AGAINST MODALITIES

ON THE PRESUMED COHERENCE
AND ALLEGED INDISPENSABILITY
OF SOME MODAL NOTIONS

by

Kaveh Lajevardi

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Philosophy, University of Toronto

Copyright © 2008 by Kaveh Lajevardi


Abstract

Against Modalities: On the Presumed Coherence and Alleged Indispensability of Some Modal Notions

Kaveh Lajevardi
Doctor of Philosophy
Department of Philosophy
University of Toronto
2008

Part I investigates the idea that kinds (as opposed to individuals) have some modal properties. I argue that concerning typical kind-essentialist claims there is a non-trivial question—the transworld identity problem—about what the relevant kind terms are supposed to refer to in non-actual possible worlds. I reject several ideas for solving the problem. The upshot is a worry about the coherence of modal talk concerning kinds.

Waiving this worry for the sake of argument, in Part II the target is the use of modal talk in the sciences. I offer a deflationary account of modalities, based on the familiar idea of reducing modalities to logical relationships between non-modal statements and non-modal background theories. I argue that this account is adequate for making sense of modal talk in the sciences. Moreover, I argue that irreducible modal properties of the world, if there are any, cannot be scientifically discovered or inferred.

Thus we have a number of arguments against modalities: the threat of incoherence, their epistemic inaccessibility, and the dispensability of modal talk in the sciences.
Acknowledgements

I am immensely indebted to Anjan Chakravartty. Qua my supervisor, Anjan is directly responsible for whatever clarity one may find in the pages to follow, for a great number of the dialectical moves I have made, for the absence of almost all the footnotes that were previously there (many of which are now incorporated into the text in a way that is much more presentable than those irresponsible, fun-to-write-but-hard-to-read footnotes), and for the correction of many ungrammaticalities. Qua metaphysician and philosopher of science, he contributed *extensively* to this project.

I was very lucky to have had logician and philosopher of language Philip Kremer, and (for a while) philosopher of science Margaret Morrison on my thesis committee; my thanks to Phil and Margie for the discussions and encouragement.

I am honoured to have had Ian Hacking in my committee. In the course of several discussions over the past three years I benefited from his unbelievably vast knowledge and his criminally sharp mind. Despite his overall agreement with me on the first part of the dissertation, he never ceased to offer criticism; and despite his disagreement with me on some methodological features of the second part, he never deprived me of his guidance. My deepest thanks to him, with admiration.
Two senior professors of the department, James Robert Brown and Bernard D. Katz kindly checked some parts of the dissertation for the accuracy of presentation. I also cheerfully remember the pleasant and stimulating discussions with Jim on thought experiments and philosophy of mathematics, and with Bernard on many topics, many of which were related to modalities. I was most pleased, quite flattered indeed, when Bernard once asked me “Did you just receive a postcard from Quine?”

My thanks to Eran Tal and Jessica Wilson for their detailed comments on what now constitute parts of Chapter 4 and Chapter 1, respectively.

Though not a philosopher by training, Farzaneh Sarafraz, with her acumen and common sense, patiently and critically listened to some of my theses and arguments while they were taking shape. (Based on a comment of hers on what is now Appendix 5, I was tempted for a while to choose ‘Unnecessary Falsehoods’ as the title of the dissertation.) I think no first-order description will do justice to what I owe her on other grounds; so I will not try. Thanks, fahym.

# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>PART I: METAPHYSICS</strong></td>
<td></td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td>20</td>
</tr>
<tr>
<td>Essentialism about kinds: the transworld identity problem</td>
<td></td>
</tr>
<tr>
<td>1. Kripke's use of 'essential' 21</td>
<td></td>
</tr>
<tr>
<td>2. The problem of transworld identity: individuals vs. kinds 28</td>
<td></td>
</tr>
<tr>
<td>3. Some ideas for a solution 39</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER 2</strong></td>
<td>47</td>
</tr>
<tr>
<td>&quot;That kind of thing&quot;</td>
<td></td>
</tr>
<tr>
<td>1. Sameness of kind 48</td>
<td></td>
</tr>
<tr>
<td>2. Importance and essentiality 50</td>
<td></td>
</tr>
<tr>
<td>3. Later Putnam 56</td>
<td></td>
</tr>
<tr>
<td><strong>CHAPTER 3</strong></td>
<td>64</td>
</tr>
<tr>
<td>Stipulation</td>
<td></td>
</tr>
<tr>
<td>1. The stipulative idea 67</td>
<td></td>
</tr>
<tr>
<td>2. Stipulation: a <em>sine qua non</em> 72</td>
<td></td>
</tr>
<tr>
<td>3. Shoemaker's causal theory of properties 81</td>
<td></td>
</tr>
<tr>
<td>4. The applicability of Shoemaker's theory 89</td>
<td></td>
</tr>
<tr>
<td>5. Conclusion 93</td>
<td></td>
</tr>
<tr>
<td><strong>PART II: PHILOSOPHY OF SCIENCE</strong></td>
<td></td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td>96</td>
</tr>
<tr>
<td>Modal talk in the sciences: a rational reconstruction</td>
<td></td>
</tr>
<tr>
<td>1. Mathematics 98</td>
<td></td>
</tr>
<tr>
<td>2. Physics 109</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5
The epistemic inaccessibility of non-relativistic modalities
1. Modal features of the physical world? 135
2. What scientists cannot do 137

CHAPTER 6
Realist views of modality
1. Cartwright 154
2. Shoemaker’s theory revisited 171

APPENDICES
1. Modality de re and de dicto 178
2. Haecceitism 179
3. Adams’s observation 183
4. The history (or lack thereof) of modal relativism 187
5. The failure of □φ → φ 190

References 193
Introduction

This dissertation consists of a presentation and a defence of two theses which, if convincing, undermine some current uses of modal notions. The theses are:

A. There is a fundamental difficulty — the transworld identity problem — in attributing modal properties to kinds and substances. For instance, there is a fundamental difficulty in saying that water is necessarily H$_2$O.

B. A deflationary account — one that reduces assertions about necessity and possibility to assertions about logical relations between statements and non-modal background theories — suffices for understanding modal talk in the sciences.

Also, as part of the argument for B, I will try to establish a third thesis, namely:

C. Even if there are genuine modal facts about the world (i.e., modal facts whose truth-conditions are not reducible to non-modal ones), such facts are beyond the epistemic reach of empirical scientists.
Thesis A attacks the coherence of a kind of modal talk. B holds that in order to make sense of modal talk within the sciences (be it of the kind which is the subject of A, or other kinds like ‘It is impossible to move faster than light’), we need not understand such talk as being about genuinely modal features of the world: we may understand a scientist’s assertion that such and such is necessary as an assertion to the effect that such and such follows from our scientist’s background theory. Thesis C says that even if there are genuinely modal features of the world, scientists cannot discover or infer them, so that such facts, if any, cannot be parts of the sciences. Hence further support for B.

This is a war against modalities, fought on two fronts: metaphysics (thesis A), and the philosophy of science (theses B and C). Chapters 1-3 are devoted to A, Chapters 4-6 deal with B and C.

In the remainder of this introduction I will try to motivate the discussions, flag some relevant issues not investigated in this dissertation, make a comment or two on some chapters, and present some conventions that I will use.

I consider theses A-C as, at least in part, results of investigating my qualms about certain views expressed in the third lecture of Saul Kripke’s masterpiece, *Naming and Necessity*. Here are two examples of the kinds of the views that worry me:
Let us suppose the scientists have investigated the nature of gold and have found that it is part of the very nature of this substance, so to speak, that it [has] the atomic number 79.

‘Cats are animals’ has turned out to be a necessary truth. [...] In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind.¹

So, according to Kripkean essentialism, at least some kinds or substances have some of their properties necessarily, and at least some of these necessary properties are discovered by scientists. (Throughout the dissertation, I will use ‘essentially’ and ‘necessarily’ and their cognates as synonyms; in particular, I will understand essentiality as a modal notion – see the first section of Chapter 1 for a discussion.)

A.

Let me first say what I am not primarily concerned with in the chapters to follow. As is well known, in the first lecture of Naming and Necessity Kripke argues that true identity statements between individuals (like ‘Hesperus is

¹ Kripke (1980: 124, 138, my square brackets). To my knowledge, the first passage is the only place where Kripke of Naming and Necessity modifies his use of ‘nature’ by adding a ‘so to speak’-like qualifier.
Phosphorus’) are metaphysically necessary — i.e., true in every possible world. One important idea in the argument is rigidity: Kripke defines a designator as **rigid** iff it refers to the same thing in every possible world wherein the designator, as we use it in the actual world, has a reference. He argues that proper names are rigid designators, and it is but one short step from the thesis of the rigidity of proper names to the inference that true identity statements in which ‘=’ is flanked by proper names are metaphysically necessary. With the brevity of his exposition growing exponentially, towards the end of the lectures Kripke says that designators like ‘molecular motion’ and ‘heat’ are rigid, and therefore ‘Heat is molecular motion’ is necessary, if true. However, at least at first glance, ‘heat’ is very different from ‘Hesperus’ as the latter is a singular term while the former is not. Thus, given the definition of rigidity that Kripke gives us, it is not quite clear what he means by saying that ‘heat’ is a rigid designator.

Given that there is such an important lacuna in the argument for the necessity of true statements of scientific identification, it is interesting that its conclusion seems to enjoy such widespread acceptance, at least among philosophers who are not experts in this particular area — and I think this is a tribute to the elegance, eloquence, and cogency of the first two lectures of *Naming and Necessity*, together with its author’s deserved fame as a genius who is also the unrivalled authority on modal logic. One may surmise that many philosophers just assumed that there was a way of filling in the gaps in
the argument, only Kripke did not have time to do it. Kripke himself never explained how rigidity was supposed to be defined for terms like ‘heat’, ‘pain’, ‘water’, and ‘gold’ — at least not in print.

In Kripke, then, talk of rigidity regarding general terms remains at a hand-waving level, and this soon caught some philosophers’ attention. As early as 1973 (barely two years after the publication of Kripke’s first philosophical paper, “Identity and necessity”, and one year after the publication of the paper version of Naming and Necessity), Donnellan expressed worries about the way we should define the rigidity of common nouns — see also his (1983). As yet, there is no uncontroversial definition of rigidity for non-singular terms. In the new millennium, we have Soames’s (2002) attempt, criticized by Salmon (2003), who presents his own account (2005a) — the discussion goes on, without everyone’s satisfaction.

Now the notion of rigidity of general terms is not what I find most worrisome. My main worry is more fundamental, and not limited to cases where one asserts that two kinds are necessarily identical, whatever we may mean by two kinds being identical. Prior to an examination of a claim to the effect that a kind is necessarily so and so, I want to know what it is for that kind to be necessarily so and so. When we say that in a possible world water is H₂O (or is green, or is heavy, or is precious), what is the thing, or the kind of thing, about which we are saying that it is H₂O (or is green, etc.) in that possible world? As we use the term ‘water’, what is its reference in non-actual
possible worlds? I think it is only after answering these questions that one can move on and talk about the rigidity of ‘water’ or defend an essentialist claim about water.

My main question in the first part of the dissertation, then, is the **transworld identity problem for kinds**, here for the kind water: What is the kind water in a non-actual possible world? Or (what turns out to be a closely related question in Kripke’s treatment of his essentialist claims): What is it, in a non-actual possible world, to be an instance of the kind water? (We should keep in mind that to satisfy a Kripkean essentialist, the answer should be such that ‘Water is H$_2$O’ turns out to be necessary *a posteriori*. Thus in this context one cannot say that by ‘water’ one just analytically means H$_2$O.)

In Chapter 1 I will argue that there is a transworld identity problem for kinds. What makes the issue perhaps worthy of more careful examination in the context of Kripkean essentialism is that, of course, one may ask the transworld identity question *for individuals* as well: What is it for something to be Hesperus in a non-actual possible world? In order to justify the coherence of rigidity talk—which requires talking about *the same individual* across possible worlds—Kripke forcefully argues that the transworld identity problem for individuals is a pseudo-problem, in the sense that while talking about different possible worlds containing (say) Nixon, it is quite obvious what we are talking about: we are talking about *Nixon, Nixon himself*, says Kripke. Part of what I do in Chapter 1, after fixing the terminology, is to show
that even granting that there is no difficult transworld identity problem for individuals, there is still a non-trivial transworld identity problem for kinds. I will also suggest that Kripke’s trivialization of the transworld identity problem for individuals may be seen as at least partly rooted in the standard, possible-worlds semantics of first-order modal logics, developed by Kripke et al. in the 1950s and early 1960s, and I will argue that, insofar as this semantics is concerned, there is an asymmetry between the case of individuals and kinds. By looking at the standard semantics, the transworld identity problem for individuals can be solved in an obvious way; not so for the transworld identity problem for kinds.

In fairness to Kripke, it should be said that although he does say that the general term ‘gold’ is rigid, in his argument for the claim that ‘Gold has atomic number 79’ is necessary if true, he does not appeal to the alleged rigidity of ‘gold’ — he seems to be talking about instances of gold, arguing that they necessarily have atomic number 79. (That he does not directly talk about the kinds he is essentialist about and that he does not use the alleged rigidity of corresponding kind terms can be better seen in his argument that cats are necessarily animals if they are actually animals—see Kripke (1980: 122, 125 f.).) So perhaps Kripke really need not define rigidity for general terms in order to argue for his essentialism about natural kinds. On the contrary, in his arguments for the necessity of true statements of identity for natural phenomena he heavily appeals to rigidity claims, as is manifest in the case of
pain in his argument against the mind-body identity theory (1980: 148 f). Thus perhaps the lack of a clear definition of ‘rigid’ for general terms does not threaten Kripke’s essentialism about kinds like gold and water; what does pose a threat to this essentialism is the transworld identity problem for such kinds. To narrow down my discussion, I will forget about natural phenomena and natural-phenomenon terms (like light and ‘pain’, respectively).

There is a problem about the non-actual references of general terms; but is there also a solution? In the last section of Chapter 1 I rather quickly deal with what I consider trivial solutions, which do not work, at least not in a way a Kripkean essentialist would be happy with. Some of these solutions fail for simple technical reasons, some because they really do not address the question, as I understand it. Thus I think someone who, in response to the transworld identity question for the kind gold, says something like “Well, by ‘gold’ in every possible world I mean the same old universal, Gold”, is not really answering the question: he is just pushing it back—we need to know what that old universal is, what the extension of ‘gold’ in a non-actual possible world is. (If I were allowed to write like David Lewis, I would ask: Does ‘universal’ just mean don’t worry?)

---

2 There might be deeper reasons for considering natural-kind terms and natural-phenomenon terms separately: Gray (2006) suggests that natural-phenomenon terms have a semantics different from that of natural-kind terms.
There are interesting solutions, though. In Chapter 2 I investigate a Putnamian solution (better put: a Putnam 1975ian solution) to the transworld identity problem for kinds, here for the kind water, according to which for every \( x \) and for every possible world, \( x \) is an instance of water in that possible world just in case in that world \( x \) shares important properties with the (paradigmatic) instances of our water, water in the actual world. I argue that this strategy cannot be favoured by an essentialist who thinks of the metaphysical essence of an entity as indicating what that entity “really is”. The reason is that importance is an interest-relative notion: what is important depends on one’s interests; but what an entity “really is” presumably transcends one’s interests. There are other problems with this attempt to solve the transworld identity problem; here one ally — who, to my knowledge, is one of the very few number of philosophers who recognize the transworld identity problem for kinds or substances as such — is a philosopher named Hilary Putnam, qua the author of Putnam (1992). The last section of the Chapter 2 is devoted to the later Putnam’s arguments about the solution in question.

