Fine*, pp.247–264. Oxford: Oxford University Press.

Mallozzi, A. (2021a). Putting modal metaphysics first. *Synthese 198*, pp.1937–1956.

Mallozzi, A. (2021b). Superexplanations for counterfactual knowledge. *Philosophical Studies 178*, pp.1315–1337.

Michels, R. (2020). Husserlian eidetic variation and objectual understanding as a basis for an epistemology of essence *Logos & Episteme 11*(3), pp.333–353.

Miller, A. (2007). *Philosophy of Language* (2nd edition). London: Routledge.

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: The MIT Press.

Millikan, R. G. (1996). On swampkinds. *Mind and Language 11*(1), 103–17.

Nanay, B. (2014). Teleosemantics without etiology. *Philosophy of Science 81*(5), 798–810.

Neander, K. (1996). Swampman meets swampcow. *Mind and Language 11*(1), 118–29.

Neander, K. (2013). Toward an informational teleosemantics. In D. Ryder, J. Kingsbury, and K. Williford (Eds.), *Millikan and Her Critics*, pp. 21–40. Oxford: Wiley-Blackwell.

Needham, P. (2000). What is water? *Analysis 60*(1), 13–21.

Needham, P. (2002). The discovery that water is H_2O. *International Studies in the Philosophy of Science 16*(3), 205–226.

Nimtz, J. (2017). Two-Dimensional Semantics. In B. Hale, C. Wright and A. Miller (Eds.), *A Companion to The Philosophy of Language* (2nd edition), pp. 948–970. Oxford: Blackwell.

Noonan, H. W. (2018). The new Aristotelian essentialists. *Metaphysica 19*(1), pp.87–93.

Papineau, D. (1996). Doubtful intuitions. *Mind and Language 11*(1), 130–32.

Papineau, D. (2001). The status of teleosemantics, or how to stop worrying about swampman. *Australasian Journal of Philosophy 79*(2), 279–89.

Papineau, D. (2006). Naturalist theories of meaning. In E. Lepore and B. Smith (Eds.), *The Oxford Handbook of Philosophy of Language*, pp. 175–188. Oxford: Clarendon Press.

Papineau, D. (2007). Kripke's proof is ad hominem not two-dimensional. *Philosophical Perspectives 21*, 475–494.

Papineau, D. (2016). Teleosemantics. In D. L. Smith (Ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism*, pp. 95–120. Cambridge: Cambridge University Press.

Peacocke, C. (1999). *Being Known*. Oxford: Clarendon Press.

Peters, U. (2014). Teleosemantics, swampman, and strong representationalism. *Grazer Philosophische Studien 90*(1), 273–288.

Post, J. F. (2006). Naturalism, reduction and normativity: Pressing from below. *Philosophy and Phenomenological Research 73*(1), 1–27.

Priest, G. (1979). The logic of paradox. *The Journal of Philosophical Logic 8*(1), 219–241.

Priest, G. (2021). Metaphysical necessity: a skeptical perspective. *Synthese 198*, 1873–1885.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review 60*(1), 20–43.

Quine, W. V. O. (1960). Carnap and logical truth. *Synthese 12*, 350–374.

Roca-Royes, S. (2021). The integration challenge. In Bueno, O. and Scott A. Shalkowski (eds.). *The Routledge Handbook of Modality*, pp.157–166. New York: Routledge.

Romero, C. (2019). Modality is not explainable by essence . *The Philosophical Quarterly 69*(274), pp.121–141.

Rosen, G. (2015). Real Definition . *Analytic Philosophy 56*(3), pp.189–209.

Salmon, N. (1979). How not to derive essentialism from the theory of reference. *The Journal of Philosophy 76*(12), 703–725.

Salmon, N. (1981). *Reference and Essence*. Princeton, NJ: Princeton University Press.

Salmon, N. (2003). Review: naming, necessity, and beyond. *Mind 112*(447), 475–492.

Schulte, P. (2020). Why mental content is not like water: Reconsidering the reductive claims of teleosemantics. *Synthese 197*(5), 2271–2290.

Smith, D. L. (Ed.) (2016). *How Biology Shapes Philosophy: New Foundations for Naturalism*. Cambridge: Cambridge University Press.

Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of* Naming and Necessity. New York: Oxford University Press.

Soames, S. (2006). Reply to critics. *Philosophical Studies 128*(3), 711–738.

Soames, S. (2007). The substance and significance of the dispute over two-dimensionalism. *Philosophical Books 48*(1), 34–49.

Stalnaker, R. (1968). A theory of conditionals. In Nicholas Rescher (ed.). *Studies in Logical Theories (American Philosophical Quarterly Monographs 2)*, pp.98–112. Oxford: Blackwell.

Vaidya, A. J. (2010). Understanding and essence. *Philosophia 38*(4), pp.811–833.

Van Fraassen, B. C. (1969). Facts and tautological entailments. *The Journal of Philosophy 66*(15), 477–487.

Williamson, T. (1999). Truthmakers and the converse barcan formula. *Dialectica 53*(3-4), 253–270.

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

Wittgenstein, L. (1956). *Remarks on the Foundations of Mathematics*. Oxford: Basil Blackwell.