Another attempt to solve the transworld identity problem for kinds is a very Kripkean one: stipulation. This is the subject-matter of Chapter 3, where I argue that an essentialist should subscribe to Shoemaker’s causal theory of properties, and this, I argue, restricts what one can stipulate about kinds or properties in possible worlds, thereby depriving one of a stipulative solution
to the transworld identity problem for kinds. This is a difficult area for me, partly because I do not see clearly what we mean by stipulating something about a possible world. To a large extent, I will be satisfied with the metaphysical part of the dissertation if it makes it plausible that there is a transworld identity problem which cannot be trivially solved, and if I manage to clarify the issue of stipulation to some extent.

Penelope Mackie (2006: 174) writes, “The most important question about the natural kind essentialism of Kripke, Putnam, and others is surely: is it true?” If I am right— if there is a transworld identity problem for kinds, and if the ideas that I examine for solving it really do not work— then it is perhaps fair to say that the most important question about Kripkean natural-kind essentialism is rather: What does it mean? If my arguments go through, then the very coherence of the kind-essentialist position is at stake.3

If a warranted conclusion of Chapters 1-3 is that essentialist talk about kinds is in trouble, then this is probably an unexpected result.

---

3 Mackie also refers to Putnam as an essentialist, which is in accordance with the very common practice of talking about “Kripke-Putnam” or “Putnam-Kripke” essentialism about kinds. Hacking (2007) argues that there is no such thing as Kripke-Putnam or Putnam-Kripke essentialism, properly so-called: even Putnam of (1975a) is not, strictly speaking and some of his phrases notwithstanding, an essentialist. (Having said that, perhaps I should add that I think there is such a thing as the Putnam-Kripke theory of natural-kind term reference, as developed in Putnam (1973a).)
Animadversions concerning individual-essentialism have always been stronger than those concerning kind-essentialism. It is an old idea that what is essential to an object is a matter of how you describe or categorize it—it has been argued that ‘essential’ is applicable, if at all, to properties of kinds, not properties of individuals. Quine has the famous example of an individual who is both a cyclist and a mathematician. Is this individual essentially rational? Is he essentially biped? The answer depends on whether you consider him qua a mathematician or qua a cyclist. Moral: ‘essential’ cannot be applied to the properties of an individual; it may describe the way a kind has a property. Two and a half centuries before Quine we have John Locke, who thinks that it is only to “Sorts” that one can significantly attribute essences—this is not unexpected in the case of nominal essences (Essay, III.vi.4), but he also holds the same view regarding real essences (III.vi.5). It is partly in opposing this tradition that Kripke’s talk about essential properties of individuals (like the necessity of origin) was revolutionary. But now, if I am right, it seems that the traditional essentialism (about kinds) faces a difficulty which might not affect the more revolutionary one.

---

4 Quine (1960: 199). Quine uses this example to argue against the coherence of quantifying into modal contexts. I think his point is conclusively refuted by Plantinga (1974: 23 ff).
B.

Returning to the quotes I gave from Kripke at the beginning of the
Introduction, in Chapters 4-6 I will, most of time, assume that modal talk
about kinds does make sense. So, in these chapters, let us assume that we do
understand what it is for something to be an essence of (say) gold and what it
is for gold to have some of its properties essentially; also, let us assume that
gold does have some of its properties essentially.

But do scientists deal with essential properties? In one sense they may: if
some of the properties of gold are essential to it, then it is imaginable that,
while investigating properties of gold, scientists discover that gold has some
of those properties. We may therefore say that scientists may thereby discover
some essential properties of gold. This is a kind of discovery that John V.
Canfield calls extensional—thus Columbus’s discovery of the new land was
extensional.5 My worry, however, is whether scientists may discover that so
and so is an essential property of gold (what Canfield calls intentional
discovery). This is the main question in the second part of the dissertation.

5 Canfield (1983: 107). The analogy of Columbus is perhaps imperfect: it might be said that his
discovery of the new land was merely extensional because he was mistaken about the thing
he discovered (he thought it was a place in Southeast Asia); but of course in principle he
could have discovered that America is the new land. One of the things that Canfield argues
against is the Kripkean idea that scientists can discover that so and so is the essence, or part of
nature, of gold.
Before explaining the worry, I would like to mention a related question, not dealt with in this dissertation. Essence discovery is, for Kripke, a *discovery*: it is not that after finding out that a designated few of those shiny, malleable, highly durable things have atomic number 79 we changed the definition of ‘gold’ so that it is now part of the definition of ‘gold’ that gold has atomic number 79. In fact, it is supposed to be one of the great advantages of the causal theory of reference for kind-terms, as developed by Kripke (1980) and Putnam (1973a), that it does not imply — and does allow one to deny — that meanings of terms like ‘gold’ change as sciences progress. However, there is a rival view here. Surely it is not a piece of antediluvian knowledge that what we call ‘gold’ is an element having atomic number 79; but — so says the rival account — after we made this discovery, the scientific community decided to change the definition of ‘gold’ so that now nothing that fails to have atomic number 79 lies in its extension. Canfield (1983) suggests that scientists change the meanings of scientific terms; he does so partly via rejecting the causal theory of reference. Joseph LaPorte (2004), too, argues for meaning-change, but he also argues that the causal theory of reference cannot block meaning-change, and offers a detailed account of how change of meaning does not lead to relativism and irrationality of theory choice. For my conviction that scientists do not discover essences I refer to these authors. I will therefore not investigate one main feature of Kripke’s essentialism, namely his claim about the constancy of meaning. I will not talk about essences either. Here I am
interested in the claim that sciences deal with necessary or essential
properties of kinds and substances, but I will not examine the more particular
claim of the alleged scientific discovery of essences. (A reminder of the
difference: according to Kripke, cats are essentially animals; but surely being
an animal is not the essence of the kind cat—at most it is part of its essence.)

Neither will I touch on Kripke’s realism about so-called natural kinds. I do
not think that there is a uniquely correct way of grouping things, but I do not
examine this topic here—and I have nothing to add to John Dupré’s
arguments for his doctrine of promiscuous realism as developed in his (1993).

To turn to a worrisome aspect of the quoted passages from Kripke that I
deal with in this dissertation, my question is: Do the sciences deal with
modalities? Of course scientific discourse, at least at the level of exposition for
the layman, is replete with ‘cannot’, ‘must’, ‘possible’ and the like—‘It is
impossible to build a perpetual motion machine’, ‘You cannot measure both the
momentum and the position of an electron’, and so on. But how should we
understand these modalities? Under what conditions does a scientist assert
that something is a necessary property of gold? How could a scientist find out
that something is necessarily (but not analytically) true of gold? How does a
scientist argue that moving faster than light is (physically) impossible? How
can he argue that a non-actual situation is possible?

My view, a relativistic account of modalities which I present as a rational
reconstruction of scientists’ modal talk, is that modal talk in the sciences can
always be understood as relative to a (highly context-sensitive) background theory. In a given context, a statement is considered necessary if and only if it follows from the background theory, possible if and only if it is compatible with the background theory. If my rational reconstruction works, then one need not appeal to any genuine modalities in order to understand scientific practice: all the apparently modal talk is reducible to non-modal talk.

I develop the view in Chapter 4, which begins with the special and more manageable case of necessity and possibility in mathematics. The chapter then turns to the case of physics, and tries to make the relativistic account look more plausible via considering two examples (Newton’s first law of motion, and the second law of thermodynamics), and examining some objections to the relativistic view.

C.

Thesis B holds that modal talk in the sciences is reducible to logical relationships of non-modal statements and background theories. But one may ask if there are modalities in background theories themselves. In Chapter 5 I argue that even if there are modal facts about the world whose truth-conditions are not given in terms of any non-modal background theories (if, in short, there are irreducible modal facts about the world), the empirical scientist cannot know them, not even by way of inferring them. Hence there
are no modalities in the background theories either. If I am right, then irreducible modalities are irrelevant to scientific practice.

It might be illuminating to compare the position advocated here with two different kinds of interpretations of Hume. Regarding necessity, in the *Treatise* and the first *Enquiry* Hume is mainly concerned with necessary *connexion*, which, in the context of these works, might perhaps be best understood as necessary connexions *between events*. Nevertheless, one might suppose that Hume would say similar things about attributing necessary properties to substances and kinds, insofar as these attributions are claimed to be matters of fact, not relations of ideas. It is not absolutely clear that Hume’s notion of necessity is modal, but let’s assume it is—my objective here is not to understand the historical Hume.

There are at least two kinds of interpretations of Hume on necessities in nature; following Blackburn (1990), let us call these kinds of interpretations *positivist* and *sceptical realist*, the latter term being John P. Wright’s (1983). According to the positivist interpretation of Hume, which is by far more common than the other kind of interpretation, Hume thinks that no necessity talk is intelligible, and that talk of necessities is meaningless. The reason is that, according to him, ideas are (combinations of) copies of sense impressions; given that we have no impressions of necessities (a position held by both the positivist and the sceptical realist Hume), it follows that we have
no idea of necessities. Necessity talk is just illegitimate, unintelligible, and meaningless.

In the 1980s there was something of a movement dedicated to understanding Hume as a sceptical realist, a movement perhaps culminating in the publication of Galen Strawson’s (1989). According to this reading of Hume, he does not denounce necessity talk as meaningless. In fact, he even thinks that there are necessities in nature, only we human beings cannot know them.

Regardless of the degrees of faithfulness of these interpretations to the historical Hume, I think my position in this dissertation lies somewhere between them. Unlike the positivist Hume, I do not subscribe to a theory of meaning according to which simple legitimate ideas are copies of sense impressions. But I have another worry—the transworld identity problem—about the coherence of talking about necessary properties of substances. So in denying the coherence of modal talk about substances and kinds I am with the positivist Hume, though for a different reason.

In the second part of the dissertation, I waive my worries about the coherence of modal talk about substances and kinds; thus, for the sake of discussion, in Chapter 5 I assume that, contrary to the positivist Hume’s position, necessity talk is legitimate. Now unlike the sceptical realist Hume, I do not think that there are necessary properties of substances—my official
position here is agnosticism. But like the sceptical realist Hume, I do think that we cannot discover such modal facts (if they exist).

**Conventions.**

- In definitions only, I use Paul Halmos’s ‘iff’, which reads ‘if and only if’. When I define a term I write it in boldface letters.
- Single quotation marks are used to build names of linguistic objects. Double quotation marks are used for quoting and as scare quotes.
- Following Soames (2002), instead of enclosing a sequence of symbols in Quine corners I will write it in boldface italics (see below for an example).
- By *talking about* \( x \) in possible world \( W \) I mean talking about the following: \( x \) in \( W \). I am nowhere concerned with talking-about-\( x \) in \( W \), if \( W \) is not the actual world.
- I do *not* always italicize Latin phrases and their abbreviations like ‘e.g.’ or ‘a posteriori’; italics, should they occur, show my emphasis.
PART I
METAPHYSICS
CHAPTER 1
Kind-essentialism: the transworld identity problem

It is generally agreed that Kripke’s essentialism about individuals was more revolutionary than his essentialism about kinds—perhaps this partly explains why, in his published works, discussions of the latter are much sketchier than the former. Nevertheless, in this chapter I will argue that the less revolutionary part of Kripkean essentialism suffers from a difficulty concerning the transworld identity problem, which makes it at least as problematic as essentialism about individuals. Kripke argues (1971, 1980) that the transworld identity problem is a pseudo-problem when we talk about essential properties of individuals; my main goal here is to argue that while talking about essential properties of, say, water, it does make sense to ask what water in a non-actual possible world is—the transworld identity question for kinds is not a pseudo-problem. Then, mainly in the next two chapters, I will argue that it is not clear that the Kripkean essentialist can solve the problem to his own satisfaction.

Section 1 is a brief, partly exegetical, discussion of Kripke’s use of ‘essential’ when the term is applied to properties of individuals. It also comments on Kripke’s understanding of the attribution of essential properties
to kinds. Section 2 deals with the problem of transworld identity for kinds or properties, and argues that, insofar as the standard, first-order possible-worlds semantics is concerned, there is an asymmetry between the case of the transworld identity problem for kinds and the transworld identity problem for individuals. Section 3 examines some of the easier suggestions for solving the problem, and argues that none of them works. More promising suggestions will be examined in Chapters 2 and 3.

1. Kripke’s use of ‘essential’.

In *Naming and Necessity*, Kripke does not offer an explicit definition of ‘essential’; in a companion paper, “Identity and necessity”, he does (1971: 151-152). Here is the definition, with my wording.

**Definition 1.** A property is *essential* to an object iff the object has it in every possible world wherein the object exists.

From the great emphasis Kripke puts on the identity of material objects in *Naming and Necessity* when discussing their essential properties,¹ one might think that the definition he has in mind in that work is something like this:

---

¹ Examples include the discussions of the necessity of origin in Kripke (1980: 113, 142).
**Definition 1.** A property \( P \) is **essential** to an object \( a \) iff for every possible world \( W \) and for every object \( b \), if \( b \) does not have \( P \) in \( W \) then \( b \) is not identical to \( a \).

In fact, in “Identity and necessity” he offers the two formulations in a single sentence (ibid.), and observes in a footnote that they are extensionally equivalent. I think the second formulation better reflects the way Kripke argues about essential properties of things.

Kripke notes that Definition 1 has the perhaps undesirable consequence that existence, if counted as a property, is trivially an essential property of everything, and some modifications have to be made if one wants to say that only necessary existents have the property of existence essentially (1971: 151n11). To make the discussion more manageable, here I will “forget” about the question of existence—in fact, in the preface to *Naming and Necessity*, Kripke himself explicitly says that he avoids talking about existence in that work.²

---

² Kripke (1980: 21n21). Judging from the title, Kripke has discussed the issue of existence in his John Locke Lectures, “Reference and existence”, at Oxford University in the academic year 1973-74. Officially, a transcript of lectures is only available at Oxford Philosophy Library, and one may not borrow or photocopy or quote from it—see [http://www.bodley.ox.ac.uk/guides/philosophy/Holdings.htm](http://www.bodley.ox.ac.uk/guides/philosophy/Holdings.htm) [viewed 17 January 2008].
What about essential properties of kinds? Kripke says (1980: 138) “In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind”. Assuming that the essence of an entity is among its essential properties, it follows that, for Kripke, at least some kinds have some of their properties essentially. But how should we understand the attribution of an essential property to a *kind*, as opposed to an object?\(^3\)

For Quine (1969: 118), kinds are sets of objects—though, of course, perhaps not every set of objects is a kind. Kripke does not explicitly address the question, ‘Is every kind a set of individuals?’; yet, whatever his considered opinion on the ontological status of kinds might be, it seems that what he says is compatible with the view that kinds are sets of objects. I have two reasons for this claim. First, in (1980: 121) he says that creatures like tigers, “living together, looking alike, mating together, do form a kind”; so, for Kripke, things of a given kind do *form* that kind. I take this to be evidence for the claim that for Kripke, things of a given kind are members of the kind,

---

\(^3\) Throughout this dissertation I use ‘object’ and ‘individual’ interchangeably. Defining what is exactly meant by these terms is, I think, not easy. Also, when one is talking about, e.g., the water in this glass, it sounds awkward to describe the water as an ‘object’ or ‘individual’. [Cf. Mill (1872: 579, my italics): “We perceive an object; say, for instance, water. We recognize it to be water, of course by certain of its properties.”] At any rate, I think the difference between *the water in this glass* and *the kind (or the substance) water* is intuitively clear.
and that, hence, kinds are sets. Second, and more importantly, while he says that we use ‘gold’ as a term for a kind of thing and ‘tiger’ as designating a species (1980: 118, 121) and so on, in examining essential properties of such kinds he invariably talks about their instances. For example, he does not argue that being animal is an essential property or a necessary property of the kind cat; he talks about statements like ‘Cats are animals’ (1980: 124, 138). In Kripke’s essentialism, in order to talk about essential possession of a property by a kind one may talk of attributing essential properties to instances of that kind, thus attributing an essential property to members of a set of objects (in the objectual sense of ‘essential’ discussed above). Moreover, this is the only way that Kripke actually argues for the claim that a given kind has a certain property essentially, and he never pauses to say how from the proposition that cats are essentially so and so it follows that the kind cat is essentially so and so — this is simply taken as obvious, suggesting that it is compatible with what he says to assume that kinds are sets of individuals.

One can even argue for a stronger claim: it seems that considering kind terms as predicates — and not as singular terms — is the default position, or at least so it was in 1970, when Kripke delivered his lectures on naming and necessity. It is therefore perhaps safe to say that had Kripke thought otherwise he would have commented on the issue. In fact, he says (1980: 127, my italics) “The old term ‘common name’ is thus quite appropriate for predicates marking out species or natural kinds, such as ‘cow’ or ‘tiger’.”
This being so, I will use *K is essentially so and so* and *Essentially, Ks are so and so* synonymously, where *K* is a kind term. By the second boldly italicized sentence I mean this: in no possible world is there anything which is an instance of *K* in that world but not so and so in that world. As an example after the next definition shows, this de dicto statement must not be confused with *Ks are essentially so and so* in the de re reading.4

Now I think for sets of objects (in particular, for kinds) Kripke uses a similar implicit definition. Thus (1980: 126) suppose we have somehow come to know that it is part of “their very nature” — as Kripke puts it — that cats are animals, and consider a counterfactual case where we are qualitatively in the same epistemic situation with respect to some species of demons as we actually are situated with respect to this-worldly cats. For Kripke, this is not a situation involving *cats*; rather, it involves cat-like *demons*. So I suggest the following definition as something that might have been presented by Kripke if he were to give us his definition of the term ‘essential’ when applied to names of classes of objects.5

---

4 A de re formula attributes a modal property to an *individual*; a de dicto formula is one which is not de re: if it there is any modality in a de dicto formula, it is attributed to a *sentence*, not an individual. See Appendix 1 for more on this.

5 See also the discussion of gold (124) and light (130 f) in Kripke (1980).

In the phrase quoted above Kripke is talking about the “nature” of things; also, some sentences in *Naming and Necessity* might suggest that for Kripke the *nature* of a thing is not
**Definition 2.** A property \( P \) is **essential** to a kind \( K \) iff for every possible world \( W \) and for every object \( b \), if \( b \) does not have \( P \) in \( W \) then \( b \) is not a \( K \) in \( W \). (In other words, \( P \) is essential to \( K \) iff, necessarily, \( Ks \) are \( P \), where the scope of ‘necessarily’ is the sentence \( Ks \text{ are } P \).)

Thus having atomic number 79 is presumably an essential property of (instances of) gold, being animal is presumably an essential property of cats, and so on. Also (not Kripke’s example), according to this definition, being unmarried is essential to the kind bachelor (or, to use a familiar phrase, it is essential to bachelors *qua bachelors*): if Travis is not unmarried in a possible world, he is not among the bachelors in that world—though, of course, according to the previously given definition of what is essential to an object (Definition 1, here applied to Travis), being unmarried is presumably not an essential property of Travis: presumably, there are possible worlds wherein he is not unmarried.

---

the same thing as its *essence*—here is an example already quoted (1980: 138, italics mine): “In general, science attempts, by investigating basic structural traits, to find the *nature*, and thus the *essence* (in the philosophical sense) of the kind”. But nowhere does he elaborate on his understanding of the difference between the two, and I am under the impression that he uses the two terms interchangeably, with ‘nature’ being used with a higher frequency. (He might have had in mind some ancient distinction between the two notions.)
Exploring Kripke’s (implicit) definition of *essence*—as opposed to *essential properties*—of a kind or an object will distract us from the main theme of the chapter; very briefly: one could argue that Kripke’s text is compatible with a Plantinga (1974) type of defining essence: a property $e$ is an *essence* of an object $x$ iff, in every possible world, $y$ is identical to $x$ if and only if $y$ has property $e$ in that possible world.

Let us wrap up the terminological discussion. Do the above definitions capture the idea of essentiality? Kripke (1971: 151n11) and Putnam (1983:54) would answer this question affirmatively, while Kit Fine (1994) argues that defining the notion of an essential property in terms of possible worlds is misguided. In addition, we have James Van Cleve’s (1995) worries about using the term ‘essential’ both for some properties of kinds and for some properties of objects.

However, these debates should not affect my main business in this chapter, which is not defining ‘essential’. My task is to examine the view—which I call *kind-essentialism*—that some kinds have non-trivial essential properties, in an intuitive sense of ‘non-trivial’. To this end, given that essential properties of something (whatever the correct formulation of the notion might be) are surely necessary for it, throughout what follows one

---

6 For more on triviality in this context, see Kripke (1971: 151). See also McMichael (1986: 33) for elaboration on what he calls ‘interesting’ essential properties.
might replace ‘essential property’ by ‘necessary property’, with the above definitions understood as defining the latter term. If there are, as I believe, no non-trivial necessary properties of kinds, then, a fortiori, there are no non-trivial essential properties of kinds either. So I will focus on criticizing the (probably weaker) view that kinds have non-trivial necessary properties. In fact, using ‘essential’ and ‘necessary’ synonymously is a commonly accepted thing to do in modern, analytic-style discussions of essentialism.7

2. The problem of transworld identity: individuals vs. kinds.

It is now a platitude—and Lecture I of Naming and Necessity, especially pp. 42 ff, might be partially responsible for its common acceptance—that one source of difficulties people have had with essentialism about individuals is its relationship with de re modality. While, in accordance with the definitions presented in the previous section, talking about essential properties of kinds is to make de dicto statements (e.g., □(x)(Kx → Px), which says that essentially, Ks are P), talking about essential properties of objects clearly involves de re modality—we are saying of an object that it necessarily is so

7 Examples include Plantinga (1976: 141), McMichael (1986: 33), and Chandler (1986: 381). Also Kripke (1980: 135): “though each of these items is, indeed, essentially (necessarily) gold, gold might have existed even if the items did not.” See also Mumford (2004: 123) for some reservations which lead him to talk about necessary, rather than essential, properties of natural kinds.
and so. And philosophers have long had difficulties with the very notion of 
de re modality—difficulties which seemed to be more serious than problems 
they had with de dicto modality.8 I am here concerned with one such 
difficulty, and I will use this discussion as an introduction to explaining my 
own problem with essentialism about kinds.

One might suspect that the very notion of an essential property of an 
object—as well as any other notion involving de re modality—is problematic 
or even incoherent unless it comes with a solution to the so-called problem of 
transworld identity for objects: “If you want to investigate the question 
whether, for instance, having green eyes is essential to Travis,” the sceptic 
might say, “you must first be able to identify Travis in other possible worlds 
where he exists. It is only after doing this that you might be able to examine 
the colour of Travis’s eyes in those worlds. And it is not clear at all how one 
could, even in principle, do the first of these tasks—of many individuals in a 
given possible world who look like Travis from different aspects, it is not 
clear which one is to be identified with Travis.” To this epistemological-
sounding formulation of the worry, the sceptic may add that the very 
meaningfulness of essentialism about objects requires a solution to the

---

8 For Quine’s well-known qualms about modalities see § 41 of his (1960) and references 
thereof; for the particular case of de re modality, see p. 199 of that section. For an examination 
of some worries about the notion of de re modality (including those of Quine’s) see Chapter 
two of Plantinga (1974).
problem of transworld identity. If we push him on this issue, the sceptic might say that this is his main worry: he is mainly concerned with the metaphysical problem of what Travis is in another possible world, rather than with the problem of how could we tell what Travis is there.

Note that the idea that there is no as-a-matter-of-fact answer to the transworld identity problem for objects is just one of the motivations for de re scepticism (others include Quine’s position that quantification into modal contexts involves the use-mention confusion). However, technical results obtained by Fine (1978) show that there is a strong relationship between the transworld identity problem and de re scepticism: roughly speaking, a formula is basically kosher for the de re sceptic (i.e., it has a provably de dicto equivalent) if and only if the determination of its truth-value does not require a solution to the relevant transworld identity problem.

To see how significant the transworld identity problem was supposed to be, one only has to recall that at least one prominent philosopher of the ante-“Naming and necessity” era once considered transworld identity as the main problem in modal logic—here is what David Kaplan wrote in late sixties (1979: 94, my square brackets):

this [the transworld identification problem] is the central problem of philosophical interest in the development of intensional logic. The other problems are all technical.
Kripke argues (1980: 43-47, 1971: 146-149) that there is no genuine problem of transworld identity for individuals. Suppose we are interested in what could have happened to Nixon, or suppose we are talking about the properties he has in another possible world. Thus suppose we are talking about a possible world wherein Nixon loses the 1968 U.S. presidential election. Here, says Kripke, it is part of the very description of that possible world that there *Nixon* loses the election. Possible worlds, in the Kripkean understanding of them, need not be described purely qualitatively: we do not have to give a description, without using proper names, of Nixon in a non-actual possible world in which he loses the election—we do not have to say that a man who looks like this and has a dog like that, etc. loses the election. Rather, in talking about such a possible world, it is *given* that we are talking about Nixon; we simply stipulate that we are talking about him, about *Nixon himself*, and that *he* loses the election there. There is no question about the identity of Nixon in the possible world in question. And we do not “look” at this possible world, as if with a modal telescope, to see that such and such a person loses the election. The famous slogan here is “‘Possible worlds’ are *stipulated*, not *discovered* by powerful telescopes” (1980: 44).

I think the way Kripke deals with the issue of transworld identity of individuals has its roots in the way we do the possible-worlds semantics for modal logic. (And I take him to argue, in Lecture 1 of *Naming and Necessity,*
that this is the way we should talk about possible worlds.) If so, no wonder, then, that Kripke, one of the founding fathers of this semantics, favours this approach. I will briefly explain how the possible-worlds semantics turns the transworld identity question for individuals into a pseudo-problem, and then argue that there is a problem with transworld identity of kinds, as opposed to transworld identity of objects, when looked at through the lens of this semantics.

The following catechism may help me in making my point.

**Q.** What is it for the individual $a$ to be an $F$ in a possible world $W$?

More precisely, what is it for the formula $'Fx'$ to be true in $W$ when the value $a$ is assigned to the free variable $'x'$?

**A.** The truth-value in question is $T$ just in case $a$ lies within the extension of $F$ in $W$. (There are considerations about a relation, which logicians call ‘accessibility’, between possible worlds; but here I ignore that issue—vide the Article published by Master Kripke in MCMLXIII.)

**Q.** What is $F$ in $W$?

**A.** It should be given at the outset. In fact, $W$ is not specified as a possible world unless the extension of $F$ in it is specified.

**Q.** What is $a$ in $W$?

**A.** It is just $a$. 
In the standard semantics, in order to say what it is for \( a \) to be an \( F \) in a possible world \( W \), there is no question at all about \( a \): you look at \( a \) itself and see if it lies in the extension of \( F \) in \( W \). The individual \( a \) is the same in every possible world. In the possible-worlds semantics, the name of an individual, if it has a name, is a rigid designator. (The same is the case for free variables: as Kripke (1980: 49n16) says, “free variables can be used as rigid designators of unspecified objects.” In fact, Kaplan thinks that variables are “paradigms of rigid designation” (1989: 493). So it really does not matter whether or not in our language we have a name for the individual in question.) This, I submit, is a fundamental presumption of our standard model theory for modal logic.\(^9\)

On the contrary, even the meaningfulness of the question of \( a \) being an \( F \) in \( W \) requires a determinate answer to this question: What is the extension of \( F \) in \( W \)? The question of transworld identity does arise for \( F \). From the point of view of the standard semantics for modalities, when talking about \textit{this object} (Richard Nixon) in other possible worlds, there is no problem about the thing we are talking about (and ‘Richard Nixon’, as used by us, designates the same object, if any, in every possible world). But if we want to talk about cats being so and so in another possible world \( W \), we \textit{have to} specify the extension of the predicate ‘is a cat’ in \( W \), and the extension may well differ from world

\(^9\) Kripke is not explicit about the relationship between possible-worlds semantics and rigidity. The relationship between the two has also been observed by Jason Stanley: “rigidity arose in the development of the semantics of Quantified Modal Logic” (1997: 558).
to world. There is an asymmetry between the case of individuals (referents of proper names) and kinds (referents of kind terms, or referents of some predicates); the transworld identity problem for the latter does make sense, and solving it is a prerequisite for talking about $Ks$ in other possible worlds.

Admittedly, one might argue that some essentialist claims about kinds are unproblematic with respect to the transworld identity problem, due to the logical form of the statements themselves. Thus one might say that being liquid-or-non-liquid is essential to water since no matter how we solve the transworld identity problem for water (and the transworld identity problem for liquid), in every possible world each instance of water is either liquid or non-liquid. But, of course, this is not a typical essentialist claim—after all, insofar as it has a truth-value, the statement that water is essentially liquid-or-non-liquid is knowable a priori. (Besides, one may coherently hold that if there is no answer to the transworld identity problem for water, then even such a tautological-sounding essentialist claim about water lacks truth-value—cf. the Strawsonian idea that the statement that the present king of France is either bald or not bald lacks truth-value.) Thus if not all, at least every “interesting” a posteriori essentialist claim about kinds requires a solution to the relevant transworld identity problem.

So, working in the framework of the standard semantics, there is a problem with understanding the claim that gold or tiger is essentially so and so, because we need to be told what an instance of gold is or what a tiger is in
other possible worlds. The problem is not the epistemological problem of identifying (instances of) these kinds; it is the metaphysical problem of transworld identity. In the next section I will consider some ideas for solving the transworld identity problem of properties, a task that will be continued in the next two chapters.

It will not escape notice that, despite my (dialectical) sympathy with the Kripkean idea that the transworld identity for individuals is just a pseudo-problem, I must say that there are problems even with essentialist claims about objects. In the case of an essentialist statement ‘a is essentially a $P$', although there might not be a transworld identity problem for $a$, there is, in light of what I have said so far, a transworld identity problem for $P$; thus I should admit that there is a problem with the meaningfulness of many de re essentialist claims (such as ‘This object is essentially gold’) as well.

To a large extent, I think that there is such a problem—hence my statement at the beginning of this chapter: essentialism about kinds is at least as problematic as essentialism about objects. But there are essentialist claims about objects with which we do not have a problem: the (presumably false) sentence that ‘this object [my earring] necessarily exists’ is an example of an unproblematic de re claim when understood in the framework of the standard semantics—it is not problematic because, thanks to the modal logical axioms for the predicate ‘exists’, we know what it is for $a$ to exist in
another possible world (we know that in each possible world \( W \), the extension of the predicate ‘exists’ is the whole domain of \( W \)). In general, if we know enough “logical facts” about \( P \), we know about its extension in other possible worlds and we can make sense of the essential attribution of \( P \) to objects.

But a great number of philosophers, certainly Kripke amongst them, agree with Putnam of the 1960s and 1970s that the great majority of kind terms do not have analytical definitions.\(^{10}\) This being so, it follows that for the majority of kind terms \( K \) we have a transworld-identity barrier in assigning a truth-value to \textit{Essentially, \( K \)s are \( P \)}, at least when \( P \) is an “interesting” property, i.e., not something like the property of being liquid-or-non-liquid, discussed above.

\(^{10}\) See especially Putnam (1970), where he argues at length that properties like that of being a lemon are not analytically definable. For Kripke’s similar position, see his discussions of gold and tigers (1980: 118 ff).

Putnam admits that there are some “one-criterion” concepts, such as bachelorhood or the property of being a vixen, but he estimates that their number is about three hundred, while, according to him, other concepts we use amount to tens of thousands (1970: 141). Note that even for the one-criterion concept of being a vixen (vixen = female fox), we still have the transworld identity problem, for, if Putnam is right, the concept of being a fox cannot be defined analytically.
Let me conclude with some remarks regarding the problem of transworld identity for kinds before I start examining some ideas for a solution.

First, note that the transworld-identity trouble with statements like ‘Essentially, tigers are mammals’ does not—at least not on the face of it—depend on whether we consider the kind tiger as a natural kind. Whenever we have non-trivial statements of the form *Essentially, Ks are so and so*, we encounter the transworld identity problem for *K*; for that matter, *K* could be *any* class of objects which does not have an extensional definition: cats, gold, grue objects, yellow circular objects of diameter less than a mile, etc. — insofar as the kind in question is not given by an exhaustive enumeration of its instances, in which case arguably the transworld identity problem for the kind will be reduced to a bunch of transworld identity pseudo-problems for the individuals that form that kind, we have a transworld identity problem. So the *problem* will arise for any kind of kind. But in order to reject one of the *solutions* I will later examine (the stipulative suggestion, Chapter 3 below), it will be required that the kinds in question are “real”, not gruesome. This, though it restricts the scope of my scepticism about kind-essentialism in its most general form, does not weaken my argument against Kripkean kind-essentialism, for Kripke is concerned only with natural kinds.

Secondly, insofar as my worries about understanding *Essentially, Ks are so and so* are concerned, it does not matter whether the name for *K* is a count noun or a mass term. Surely the *answers* to the transworld identity problems
for kinds, if there are answers, might not be the same for the two categories of common nouns; but what makes kind-essentialism problematic is that the name for $K$ is a general term, be it a count noun or a mass term. Besides, in all cases where Kripke argues that a given kind has some of its properties essentially, his argument seems to be independent from the category of common noun the name for the relevant $K$s belongs to—and he summarizes his view by saying “This conclusion [that natural-kind terms are closer to proper names to a greater degree than it was generally assumed before Kripke] holds for certain for various species names, whether they are count nouns, such as ‘cat’, ‘tiger’, ‘chunk of gold’, or mass terms such as ‘gold’, ‘water’, ‘iron pyrites’” (1980: 134, my square brackets; cf. 1980: 127).

Thirdly, of course one should not over-emphasize the importance of the semantics of formal modal logic here. True, there is an asymmetry between the case of individuals and that of kinds, so far as the standard semantics is concerned; yet, in itself, this might be more of a problem for the semantics of modalities, rather than a difficulty for essentialism about kinds. But I hope that, even apart from purely logical questions, the problem of transworld identity for kinds is motivated enough—when talking about necessary properties of, say, gold, I find it quite natural to ask: What is gold in other possible worlds? That is to say: What is the entity to which we are attributing some properties in some possible worlds? On the other hand, I think one should not underestimate the semantical problem either: if the kind-
essentialist cannot solve the transworld identity problem for kinds, and if he claims that the real culprit here is the standard semantics, he owes us a semantics suitable for his essentialist talk.\footnote{Also, I believe that the Kripkean intuition about the absence of any real transworld identity problem for individuals is at least partially grounded in the fact that there is no corresponding problem in the model theory of modal logic. The intuition is at least weakened in the case of kinds when we see that there is a transworld identity problem for kinds in the standard semantics.}

3. Some ideas for a solution.

The problem, then, is how to make sense of essentialist claims about kinds, given that it does make sense to ask about the extension of a kind term in another possible world—and the question, to repeat, is metaphysical, not epistemological. This section examines some candidates for solving the transworld identity problem for kinds. I will be relatively short with these ideas—more promising ideas will be considered in the following chapters.

1. Let’s try the obvious. This readily comes to mind: “Grant me that you know what it is to be an instance of gold in the actual world. Now by gold being essentially so and so, what I mean is that each actual piece of gold is essentially so and so. Thus I reduce the de dicto claim about gold to a set of
de re claims about instances of gold, and you say you do understand the 
latter statements."

To explicate, let \( g \) be the kind gold (or the property of being gold), and let 
\( P \) be a property, like that of having atomic number 79, which is said to be 
possessed by gold essentially. Let \( G \) be the predicate ‘is gold’. According to 
the solution under consideration, ‘\( g \) is essentially \( P \)’ means the following: for 
each actual object \( x \), if \( G(x) \) is true then the property \( P \) is essential to \( x \), in the 
sense defined in Section 1. This account thus reduces kind-essentialism to 
object-essentialism, which I am assuming, at least for the sake of argument, to 
be unproblematic. (For the sake of illustration, I also set aside the question 
about the reference of \( P \) in other possible worlds.)

However, Kripke believes that there could have been things which do not 
exist in the actual world. Thus in the Appendix (a) to Naming and Necessity 
(1980: 158) he approvingly quotes from his (1963):

The substantial point I was trying to make, however, remains […]. The point 
was that, in other possible worlds ‘some actually existing individuals may be 
absent while new individuals … may appear’ […].

Therefore, if I am right in my understanding of what Kripke means by a 
property being essential to an object, the solution presented above will not do 
the job. For to say that a property \( P \) is essential to gold is to say that even non-
actual pieces of gold have property \( P \) in every possible world, or at least in every possible world that they exist.\(^{12}\) If one believes—and I think Kripke would agree with this—that there could be instances of gold numerically distinct from each actual piece of gold, then restricting the essentialist claim to the essential properties of actual pieces of gold makes it weaker than what Kripke means by it.

2. A modified version of the previous suggestion. One remedy is this. “By gold being essentially so and so, what I mean is that each piece of gold, be it actual or not, is essentially so and so.” This suggestion is close to Monte Cook’s (1980) solution to the problem of making sense of the notion of rigidity of non-singular terms. Cooks asks, “If ‘cat’ is a rigid designator, what does it designate?” His answer: ‘cat’ designates the class of all cats in all possible worlds.

Whether or not this solves the problem of making sense of the rigidity of ‘cat’ is not the issue here: my worry is the reference of general terms. If,

\(^{12}\) By a ‘non-actual piece of gold’ I mean an object which does not exist in the actual world but exists in at least one possible world and is gold there (whatever being gold in another possible world may mean). The other understanding of the phrase is not relevant here: I am not concerned with actual objects, if any, which are possibly but non-actually gold—that could be of some interest if we were to consider statements like the (false) claim that jewels are essentially gold.
independent of the question about what is designated by ‘cat’, we could make sense of a property (say cathood) being essential to cats, then Cook’s solution would prima facie have a chance to do the job. But my question is not “What it is for ‘cat’ to be rigid?”; rather, it is the more basic question, “What is designated by ‘cat’?”. To answer this question, which is Cook’s title question, one cannot appeal to alleged essential properties of cats because attribution of essential properties to cats \textit{presumes} that we know what is designated by ‘cat’ in other possible worlds.

3. The abstract kind / the same old universal. At least since the time of Donnellan’s (1973), philosophers have been aware that the notion of rigidity, originally defined by Kripke (1980: 48) for singular terms, has no obvious analogue for common nouns—an important gap in Kripke’s argument for his claim that many theoretical identifications (say ‘heat is the motion of molecules’) are necessary, if true. Donnellan (1983: 91) suggests that

construing terms for kinds, such as ‘water’, ‘tiger’, etc., as rigid designators \textit{and} giving the Kripke-Putnam view the best run for its money is to think of them as

\footnotesize{\textsuperscript{13} See also Soames (2002: 251-258) for a criticism of the essentialist approach to defining rigidity for general terms (with no reference to Cook). I find Soames’s criticisms sound and applicable to Cook’s view. For a criticism of Soames on the essentialist approach see Gómez-Torrente (2006).}
what Mill calls “abstract” nouns. ‘Tiger’ is not to be thought of as designating its extension. Rather, it designates (is the name of) a certain species. ‘Water’ designates, not *its* extension—puddles, pools, ponds of stuff—but the substance, water. Thought of in this way, kind terms are in one way more like proper names: they designate a single entity, albeit an abstract entity—a species or a substance in these cases.\footnote{Using Mill’s terminology here is not totally unproblematic: according to Mill (A System of Logic, I.ii.4), not every abstract name is a singular term. In fact, he suggests that one may consider abstract names as “neither general nor individual, and to place them in a class apart” (1872: 30).}

In order to define rigidity for general terms, one then applies the original definition of rigidity to the corresponding singular terms: a general term is **rigid** [according to this suggestion] iff, as an abstract singular noun, it designates the same entity in every possible world. More generally, following Soames (2002), one may think of a way *t* of mapping general terms to singular terms, and define a general term *K* to be **rigid** iff *tK*, as a singular term, is rigid. Here the referent of *tK* need not be the property of being a *K*: thus Kaplan (1973: 518n31) suggests that ‘red’ rigidly designates not the property being red, but the colour red.

It has been argued that this suggestion has the undesirable consequence of the collapse of the rigid/non-rigid distinction for kind terms, for, under this interpretation of referents of general terms, even ‘bachelor’ turns out to be
rigid: it denotes one and the same abstract kind in every possible world. This is examined by Donnellan (1983: 91) and, for the case of the more general strategy mentioned above, by Soames (2002: 259 ff). However, not every philosopher agrees that the suggestion in question has this drawback.

Defending this understanding of what it is for a kind term to be rigid, Joseph LaPorte rejects the claim that this account makes every kind term rigid. Thus LaPorte argues (2000: 297 f) that although the phrase ‘the insect species that is typically farmed for honey’ actually picks out the same kind as the one that ‘the honeybee’ actually picks out, it could have been the case that bumblebees, and not honeybees, were usually farmed for honey; and in a possible world where this is the case, ‘the insect species that is typically farmed for honey’ refers to the kind bumblebee (while, as in any other possible world, ‘the honeybee’ — as used by us — still denotes the kind honeybee). Hence, if this is what rigidity for kinds is, not every kind term is rigid. And then one can go on and try to argue that a true statement of identity in which the identity sign is flanked by rigid kind terms is necessarily true. Insofar as the application of rigidity is concerned, LaPorte argues, the above generalization of the notion of rigidity to kind terms works. (See also LaPorte’s (2006) reply to Schwartz (2002).)

One might, then, think that this may solve the transworld identity problem for kinds (or the transworld identity problem for the properties of the form the property of being a K, or K-ness, where K is a kind term): the
referent of ‘gold’, for example, is the substance, or the abstract kind, gold—or maybe another abstract object if one is able to provide another (preferably systematic) way of mapping kind terms to singular terms.

For the purposes of this chapter, the main question to ask in evaluating this solution is: What is the “abstract kind”, gold? His main concern being rigidity, LaPorte is justly silent on this issue, and explicitly so: “Just what sort of abstract entity might be designated by a natural kind term is a matter of dispute, one that for the purposes at hand does not require a resolution” (2000: 311n2). But, as in the case of the suggestion (2) above, my concern here is more basic than what it is for ‘gold’ to be rigid—my worry is a worry about the extension of ‘gold’ in non-actual possible worlds. To make sense of a statement to the effect that gold is essentially so and so, we need to have an answer to the transworld identity question, ‘What is to be an instance of gold in a possible world?’ The the-abstract-kind suggestion does not resolve this worry. To say that being gold is a matter of instantiating the abstract kind gold is just to push the problem back.

Recall that when Kripke argues that so and so is essential to, say, the kind cat, what he does is argue that, individually, cats are essentially so and so—this is how, in Section 1, I arrived at the definition of what is essential to a kind. That definition requires an answer to the question about the extension of, say, ‘cat’ in non-actual possible worlds. Whatever other merits it might have, to say that ‘cat’ refers to the abstract kind cat does not answer this
question. Of course the essentialist might say that the definition I presented is not a good one; but then he has to present his own definition, and also to explain how, according to his definition, the kind *cat* being essentially so and so follows from *cats* being essentially so and so.

Likewise, if one says that in every possible world the term ‘gold’, as we use it in the actual world, refers to the same *property* (that of being gold), this does not solve the relevant transworld identity problem for kind-essentialism—we are really back to our main metaphysical question: What is the extension of this property in a non-actual possible world? The same question obtrudes if one says that ‘gold’ picks up the same old universal, Gold, in every possible world wherein this universal exists or instantiates: we need a positive account of what that old universal is and, more to the point, we need to know what it is for an individual to be an instance of gold in a non-actual possible world, and how does *gold* being essentially so and so follows from *instances of gold* being essentially so and so.\(^{15}\)

\(^{15}\) But perhaps we have an extra problem for the the-same-old-universal version of the solution. It seems that even if one believes in the existence of universals, one might deny that corresponding to *every* kind there is a universal—is there a universal *Cat*?
CHAPTER 2
“That kind of thing”

So far I have argued that there is a question the essentialist about kinds has to answer before he can start arguing for his doctrine. The question is that of the transworld identity for kinds: What do you mean by, say, ‘cat’ when you say that something is essential to the kind cat? In the last section of Chapter 1 I rejected a number of rather obvious attempts to answer this question.

Here I will examine a more promising attempt to solve the transworld identity problem for kinds. The idea of this attempt is developed in Putnam’s modern classic, “The meaning of ‘meaning’” (1975a), and can also be found in Kripke’s Naming and Necessity (1980), from a passage of which the title of this chapter is taken.¹ The idea is that, in every possible world, something is an instance of a kind \( K \) just in case it is of the same kind of the actual-world \( K \)s.

After presenting the idea in Section 1, I will offer two arguments against it. Section 2 focuses on the interest-relativity of this way of transworld identifying a kind, and argues that this goes against the idea of metaphysical essentiality. Section 3 presents another argument, based on Putnam’s later thoughts (1992) about the suggestion.

¹ Kripke (1980: 122): “The original concept of a cat is: that kind of thing, where the kind can be identified by paradigmatic instances.”
1. Sameness of kind.

In the last chapter I argued that the the-abstract-kind idea for solving the transworld identity problem for kinds does not work because it does not tell us what the abstract kind in question is in a non-actual possible world. But perhaps we need not know the answer to that question—perhaps all we need is a well-defined relation of the sameness of the kind defined for any two objects. (If need be, after getting such a relation we can abstract the notion of a kind out of it à la Frege: perhaps kinds can be thought of as equivalence classes, spread across possible worlds, of objects under the relevant same-kind relations.) Let’s go into the details.

In his (1975a), Putnam develops the idea that being an instance of a natural kind $K$ in a possible world is a matter of bearing a suitable relation to the stuff that we, in the actual world, call by the name $K$. Thus, in every possible world, an object lies in the extension of ‘gold’ if and only if it is of the same kind as the actual-world gold. Putnam’s most worked-out example is water, and the relevant relation is called ‘sameL’, the same liquid as. Suppose we know, or at any rate have settled the question of, what we mean by ‘water’ in the actual world—e.g., suppose we had previously pointed to a glass and said, as a matter of ostensive definition, ‘this [i.e., the liquid in the glass] is water’. Then, according to Putnam (1975a: 231), the relation sameL gives us the extension of water in every possible world:
(For every world \( W \)) (For every \( x \) in \( W \)) \( (x \) is water \( \equiv \) \( x \) bears same\(_L\) to the
entity referred to as ‘this’ in the actual world \( W_1 \)).\(^2\)

Note that, by virtue of having a biconditional in its matrix, the above formula
not only tells us that water has its same\(_L\)ish properties essentially, but also says
that bearing same\(_L\) to an actual instance of water is an essence of water, in the
sense discussed in Chapter 1: not only it is the case that in every possible
world every instance of water has these properties, but also having these
properties is sufficient for being water. In his (1975a), however, Putnam
almost always just uses the left-to-right direction to show that something is
not water.

It seems that at some places in Naming and Necessity something like this
idea is at work—here is an example: “Further, there might be a substance
which has all the identifying marks we commonly attributed to gold and used
to identify it in the first place, but which is not the same kind of thing, which is not
the same substance” (1980: 119, my italics; cf. the example of polywater in 1980:
129).

The solution is not yet complete, for we have an unanswered question: What
is same\(_L\)? To say what we have said about it so far, that \( x \) bears the relation

\(^2\) Though not directly related to the current discussion, note that it seems that this account
cannot make sense of those kinds, if any, which are not instantiated in the actual world.
same\(_L\) to \(y\) iff \(x\) is the same liquid as \(y\), is just a matter of re-labelling. What I take to be the definition of this relation is presented on pages 238-9 of Putnam (1975a): “\(x\) bears the relation same\(_L\) to \(y\) just in case (1) \(x\) and \(y\) are both liquids, and (2) \(x\) and \(y\) agree in important physical properties.”

2. Importance and essentiality.

But, of course, this cannot be the end of the explanation: we have to be told which properties count as important.

“Importance is an interest-relative notion”, Putnam reminds us (1975a: 239).\(^3\) Now I think from this it follows that the that-kind-of-thing idea for solving the problem of transworld identity for kinds does not lead to typical essentialist results, even if it could enable us to make sense of \(K\) is essentially so and so. For suppose that the original sample of water we used for the ostensive definition of ‘water’ was colourless and thirst-quenching. (Alternatively, suppose that the great majority of paradigmatic instances of water had these properties.) Now if what we are interested in about liquids is just whether they are colourless and thirst-quenching—if, for some particular purpose, these are all and the only important properties of liquids—then in every possible world every liquid with such properties is water, be it \(H_2O\) or

\(^3\) Among other authors, de Sousa (1984) emphasizes the interest-relativity of kinds. See also Section III of Khalidi (1998).
not (Sprite™ in the actual world is an example). This is very far from the well-known essentialist claim that water is essentially H₂O.

This argument uses the sufficiency condition of the suggested criterion for waterhood. One might think of blocking the inference by weakening the claim: one might think of accepting only the left-to-right direction of the displayed formula of Section 1, giving only necessary conditions for being water in a possible world. If one goes this way, then, even accepting the importance of colourlessness and thirst-quenchingness of liquids, one does not commit oneself to the view that every colourless thirst-quenching liquid is water; so there might still be hope for justifying the claim that water is essentially H₂O. However, this move is not satisfactory, for two reasons.

First, from the modified formula it still follows that water is essentially thirst-quenching and colourless, and, presumably, the typical essentialist will not be happy with this. Even if the essentialist is happy with the essential attribution of these properties to water (say because—let us assume—the disposition to quench the thirst of human-like creatures follows from the microstructure of water which is, in turn, typically considered to be essential to water), I think the essentialist will be unhappy with a general aspect of the weakened formula, viz. the idea that all “important” properties of the original or paradigmatic samples of water are essential to water. To see this, suppose that all of the original samples of water were, at one time or other, either in a glass or were bottled. Also suppose that being-in-a-glass-or-being-bottled is
considered an important property of a liquid. Now it follows from the weakened formula that every instance of water is essentially such that, at some point in its history, it is either in a glass or in a bottle, and this, I think, the typical essentialist will find unacceptable: a bottleless, glassless possible world is not thereby a waterless one.

Secondly, and more importantly, the weakened formula does not solve the transworld identity problem: to make sense of essentialism about water, we need to know what water is in a possible world, and the weakened version of the formula does not tell us that—it just says what properties water cannot lack. This is less than what we wanted from a solution to the transworld identity problem.

A better move is this. The essentialist who favours the Putnamian solution might remind us that the essentialist’s claim is a conditional one: if water is actually H₂O, then water is essentially H₂O; if important properties of water are not its microstructural properties but its colourlessness and thirst-quenchingness, then water is not actually H₂O in the first place, and the essentialist’s claim about water still holds. But my point is that if one accepts the that-kind-of-thing solution for the transworld identity problem for kinds,

---

4 Note that the first assumption is true about Putnam’s story (1975a: 231) about the ostensive definition of ‘water’: the original sample was, in fact, in a glass. Also, for someone who considers thirst-quenchingness an important property of water, the second assumption is not terribly artificial.
then whatever property of a thing one deems important is to be considered essential to the kind of that thing, and the totality of such important properties constitutes an essence of the kind. Given the interest-relativity of importance, this seems odd, for, pre-analytically, the (metaphysical) essence of a kind is supposed to be what the kind “really is”, something loftier than what is picked out by our interests, something that transcends our state of scientific knowledge, something independent of our minds.

The essentialist can try to get rid of the interest-relativity of importance. Via appealing to the notion of True Science, he might try to substitute ‘whatever properties of a thing one deems important’ with ‘whatever properties of a thing that are important’: the essentialist might say that important properties of a thing (or a kind of thing) are those which are so considered by, say, the true, Ultimate Physics or Chemistry. Thus the essentialist might suggest that a property is important just in case at least one formulation of one of the laws of the True Science uses a name for that property in a non-superfluous way.5

Now there are qualms about the coherence of the idea of True Science. Also, True Science, even if it exists, might say nothing about samenesses and similarities. Yet, for the sake of argument, let us set aside such anti-realistic

---

5 The qualification ‘in a non-superfluous way’ is here to prevent every property from being referred to in at least one statement of every law. For example, one may formulate Newton’s first law thus: ‘No object, whether or not it is grue, accelerates unless a force acts on it.’
and nominalistic scepticisms and assume that there is such a thing as True Science, and that it does talk about samenesses. Still it seems plausible to say that it is in general hard to know if one’s science *has* reached the level of True Science. In particular, it seems to be uncontroversial that our current sciences have *not* reached that level: even though in some cases there might be reasons to think that the properties considered important by our current sciences have counterparts in True Science, it seems outrageous to say that, for a given kind, we have discovered *all* of its really important properties. (As Putnam puts it in the particular case of sameness for liquids, “the relation same$_L$ is a *theoretical* relation: whether something is or is not the same liquid as this may take an indeterminate amount of scientific investigation to determine”$^6$.)

Thus, if importance is to be defined by reference to the True Science, we might well be unaware of the important properties of kinds. If, in response to the transworld identity question for kinds, we go for the that-kind-of-thing idea, this means that we might well be unaware of the necessary properties of a kind. But then by conditionalizing his claim the essentialist is not giving us a very substantial piece of information—it is not highly informative if I give you a set of properties of $K$s and tell you “*If these properties are necessary for being a $K$, then in no possible world is anything which lacks these properties a* $^6$ Putnam (1975a: 225). I was directed to a similar passage in Putnam (1973b) via Zemach (1976: 124).
"THAT KIND OF THING"

K." Yet, it seems to me, this captures the content of the essentialist’s conditional claim—here is Kripke (1980: 125, my square brackets):

Any world in which we imagine a substance which does not have these properties [e.g., the property of being an element,) is a world in which we imagine a substance which is not gold, provided these properties form the basis of what the substance is.

Unless the antecedent of this conditional is established, the essentialist is telling us nothing. Or perhaps qua semanticist, the essentialist might be seen as teaching us the use of ‘essential’; qua metaphysician, he is telling us nothing.7

So much for my own worries about the that-kind-of-thing idea for solving the transworld identity problem for kinds. The rest of this chapter will provide an exposition of the relevant parts of Putnam (1992). Apart from presenting

---

7 When talking about this issue in Naming and Necessity, Kripke usually talks about conditionals of the form discussed above (see, e.g., pp. 121 on tigers, 124 on gold). There is also this claim, with no further explanation: "Whether science can discover empirically that certain properties are necessary of cows, or of tigers, is another question, which I answer affirmatively” (1980: 128). For an examination of the idea that science can discover essences see Canfield (1983).
another argument, I hope the next section will shed some light on the
dialectic of Putnam’s (1992), and will also clarify some points in his (1975a).

3. Later Putnam.

As evidenced by one of the quotations I gave, Putnam (1975a) is perfectly
aware of the interest-relativity of sameness relations. Perhaps one reason that
he was not moved by what I have argued for— that same gives us less than a
full-fledged metaphysical essentiality — is that, even in “The meaning of
‘meaning’”, Putnam is really not interested in essential properties in the
metaphysical sense that Kripke is interested in. (See Hacking (2007) for more
on this.) With the wisdom of hindsight, we now know it more clearly that in
that paper Putnam talks about possible worlds only in connection with the
discussion of Kripke’s views, and he simply takes possible worlds as playing
the rôles of hypothetical distant planets like his celebrated Twin Earth. He
tells us that in his (1975a), he “did not think through the consequences” of
possible-world talk (1992: 443). However that may be, let us now briefly
consider Putnam’s (1992) worries about the that-kind-of-thing idea for
solving the transworld identity problem for kinds.

It has been argued that, even within the actual world, there is no
sameness-oriented solution to the problem of the identity of a kind in the
realm of biology;\(^8\) so let us just see if the suggestion works for substances—after all, unlike Kripke, who frequently considers small and large felines, Putnam (1975a) mainly talks about things which are studied by chemistry and physics.

Immediately after saying that importance is interest-relative, Putnam writes (1975a: 239),

Normally the ‘important’ properties of a liquid or solid, etc., are the ones that are *structurally* important: the ones that specify what the liquid or solid, etc., is ultimately made out of – elementary particles, or hydrogen and oxygen, or earth, air, fire, water, or whatever – and how they are arranged or combined to produce the superficial characteristics.

In fact, it seems that for Putnam structural properties are just what is predominantly *meant* by ‘important’ — note the ‘i.e.’, and not ‘e.g.’, in:

“Suppose, now, I have not yet discovered what the important physical properties of water are (in the actual world) — i.e. I don’t yet know that water is H\(_2\)O” (1975a: 232). Even ages before the new science, when, say, Archimedes said that something was gold, “he was saying that it had the same hidden structure (the same ‘essence’, so to speak) as any normal piece of local gold”, says Putnam (1975a: 235).

\(^8\) The locus classicus of such a criticism is Dupré (1981). See Chapter 3 of LaPorte (2004) for criticism of both early Putnam and Dupré.
Enter Putnam (1992). Reviewing his (1975a), Putnam says that one can take the criterion of substance identity to be “has the same physico-chemical composition and obeys the same laws” (1992: 435). At first sight, one might think that there is some redundancy in this criterion—is it not the case that if two substances have the same composition they also obey the same laws? (And is this not one of the facts that make the microstructure important?) Would not having the same microstructure be enough for substance-identity? Moreover, in his (1975a) Putnam just talks about microstructures, not microstructures and laws.

But there is no redundancy—at least not an obvious one. In fact, Putnam says (1992: 435) that Kripke’s not mentioning laws is a defect of Kripke’s version in comparison to Putnam’s. Why is that so? Because Putnam believes that the following two scenarios are possible. He does not explicitly talk about the differences and the relationships between the two, but the possibility of each one of them shows the irredundancy of the second criterion.

Scenario 1 (1992: 453n8) is that the composition of a material object does not determine its behaviour. Discovering that substances of the same composition can obey different laws has grave consequences indeed: “our whole picture of the world—no just our philosophy—will have to be revised”
in the aftermath of such a discovery. Now, what if in a possible world, wherein the microstructure does not determine the behaviour of substances, we have some H$_2$O molecules which do not behave like the actual-world water (they do not fall as rain, they do not quench thirst, etc.)? Are those H$_2$Os instances of water? Putnam finds Kripke’s presumably affirmative answer “utterly arbitrary”. He does not elaborate on his use of ‘utterly arbitrary’; besides, utter arbitrariness may, in itself, not be a serious defect. His point would appear to be that, in such a circumstance, the way we fix the reference of ‘water’ has nothing to do be what water “really is”; hence we cannot use this way of fixing the reference of ‘water’ (and solving the transworld identity problem for water) to make sense of metaphysical necessity.

The claim that the composition of a thing need not determine its behaviour sounds plausible: as Putnam observes (1992: 453n8), it is certainly not a priori that objects with the same microstructure obey the same laws, and perhaps there is no incoherence in the idea that things of the same microstructure behave differently. An essentialist who wants to escape Putnam’s charge of utter arbitrariness—the charge of not capturing what things “really are”, in my interpretation of Putnam—has to deny the prima facie possibility that

---

9 Cf. Kuhn (1990: 310). Kuhn thinks that, upon the discovery that the watery stuff on Twin Earth is not H$_2$O, the report the visitors from the Earth send home must be something like, “Back to the drawing board! Something is badly wrong with chemical theory.”
Putnam is talking about. At any rate, here I want to make a point about the relationship between Putnam’s arguments in his (1975a) and his claim in his (1992) that it is not necessary that the sameness of composition comes with the sameness of behaviour: I think that, unlike what might appear to be the case, the Twin Earth thought experiment does not show the possibility of Scenario 1 (nor do I think that it was intended to do so).

First, Twin Earth is not a possible world: it is a planet (well, a hypothetical one), “somewhere in the galaxy”.\(^\text{10}\) If there is such a planet, then there is a place *in our world* wherein liquids which are not H\(_2\)O have all the superficial properties of H\(_2\)O. Second, and more importantly, on Twin Earth we do not have H\(_2\)O with a non-water-like behaviour; rather, we have a *non*-H\(_2\)O substance with water-like behaviour. Nor does the existence of Twin Earth directly show that laws are less universal than what was previously supposed. What it does show is the multiple realizability of the superficial properties of water. (For Putnam’s pragmatic treatment of what may be considered as an actual example of multiple realizability of superficial properties of substances, see the case of jade in his (1975a: 241).)

Putnam’s Scenario 2 (1992: 443) is that there is a world that obeys laws different from those which govern our world. Here the difference between

---

\(^{10}\) Putnam (1975a: 223); cf. (1992: 443). That a spaceship from the Earth visits the Twin Earth shows that the latter is at least not a possible world, distinct from ours, in the sense of possible world advocated by Lewis (1986).
the laws governing that world and those governing the actual world results in a behaviour of H$_2$O molecules different from the behaviour of actual H$_2$Os (or actual water). Putnam asks, “Is it clear that we would call a (hypothetical) substance with quite different behavior water in these circumstances?” The intended answer to this rhetorical question is No.

The morals of the two scenarios are the same. If there is a world wherein a substance’s behaviour is not determined by its microstructure, or if there is a world obeying different physical laws—and both seem to be “metaphysically possible”, according to Putnam—then “the question, ‘what is the necessary and sufficient condition for being water in all possible worlds?’ makes no sense at all. And this means that I now reject ‘metaphysical necessity’” (1992: 443).

As in the case of Scenario 1, I think if Scenario 2 is possible then we have a good case against the coherence of the idea of metaphysical necessity (and, a fortiori, against kind-essentialism). But it might appear that here the case is weaker than that of Scenario 1: it seems that the essentialist can deny that Scenario 2 is possible. Thus for Brian Ellis (2002: 101), “the laws of nature spell out the essential properties of the natural kinds […] the laws of nature are all metaphysically necessary”. Therefore an essentialist like Ellis may say that water cannot be governed by different laws, and surely presenting a metaphysically impossible scenario is not a good argument against a doctrine.

---

There are two problems with this essentialist move. A minor one is that the properties Putnam is talking about (falling as rain, quenching thirst, etc.) do not seem to be essential properties of water on any account; so Ellis should see no problem in thinking of the possibility of water lacking these properties—it seems that, insofar as Scenario 2 involves such properties, Ellis should admit that this scenario is possible. However, perhaps this difficulty is not insurmountable: perhaps Ellis could say that even a non-thirst-quenching etc. substance is water if it is H₂O, because that is what water really is. Perhaps this is something one would expect to hear from Putnam (1975a).

But now a major difficulty obtrudes: as I understand it, Putnam’s (1992) real challenge for the essentialists is to say what they mean by ‘water’ when they talk about substances in non-actual possible worlds; this question has to be answered prior to making the claim that so and so is essential to water. The question is how to solve the transworld identity problem for water, and before solving this problem all the essentialist knows is what water in the actual world is. Simply to say that by ‘water’ in every possible world we mean H₂O seems to be utterly arbitrary if this does not come with a justification. And what would a justification look like? If the essentialist is tempted by the that-kind-of-thing idea, then he will be faced by the problems discussed earlier in this chapter.
The that-kind-of-thing suggestion is the only attempted solution to the problem of transworld identity for kinds that is developed by Putnam in his (1975a) and criticized in his (1992). Given that Putnam’s conclusion in the latter work is the rejection of metaphysical necessity altogether, it seems safe to infer that Putnam considers this suggestion as the only plausible candidate for a solution to the transworld identity problem for kinds. However, in the next chapter I will examine another, perhaps even more promising, idea for solving the problem of transworld identity for kinds: stipulation.
CHAPTER 3  
Stipulation

The transworld identity problem for kinds asks about the references of kind terms when we use them to refer to groups of things in non-actual possible worlds. I introduced the problem in Chapter 1, and argued that the kind-essentialist—someone who wants to attribute the necessary possession of some interesting properties to some kinds—cannot make sense of his view, let alone defend it, prior to solving this problem. In the last section of that chapter I also rejected some ideas for solving the problem, and in the previous chapter I argued that the Putnamian idea of the sameness of the kind does not provide a solution either—at least not a solution to the satisfaction of the essentialist. The last idea that I will examine in regard to the transworld identity problem for kinds is the suggestion that one might solve, or dissolve, the problem analogously to the case of transworld identity for individuals, viz. by stipulation.

Unlike my discussions of the previous suggestions for solving the transworld identity problem for kinds, here I will not present a (what I would consider) decisive argument against the suggestion at hand. Part of the reason is that in the course of my argument I will appeal to a certain theory of the identity of properties, and this theory—Sydney Shoemaker’s causal theory of properties—is not uncontroversial. Thus perhaps not every essentialist will
agree that my argument against the stipulative response is sound, even if he admits that it is valid. (I will briefly argue for Shoemaker’s theory in this chapter, but, as this is not my main concern here, I will not consider arguments against it in any detail.)

The other, more serious difficulty is with the very idea of stipulation: it is not quite clear what stipulation is exactly in the context of modal talk. Therefore, I acknowledge the possibility that my argument here might turn out to miss the real point of stipulation, should there be a clear theory of stipulation. What I will do here is argue that stipulation, as I understand it, requires that a condition be met about identity—a condition which, I will argue, is not satisfied in the case of a stipulative attempt to solve the transworld identity problem for kinds. This strategy may serve as an argument against the stipulative move. On the other hand, an essentialist who wants to stipulatively solve the transworld identity problem for kinds has to say more about stipulation: What is it for us to stipulate something in a possible-worlds discussion? How does stipulation work? For what kinds of questions is stipulation permissible? I think even in the case of the transworld identity problem for individuals we need to hear more about stipulation: it seems to me that when Kripke forcefully argues that there is an obvious stipulative answer to the transworld identity problem for individuals (like the case of the transworld identity for Nixon in 1980: 44 ff), much of the
argumentative work is done by means of using italics and emphatic phrases like ‘he himself’ — we are not really told what stipulation is.

This chapter is structured as follows. First, in Section 1, I present a stipulative suggestion for solving the transworld identity problem for kinds. I then argue in Section 2 that stipulation about an entity presupposes some sort of primitiveness of the identity of that entity. With recourse to Shoemaker’s causal theory of properties, which I briefly argue for in Section 3, I will then (Section 4) argue that the identities of properties corresponding to natural kinds are not primitive in the required sense.

If my arguments are sound, then the overall result is that one cannot solve the transworld identity problem for natural kinds by means of stipulation. This, then, strengthens my scepticism about kind-essentialism.

Two remarks on the natural-kind talk of this chapter has to be made now. First, Shoemaker’s causal theory of properties might not be true of properties corresponding to non-natural kinds (say the property of being grue), and thus my argument might be ineffective against the stipulative suggestion for solving the transworld identity problem for non-natural kinds. However, in the context of an argument against the kind-essentialist, this does not weaken my position since typical kind-essentialist’s claims are about natural kinds (perhaps tiger, gold, water, electron), not non-natural ones.

Lastly, of course there are important and difficult questions regarding the meaning of ‘natural kind’ and about the existence and the ontological status
of natural kinds. Also, more particularly, one may ask whether this or that

group of things form a natural kind. Nonetheless, for the purpose of my
argument against kind-essentialism, I need not answer these questions; I will
take the kind-essentialist’s favourite theory of natural kinds for granted, and I
will accept the kind-essentialist’s claim about any group of objects $K$ that it is

a natural kind—all this I grant the kind-essentialist so long as it is agreed that

if $K$ is a natural kind then $K$ plays, or $K$s play, a causal or explanatory rôle.

1. The stipulative idea.

Consider the following passage from *The Blue Book*:

> Someone says, he imagines King’s College on fire. We ask him: “How
do you know that it’s *King’s College* you imagine on fire? Couldn’t it be
a different building, very much like it? In fact, is your imagination so

absolutely exact that there might not be a dozen buildings whose

representation your image could be?” — And still you say: “There’s no
doctor I imagine King’s College and no other building”.

---

1 Wittgenstein (1958: 39). I believe it was from a paper by Crispin Wright that I learnt about
this passage; unfortunately, I cannot locate that paper now, and professor Wright’s memory

is not fresh about the issue.

Wittgenstein uses the example in a different context. To motivate the stipulative idea for
solving the transworld identity problem, I will assume that ‘King’s College’ names just *one
building* — a false assumption apparently also made by Wittgenstein.
The advocate of the stipulative suggestion loves this passage. Here I take
Wittgenstein to suggest that it is part of the story, as it is told by he who is
reporting on his own imagination, that what was imagined to be on fire was
King’s College and no other building, however similar to King’s College it
might be. Prima facie, asking for a criterion of the identity of the building in
question does not make much sense. Thinking of the content of the
imagination or dream as a partial description of a possible world (and let us
assume that the description does not involve any contradiction), we may
simply stipulate that it was King’s College that was imagined to be on fire, in
the same way that Kripke (1980: 44 ff) urges that when we talk about things
that could have happened to Nixon, it is part of the story that we are talking
about Nixon, and no one else, as I explained in Section 2 of Chapter 1.

Above I treated King’s College as an individual. Now consider the
predicate ‘is on fire’. It does not seem to make much sense to ask, “Are you
sure that in your imagination King’s College was on fire? Couldn’t it be the
case that it was another phenomenon that you imagined? Couldn’t it be
another process, not burning, which King’s College was undergoing?” It
seems that in the same way that we may stipulate that the individual we are
talking about is King’s College (and no other individual), we may also assume
that the property we are talking about is that of being on fire (and no other
property). It seems that we may simply stipulate that we are talking about the property of being on fire, not any other property.

Perhaps more significantly for this discussion is a passage from *Naming and Necessity*. Here Kripke first reminds us of his position on what we may stipulate about a possible world while talking about individuals therein:

“Don’t ask: how can I identify this table in another possible world, except by its properties? I have the table in my hands, I can point to it, and when I ask whether *it* might have been in another room, I am talking, by definition, about *it*. I don’t have to identify it after seeing it through a telescope.” Then he says,

If I am talking about it, I am talking about *it*, in the same way as when I say our hands might have been painted green, I have stipulated that I am talking about greenness.²

² Kripke (1980: 52-53). It was Macbeth (1995: 263) that drew my attention to this passage.

Kripke’s talk about pointing to the table might be misleading here—it might suggest that ostension is crucial to enabling us to talk about objects in other possible worlds, or that the observability of the thing in question (here the table) plays an important rôle in the legitimacy of stipulation about it. I take Kripke’s point to be that, like proper names, expressions likes ‘it’ or ‘this table’ are rigid designators, designating the same object (if any) in every possible world—cf. Putnam (1975a: 231): “We may also say, following Kripke, that when I give the ostensive definition ‘*this* (liquid) is water’, the demonstrative ‘this’ is *rigid*."

It seems that Kripke finds the transworld identity problem for properties (here the transworld identity of the property of greenness) even more trivial than that of the transworld identity problem for individuals—he is using the former to illustrate that there is no real content to the latter. But then for any given natural kind $K$ we may consider the property of being an instance of $K$, and it will seem plausible to say that, for Kripke, there is no real problem about the transworld identity for kinds either. It seems that in Kripke’s view the transworld identity problem for kinds is only a pseudo-problem, perhaps even more of a pseudo-problem than the transworld identity for individuals. (Kripke’s position on property-stipulation is clearer in his (1971); I will give a quotation at the end of this chapter.)

So here is the issue I have to deal with if I want to maintain that the transworld identity problem for properties is a genuine problem. I have explained in Chapter 1 that, working within the standard, possible-worlds semantics, we may assume that there is a trivial answer to the transworld identity problem for individuals—so trivial that it renders the problem into a pseudo-problem. Now, given the fact that Kripke himself thinks that in considering a counterfactual situation we may stipulate that we are talking about a certain property (say greenness), and given the plausibility of Wittgenstein’s example (the modified, on-fire version), can’t we say that the same sort of trivial solution works for the case of the transworld identity problem for kinds? Can’t we just stipulate which property we are talking
about while considering a possible world with certain objects having a certain property? By switching from a kind to the property of being an instance of that kind, this idea would solve the transworld identity problem for kinds.

I have already argued (Section 2 of Chapter 1) that the standard semantics does not allow us to do so—in talking about a possible world, we have to specify the extensions of general terms in every possible world, unlike the case of the reference of individual constants which are the same in every possible world in which they refer. But perhaps this is just a problem with the semantics, and not a real metaphysical problem with the transworld identity of kinds? If we forget about the formalities of the semantics, what would prevent us from property-stipulating in the just the same way that we stipulate about individuals so easily? Kaplan (1975: 723) demonstrates that even if one’s modal logic is haecceitistic, one may interpret cross-world identity as a mere façon de parler; here one might ask the reverse question: despite the fact that our standard modal logic is not haecceitistic with respect to properties, why cannot we commit to haecceitism with respect to properties? In the final analysis, maybe even the transworld identity problem for individuals does not have the trivial solution Kripke says it does; but, at the intuitive, informal level, maybe property-stipulation is no less legitimate.

---

3 Roughly speaking, this means that it is unproblematic to talk of one and the same object existing in different possible worlds. See Appendix 2 for a brief discussion of haecceitism.
than object-stipulation? Can we solve the transworld identity problem for kinds by stipulation?

2. Stipulation: a *sine qua non*.

When is stipulation legitimate? Surely not in *all* cases. Under normal circumstances, I cannot stipulate that Travis is the tallest student in Martin’s class. Normally, if I believe that Travis is among the students in the class, I also believe that there is a fact of the matter as to whether or not he is the tallest student in the class, and the issue is not open to stipulation. If I can stipulate that Travis is the tallest student in the class, then by doing so I will be giving a stipulative *definition* of ‘Travis’: “Let us call the tallest student in Martin’s class—whoever he is—‘Travis’.” This will just be a matter of naming by description.

Or suppose that there was a person Travis whom I met two years ago. Now I cannot point to a student in the class and say that I stipulate that *this* is Travis. Here the point is that I have a theory (maybe at a non-articulate, non-formalized level) of personal identity through time, which precludes stipulation about the identity of Travis. Let us here stick with some version of the memory criterion of personal identity— for our present purposes, it really does not matter what the theory is. As one who, perhaps implicitly or unconsciously, subscribes to this theory of personal identity, I would be committed to the view that being identical to Travis is to satisfy the relevant
memory criterion; whether or not a guy in the class satisfies this criterion is not open to my stipulation. Here the reason for the illegitimacy of stipulation is not that being Travis is a matter of fact: maybe “really” (i.e., according to the True Ontology) there is no such matter of fact. Rather, the illegitimacy of stipulation on my part is the result of my commitment to a certain theory of personal identity (and the identity of Travis in particular, according to that theory).

I cannot rationally both (1) believe, because my theory of personal identity tells me so, that to be identical to Travis is a matter of satisfying certain necessary and sufficient conditions, and (2) stipulate about someone that he is Travis. Here is why. If (1) is the case, then I believe that my theory is not silent about Travisity. Now, on the one hand, if what my theory declares to be Travis is not what I stipulate to be Travis, then I am contradicting myself. On the other hand, if I am lucky and what I “stipulate” to be Travis happens to be what my theory declares to be Travis, then I am not really stipulating: I am just stating a result of my theory; nothing in particular is done by this act of “stipulation”.

It is as if one wants to stipulate that the number 125 has a certain arithmetical property, say it is equal to 68 + 57 or that it is divisible by 6. To

---

4 I assume that identity is a symmetric and transitive relation; hence two different things cannot be both X. I take the axioms of symmetry and transitivity of identity to be part of every background theory.
stipulate the latter is to contradict oneself, for it follows from simple arithmetic (which I assume to be part of every background theory) that 125 is not divisible by 6. On the other hand, to say that one stipulates that $125 = 68 + 57$ is, to say the least, weird: 125 is, in fact, equal to $68 + 57$ prior to one’s so-called stipulation; “stipulation” is idle here. One can prove or discover that $125 = 68 + 57$; one cannot stipulate it.

Of course, for whatever reason, I might have such a passionate interest in stipulating that something is Travis that I revise, in a systematic or ad hoc way, my theory of personal identity; but then I am changing my background theory: with respect to the theory I currently hold, I cannot do both (1) and (2).

In general, assuming that we are not dealing with naming or stipulative definition, I submit that we have this requirement for the legitimacy of stipulation:

(S) It is legitimate for me to stipulate that something is $X$ only if, to my knowledge, my accepted background theory $T$ does not give a non-trivial account of what $X$ is.

(A slightly more precise formulation will follow.) If this requirement is not met, then my stipulation about $X$ will meddle with what my background theory counts as a matter of fact, and this might result in inconsistency (e.g., my theory and my stipulation declare two different objects to be Travis). If
the meddling does not lead to inconsistency it is only because the “stipulation” about the identity of X is a consequence of what T says about X, in which case I am not really stipulating.

Note that the reasoning behind S is insensitive to the kind of entity X is: the argument that led to it works equally well if X is a property, as opposed to an individual. I can stipulate that something is a given property (say the property of greenness) only if, so far as I know, my background theory does not tell me what that property is. Very likely, the satisfaction of this condition is not sufficient for the legitimacy of stipulation; but for my purposes in this chapter I only need the condition as a requirement for stipulation.

I am going to appeal to S in the case of stipulation about possible worlds—that is to say, I am going to assume that while we are talking about X in a possible world W, stipulation about the identity of X is illegitimate unless we believe that from the totality of other things we have accepted (i.e., from our accepted background theory) it does not follow what X in W is. But one might wonder if, when applied to stipulation about other possible worlds, the argument leading to S somehow works with an un-Kripkean picture of possible worlds. If the example of Travis in Martin’s class is considered to be about things in a non-actual possible world W, then perhaps—one might worry—I have argued in a way as if I was looking through a kind of modal telescope at Martin’s class in W, and as if I “saw” who in that world was the
tallest student in the class. Though in the final analysis the Kripkean
conception of possible worlds might not be the best one, it is dialectically
undesirable here to talk about possible worlds and stipulation in a way that
Kripke would not approve of.\(^5\)

So, here is an objection to the idea of extending the application of S to
stipulation about non-actual possible worlds. By talking about Martin’s class
in a non-actual possible world \(W\), we are simply, as it were, weaving a story
about the class and Travis. In describing the relevant parts of \(W\) we do not
“see” beforehand who the tallest student in the class is—as we learnt from
Kripke (1980: 44), possible worlds are *stipulated*, not discovered via powerful
telescopes. What, then, prevents us from stipulating that *Travis* is the tallest
student in the class? If we so stipulate, we do not contradict anything we
consider a fact of the matter about Travis in \(W\); we are just creating, so to
speak, facts of the matter about Travis-in-\(W\), and the very act of stipulating
that Travis is the tallest student in the class might be considered as
constituting one part of that creation. Thus stipulation about Travis in \(W\) is
not problematic. (End of the objection.)

\(^5\) For Kripke on possible worlds and stipulation see Kripke (1980: 43-47) and (1971: 146-149). I
have discussed these in Section 2 of Chapter 1 above. For a critique of the Kripkean way of
dealing with the problem of transworld identity for individuals see Lewis (1983) and § 4.4 of
Lewis (1986).
I have no totally conclusive reply here. However, this need not be destructive to my project in this chapter: note that the objection, if it works, says something about stipulation about *individuals* (here Travis), but I think I can make a stronger case for my suggested requirement S in the case of property-stipulation. I will thus make two comments—more in the spirit of seeking clarification—on the objection, and will then move on to the case of stipulation about properties.

First, arguably the objection itself is based on an un-Kripkean conception of possible worlds. Nathan Salmon argues (1996: 203 f) that it is a serious misunderstanding of Kripke to think of him as implying that “we are the masters of metaphysical modality, in the sense that it is entirely for us to decide, by ‘stipulation,’ what is metaphysically possible and what is not.” Salmon thinks that what Kripke does by saying that possible worlds are stipulated is to endorse a kind of haecceitism, a doctrine I briefly explain in Appendix 2. Now it seems plausible to think that if we could just start telling stories—logically consistent stories—about Travis and thereby construct possible worlds containing him, we would thereby be creators of modal metaphysical facts. Therefore, if Salmon’s reading of Kripke is right, then what we do when we tell a story about Travis is *not* create facts about Travis in a non-actual possible world. It is not then clear what restrictions should be set on the kinds of things we may stipulate or incorporate in our story about Travis in a non-actual possible world.
Second, and more importantly, so far as the objection goes, perhaps nothing prevents us from stipulating that Travis is the tallest student in Martin’s class in $W$. But that is because thus far we know very little about Travis in $W$. We certainly may start telling the story of someone in $W$ by saying that he is the tallest student in Martin’s class, but that is not the end of the story: the story must be “complete” — at least complete with respect to the things which are “relevant” to Travis and the class. Add more relevant details (about the genetic codes of other students’ biological parents, the laws of nature in $W$ regarding how tall the children of such parents will grow, etc.), and it might turn out that the stipulation that he is the tallest student in the class cannot be about Travis in $W$: the stipulation might clash with other things we are assuming, or stipulating, about Travis and Martin’s class in $W$. My objective here is not to attack the usual way of talking about counterfactual situations; my worry is about the range of the things we can stipulate while describing a possible world. One should be careful to ensure that one’s (partial) description of a putatively possible world is possible, at least in the sense of not being contradictory.

Perhaps we can make a stronger case for S if we switch to an example about properties, as opposed to individuals. Suppose we are talking about a possible world $W$ and a set $P$ of persons in $W$. Can we stipulate that $P$ is the extension of the predicate ‘is the tallest student in his or her class’? We cannot, at least not if we can minimally meaningfully talk about the extension
of the relation of being-taller-than in $W$. The extensions of the mentioned predicate and relation have to be related in a certain way, and as soon as we have the extension of is-taller-than (or the extension of has the height $x$) in $W$, we cannot stipulate anything about the extension of the predicate ‘is the tallest student in his or her class’. Whether or not something lies in the extension of the property of being the tallest student in his or her class in $W$ is a fact of the matter determined by the extension of other things in $W$.

Stipulation is not allowed here.

Likewise, and perhaps more clearly, regarding a set $A$ in a possible world $V$, we cannot stipulate that $A$ is the extension of the property of being self-identical in $V$, because, given our accepted theory of self-identity, it is a matter of fact whether or not $A$ is the extension of that property in the given possible world – $A$ should be the whole domain of $V$. In saying so, I am not trying to “discover” $V$ via a telescope; I am saying that, given my logical theory about self-identity, I cannot stipulate that $A$ is the extension of that property in $V$. We cannot stipulate this about $A$ in $V$, whatever our conception of possible worlds might be.

Hard as it might be to give more examples at this stage, it seems plausible to say this much: if our background theory says that, in $W$, being $X$ is a matter of satisfying certain non-trivial conditions, then we cannot stipulate that something is $X$ in $W$. (And for the purpose of my argument, I only need this to be so in the perhaps more plausible-looking case where $X$ is a property.)
Let us say that the property of being $X$ is **primitive** for a theory $T$ iff $T$ has no theorem which gives a non-trivial necessary and sufficient condition for being $X$. (As is common with such definitions, the reference to the theory is dropped when it is clear from the context.) So, what I have argued for is that if I believe that being $X$ is not primitive for my background theory, then it is not legitimate for me to stipulate that something is $X$. Again, here $X$ could be any kind of entity— but in this chapter I am mainly interested in properties.

Two things must be said about this notion of primitiveness. First, the background theory $T$ might be the totality of one’s accepted philosophical doctrines on a particular subject-matter, in which case it might be more natural to talk about *providing an analysis* of being $X$, rather than having a theorem which gives a necessary and sufficient condition for being $X$.

Secondly, it might be difficult to define what it is for a theorem to be trivial, but I think we have an intuition about triviality here. Thus to say that $X = X$ is to say something trivial, while to give a purely qualitative description of $X$ is not.

Here is an example of stipulation which, given all that I have said, is not illegitimate. Suppose we want to examine some of the properties (say the non-robotness) of Felix the cat. As suggested by Danielle Macbeth (1995: 274 f), in order to infer some essentialist claims about cats from what we observe about Felix, we have to stipulate that Felix is a paradigmatic cat. Now if, like
Macbeth, one believes that there is no fact of the matter as to whether Felix is a *paradigmatic* cat, then one can stipulate that Felix is a paradigmatic cat—in doing so, one is not thereby violating the requirement I presented.

### 3. Shoemaker’s causal theory of properties.

It is debatable whether the property of being identical to an individual is a primitive notion—Kripke seems to think that it is, evidenced by his position against sortal identities.6 (A short, related discussion can be found in the Appendix 3.)

However the case of the identity of individuals turns out, it seems that being identical to a given *property* is not something primitive. It sounds plausible to say that, at least for a large class of properties including those in which the kind-essentialist is interested, being a property is a matter of playing certain rôles—a property cannot be what it is without playing those

---

6 Reporting on Kripke’s 1968 unpublished lectures on identity, Putnam (1992: 440) writes, “Identity is a *primitive logical notion*, Kripke claims, and it is a fundamental philosophical error to think that it can or should be ‘explained’.” (Kripke’s notion of primitiveness, on which Putnam does not elaborate, may or may not coincide with what I call by the same name.)

The quote from Putnam is in the context of discussing Kripke’s rejection of the idea of sortal identity, but I cannot reconstruct Kripke’s argument from Putnam’s short report. Kripke’s most detailed published discussion of sortal identity about objects can be found in footnote 58 to his (1980: 115). Gibbard (1975: 194 ff) is an advocate of sortal identities for objects.
rôles, and two properties playing the exact same rôles are identical, inter- and intra-worldly. I will shortly present an argument for this claim, but before doing so let us briefly digress and examine a point about naming and reference.

It is safe to say that the essentialist’s favourite theory of reference for natural-kind terms is the theory developed by Putnam (1973a), based on Kripke’s causal picture of proper names described in Lecture II of Naming and Necessity. Similar to what Kripke calls an initial ‘baptism’ (1980: 96), in Putnam’s theory of reference for natural-kind terms we have introducing events, and when a speaker uses a kind term, say ‘electricity’, she is connected by a certain kind of causal chain to a situation in which a description of electricity is given, and generally a causal description—that is, one which singles out electricity as the physical magnitude responsible for certain effects in a certain way.

The italics are Putnam’s (1973a: 200). Within the actual world, the causal rôle played by a natural-kind property is crucial to the way we refer to it. The fact that we manage to refer to a property is grounded in the fact that it is the property that is responsible for certain effects. Referring to, say, electricity would be problematic if there were another property having the exact same causal powers. (The fact that in the essentialist’s favourite theory of reference
for proper names there is no rôle to be played by the named individual in the initial baptism might suggest that perhaps the identity of individuals is a primitive notion.)

Of course this is not an argument to the effect that what I claimed about the individuation of properties should be accepted by the kind-essentialist—after all, one methodological lesson of Salmon’s well known (1979) is that one should be careful about inferring metaphysical theses from semantical considerations. But perhaps the way the standard theory of reference for natural-kind terms works is good evidence for the utmost importance of the rôles played by a property in an essentialist context.

How are properties or relations individuated? My intuition is that, at least for the case of inter-world individuation which is my concern in this chapter, this is just a matter of what objects have what properties or what objects are related to each other via what relations. Instead of trying to make this idea precise, I offer an oversimplified example. Consider two possible worlds \( W_1 \) and \( W_2 \), each of which has a domain consisting of objects \( a \), \( b \), and \( c \). Here is, by fiat, the complete list of the atomic propositions which are true in these worlds: in the first world, \( a \) and \( b \) are \( F \), and \( c \) is \( G \); in the second world, \( a \) and \( b \) are \( G \), and \( c \) is \( F \). Here the “rôle” played by \( F \) in \( W_1 \) (connecting \( a \) and \( b \), and

---

7 The example is from Melia (2003: 161 f), in his discussion of haecceitism. His intuitions about the identities of \( F \) and \( G \) are different from mine.
leaving $c$ in isolation) is the same as the one played by $G$ in $W_2$, and the “rôle” played by $G$ in $W_1$ is the same as the one played by $F$ in $W_2$. It seems plausible to say that, therefore, $F$ in $W_1$ is the same as $G$ in $W_2$, and that $G$ in $W_1$ is the same as $F$ in $W_2$.

However, perhaps not many essentialist philosophers share such an intuition; so I will argue for a less general version of the view, viz. the causal theory of properties, championed by Sydney Shoemaker, according to which “genuine” properties are individuated by their nomic profiles. What I do in the rest of this section is to argue that if one thinks that possible-worlds talk involving properties is legitimate (as surely a kind-essentialist does), then one must also commit oneself to Shoemaker’s theory. Hence I will here assume that inter-world talk of properties does make sense. The result of the following discussion will be used, in the next section, to argue that for any natural kind $K$, being the property of being an instance of $K$ is not a primitive notion, and is therefore not suitable for stipulation. Hence the transworld identity problem for kinds cannot be solved by stipulation.

As summarized by Shoemaker, the view is that “properties are identical, whether in the same possible world or in different ones, just in case their coinstantiation with the same properties gives rise to same powers” (1980: 221). Shoemaker’s arguments are well known; I will just briefly review them. Shoemaker’s (1980, 1998) reason for advocating this causal theory of
properties is that were this not the case, we would encounter “disastrous epistemological consequences”: in the absence of such a theory, Shoemaker argues, one cannot explain the knowledge of properties which we seem to have—neither can one explain how we are capable of talking about properties in our language.

There are two parts to the view. The first part is that the causal powers of a property are essential to it. The second is that properties sharing all their causal powers are one and the same property. (Recall the mini worlds introduced above. Shoemaker’s view implies that if the properties $F$ and $G$ are “real” properties in a sense to be explained shortly, then $F$ in $W_1$ is the same as $G$ in $W_2$.)

For the first part, a good argument can be found in Shoemaker (1980: 217 f). Consider the property of being grue which an object possesses iff it is green and the time is before 2100 CE, or else it is blue and the time is 2100 CE or after that. (This is not Goodman’s classic definition.) Being grue is not a “real” property: a green object will lose its property of being grue at the very first moment of the year 2100, but not because some real change has occurred in the object itself. Though we might not be able to define what a “real” property is (as opposed to a “mere-Cambridge” property like that of being grue), the difference is clear enough for the argument at hand. (But see also Chapter 6 below for further discussion of the distinction.)
Now this “non-real” property, grue, changes its causal powers through time: before 2100 it has the causal powers of the property of being green, afterwards those of the property of being blue. It is intuitively clear that this does not happen to a “real” property, and therefore “[t]here is no such thing as tracing a property through a series of changes in its causal potentialities—not if it is a genuine property, i.e., one of the sort that figures in causal laws” (218). But then this amounts to real properties having their causal powers necessarily, for it is not true that something might have been a $P$ unless it could become a $P$. (The possible-worlds analyses of relevant notions that I use are as follow. Something **might have been** a $P$ iff there is a possible world in which it is a $P$. Something **could become** a $P$ iff there is a possible history of that something such that an initial segment of that history is the same as its actual history, and then it diverges from the actual history).

Note how Kripkean this picture of possibility is—recall Kripke’s (1980: 112 f) argument for the claim that QEII necessarily has the biological parents she has: she has the property of having so and so biological parents right from the moment that she comes into existence; hence she cannot become a person with different biological parents. Hence the necessity of the origin, according to Kripke.

It is the argument for the second part of Shoemaker’s thesis that is epistemological. (My exposition draws on part two of Alexander Bird’s (2005).) First recall the construction used in Chisholm’s (1967) celebrated
example of Adam and Noah. Chisholm makes the plausible assumption of the transitivity of the transworld identity. Then, starting from the actual world and switching the qualitative properties of Adam and Noah one by one in a sequence of presumably possible worlds, Chisholm talks about a presumably possible world wherein Adam and Noah have switched all their qualitative properties (their life spans, their adventures, their names, etc.), and, except for this difference and what follows from it, everything in the imagined world \( w \) is just like the actual world. But, although they have exchanged all their qualitative properties, Adam and Noah have not exchanged their identities: in \( w \), it is Noah—not Adam—who has a son called ‘Cane’ who murders another son of his, called ‘Able’, etc.

Whether this description of \( w \) represents a genuine possibility is debateable. But Bird (2005: 449) uses this method to reach a (putatively) possible world \( w_{gr} \), which is like the actual world except for the fact that in that world charge and inertial mass have exchanged all their rôles. Bird’s intuition is that there is something wrong with \( w_{gr} \): “if anything exists which seems to fit our description of \( w_{gr} \), then it is just the actual world plus a decision to swap the names ‘inertial mass’ and ‘charge’” (2005: 450).

One might well not share Bird’s intuition here—in fact, the above example is basically just a scientific-looking version of the case of the three-element worlds and their \( Fs \)s and \( Gs \)s which I presented earlier. No epistemological disaster so far. But—and this is Bird’s contribution (2005: 451 f)—if real
properties do not have all of their causal powers essentially, one can use a modified version of Chisholm’s construction to infer the existence of a possible world in which two properties share all their causal powers in that world. We start with two properties $P_1$ and $P_2$, and instead of exchanging their powers one by one, we do this: in each world of a sequence of possible worlds, we keep $P_1$ as it is in the actual world, and replace one of the powers of $P_2$ by one of the powers of $P_1$. Doing this to all of the powers of $P_2$, we will finally get a possible world in which there are two properties, $P_1$ and $P'_2$, which share all their powers.

Now this is an epistemological disaster. Consider mass, for example. If not all its powers are essential to it, then for all we know we could be in a world in which we have two different properties, mass$_A$ and mass$_B$, whose causal powers are alike. You apply some force to an object, and it accelerates; now the explanation could be either that it has such and such mass$_A$, or that it has a mass$_B$ of the exact same magnitude. I agree with Shoemaker and with Bird (2005: 453) that the possibility of such a case “does serious damage to our concept of property.”

Moral: “real” properties at least, are individuated, in the actual world and across possible worlds, by their causal powers.
4. The applicability of Shoemaker’s theory.

Shoemaker’s theory says that properties are individuated, inter- and intra-worldly, by the causal rôles they play. But this is not a theory about the individuation of all properties: (a) it is only about real or genuine properties, i.e., properties whose “acquisition or loss by a thing constitutes a genuine change in that thing” (1980: 207). Moreover, (b) a property which is subject to this theory contributes to “the causal powers of the things that have it” (1980: 212). Looking back to my brief exposition of the theory, we see that condition (a) was used in arguing that properties have their causal powers necessarily, while condition (b) was crucial in arguing that no two different properties can have one and the same nomic profile.

In order to apply the causal theory of properties to the issue of the stipulative suggestion for solving the transworld identity problem for kinds, one should ask whether the property of being an instance of a given natural kind $K$ (say the property of being gold or the property of being a tiger, if gold and tiger are natural kinds) satisfies (a) and (b). That the kinds kind-essentialists are essentialist about satisfy (a) is rather obvious: it seems that nobody holds an essentialist view about the kind whose instances are exactly those beds slept in by George Washington, or about the colour grue (Shoemaker’s examples of “mere-Cambridge” properties), or the like. However, the case of satisfying condition (b) might be less obvious—after all,
one might ask if the property of, say, being a tiger plays a causal rôle in the world.

It seems to be an orthodoxy that at least part of what makes a group of objects a natural kind is that the group plays a theoretical or explanatory rôle in our science (or perhaps in the True Science). I think it follows that for any given group $K$ of material objects, if $K$ is a natural kind then the property of being an instance of $K$ does contribute to the causal powers of things that have it—how else could $K$ play an explanatory rôle?

Now one may think of explanations that use some mathematical properties like roundness. It is doubtful if anyone considers the set of round objects as a natural kind; but even if the set of material round objects is, in fact, a natural kind, the property of roundness is not a counterexample to my claim. Having the property of roundness does contribute to the causal powers of the objects that have it: it affects the way an object behaves in certain situations.

At any rate, it is the kind-essentialist, not the sceptic, who has to have an opinion about natural-kindhood status of a given set of objects, and it seems odd if the kind-essentialist thinks of a set of causally inept objects as a natural kind. If he thinks that tiger is a natural kind, he’d better also think that the

---

8 Thus Putnam (1970: 141) considers this definition: “a natural kind is a class which is the extension of a term $P$ which plays such-and-such a methodological rôle in some well-confirmed theory”. And LaPorte (2004: 19): “a natural kind is a kind with explanatory value”.

property of tigerhood does contribute to the causal powers of tigers. Thus I think it is safe to say that if $K$ is a natural kind, then the property of being an instance of $K$ is subject to Shoemaker’s causal theory of properties. Hence the property of being an instance of $K$ is a property that is not open to stipulation, and the transworld identity problem for $K$ cannot be solved stipulatively, if $K$ is a natural kind.

Here I am not suggesting that if $K$ is a natural kind then all its instances play *one and the same* causal rôle. Regarding kinds like electron or gold, it seems reasonable to think that there is a single property—perhaps something about being negatively charged or having atomic number 79, respectively—by virtue of which an object is an instance of the kind. In fact, a conception of kinds according to which a natural kind has an “essence” (a necessary and sufficient condition for being an instance of the kind) seems to be closely related to kind-essentialism.\(^9\)

However, perhaps there are natural kinds being an instance of which is not a matter of satisfying any necessary and sufficient condition. Thus in Richard Boyd’s view (1999), biological species—which are, for him, paradigmatic natural kinds—are examples of what he calls ‘homeostatic property cluster natural kinds’. For our discussion, what matters here is that

---

\(^9\) Dupré observes (1993: 270n7) that kind-essentialism is connected to the idea that natural kinds are subject to (exceptionless) laws: laws might be thought of as consequences of the common essence of instances of a kind.
for some natural kinds, being an instance of the kind can be a matter of
having a cluster of properties, and in such cases typically there is no single
property shared by all the instances of the kind. (Different as it is from the
traditional idea of defining kinds by genus and differentia, the idea of cluster
kinds does not imply that such kinds do not have essential properties—in
fact, Boyd himself believes that species have some of their properties
essentially.) Now if tiger is a natural kind of this kind, then perhaps there is
no single causally efficacious property of tigerhood to which we can apply
Shoemaker’s theory.

However, even if no single property is shared by all tigers given that the
property of being an instance of the kind tiger is a cluster of properties, still
the properties in the cluster are themselves real properties with causal
powers—how else could the cluster of those properties form something
having an explanatory rôle? Now suppose the kind tiger is a cluster of
properties \( P_1, \ldots, P_n \). In every possible world, and for every \( j \), being identical
to \( P_j \) is not a primitive notion because \( P_j \) is a real property— in every possible
world, being identical to \( P_j \) is a matter of playing certain rôles, and the
identity of the property \( P_j \) is not open to stipulation. But then the same is true
for every subset \( A \) of \( \{P_1, \ldots, P_n\} \) : in every possible world, whether or not an
object lies in the extension of each of the properties in \( A \) is a matter of fact.
That is to say, for every cluster of the properties \( P_1, \ldots, P_n \), being an instance
of that cluster in a possible world is a non-primitive notion, something about which we cannot stipulate.

If tiger is a cluster kind, then (with some oversimplifications) perhaps being a tiger in the actual world is a matter of either being a $P_1 \& P_2 \& P_{12} \& P_{34}$ or being a $P_1 \& P_7$. Perhaps in the possible world $W$ being a tiger is a matter of being a $P_3 \& P_{21}$ or being a $P_1 \& P_7 \& P_{17}$, and in the possible world $V$ being a tiger is a matter of being a $P_2 \& P_{22} \& P_{18}$ or being a $P_9$ or being a $P_4 \& P_8 \& P_9$. The relevant clusters might vary over possible worlds; but for no cluster we may stipulate its identity. Therefore possible-world stipulation about the kind tiger is not legitimate, even if tiger is a cluster natural kind.

The moral is that we cannot solve the transworld identity problem for natural kinds stipulatively.

5. Conclusion.

In “Identity and necessity” Kripke says (1971: 148),

And there seems to be no less objection to stipulating that we are speaking of certain people than there can be objection to stipulating that we are speaking of certain qualities. Advocates of the other view take speaking of certain qualities as unobjectionable. They do not say, “How do we know that this quality (in another possible world) is that of redness?” But they do find speaking of certain people objectionable. But I see no more reason to object in the one case
than in the other. I think it really comes from the idea of possible worlds as existing out there, but very far off, viewable only through a special telescope.

(Of course the real issue, as Kripke makes it perfectly clear elsewhere, is not the epistemological problem of knowing what the quality red or the person Nixon in a non-actual possible world is; it is the metaphysical question of transworld identity for redness and Nixon.)

In this chapter I tried to turn Kripke’s argument (or rhetoric) on its head: I think there are good reasons to say that one cannot simply stipulate that one is talking about the property of being red, or the property of being on fire, in a possible world $W$. Given what I find a reasonable metaphysics of properties (i.e., Shoemaker’s causal theory of properties), it follows that whether something is this or that real property in $W$ is a matter of fact about which one cannot stipulate. This, in particular, is true in the case of the property of being a $K$ in a possible world, if $K$ is a natural kind. The stipulative suggestion for solving the transworld identity problem for kinds or properties seems to be in trouble.

Stipulation about properties is problematic, and if object-stipulation is no less problematic than that (and I am not sure about this), we may be forced to revise our practice of object-stipulation as well. Why the “advocates of the other view” have never objected to property-stipulation, I do not know.
PART II

PHILOSOPHY OF SCIENCE
CHAPTER 4
Modal talk in the sciences: a rational reconstruction

In the previous part (Chapters 1-3) the focus was on the coherence of modal talk: when we say that gold or water is possibly or necessarily so and so, it is not clear what we are talking about—or so I argued. In the chapters to follow, my concern is the dispensability of modal talk in science. If the arguments of previous chapters are convincing, one might wonder about their implications for scientific discourse, in which one normally finds many instances of modal expressions. I will suggest that we need not be worried: I will argue that in order to make sense of scientific talk, as engaged in by scientists, we need not commit ourselves to the existence of any genuine modalities. Modal talk in science, I will argue, is always relativized to a non-modal discourse and is reducible to it: according to the relativistic reading of modality that I will present, the necessity or possibility of a statement is determined by its logical relationship to what I will call a ‘background theory’, which is the totality of the statements—general principles, particular assumptions, commonsense intuitions, etc.—that one assumes or presumes in a give context.

It can be seen straightforwardly that in mathematics the background theory is always non-modal. For physical background theories, I will examine some cases of putative modal principles in this chapter; in the next chapter I
will argue that even if there are genuine modal aspects of the physical world (i.e., modal properties which are not reducible to logical relationships between statements and non-modal background theories), the scientist is unable to discover them. Hence if scientific knowledge is comprised of what scientists discover, the background theory is non-modal in case of empirical sciences as well.

In Section 1 I offer the relativistic rational reconstruction of the modal talk in mathematics. The arguments of this section are not crucial to the rest of the chapter, yet I think it is instructive to see how the idea works for the simpler case of modal talk in mathematics, before proceeding to the more complex case of modal talk in physics.

Section 2 deals with modal talk specifically in physics, but the idea is meant to be applicable to modal talk in all empirical sciences. I try to establish the claim that when a physicist says that something is (physically) necessary, his claim is true just in case what he says logically follows from what he assumes as the background theory; likewise for possibility claims: a statement is (physically) possible just in case it is logically consistent with the background theory. After elaborating on my conception of background theory and explaining the main thesis, I spend the rest of the chapter answering some objections.
1. Mathematics.

Naïvely put, the question I want to investigate in this section is this: What does a working mathematician mean when she says that so and so is possible? What I am really looking for is a plausible rational reconstruction of mathematicians’ modal talk—a plausible way of making sense of their use of ‘possibly’, ‘is necessary’, ‘cannot’, and the like.

Here I am not interested in the modal status of mathematical theorems. One might argue that mathematical theorems are necessarily true, in a yet to be analyzed sense of ‘necessarily’; but this is not something a mathematician, qua mathematician, does. My topic here is modal talk in mathematics proper, not modalities which might be explored in the philosophy of mathematics.

Wilfrid Hodges (forthcoming) goes through the first 100 pages of a classic text-book in higher algebra, namely A Survey of Modern Algebra by Birkhoff and Mac Lane (1953), and finds more than 300 modal expressions, such as ‘can’, ‘cannot’, and ‘possibly’. The abundance of modal terms in a mathematics text-book is of course not unexpected, as mathematicians write in more or less ordinary natural languages, not in totally formalized ones. Thus it is quite normal to see something like We can visualize the above proof as follows..., and it seems obvious that the modality here (‘can’) has nothing to do with the subject-matter of a technical discussion in algebra.
What about modalities in statements of theorems? My understanding, in agreement with Hodges’s conclusion, is that one rarely finds modalities in the statements of theorems; and when we see them, “[t]here is no need to invoke any notion of ‘mathematical necessity’ or ‘logical necessity’ to explain these usages”, as Hodges concludes in his paper. Let me look at a specific example, which is not Hodges’s. In the course of analyzing this example, we will also see cases of modal talk in mathematical proofs and definitions.

Consider the case of a famous mathematical problem of antiquity, finally solved in the nineteenth century: the trisection of an arbitrary angle—or, in some presentations, just the 60-degree angle—by Euclidean ruler and Euclidean compass.1 After a large number of failed attempts over centuries, we naturally begin to wonder if it is in principle impossible to trisect the 60-degree angle. What does this mean? I suggest that, mathematically speaking, this means that we wonder if the existence of a method for trisecting this angle is ruled out by our accepted axioms. In other words, the trisection is considered impossible, I suggest, just in case it follows from our axioms that there is no way of doing it.

---

1 The Euclidean ruler is also called straightedge. In what follows, when I talk about trisection I mean trisection by Euclidean means. The concept of straightedge is different from a straightforward abstraction of everyday rulers, and the Euclidean compass has a more restricted function than a normal compass. For the technicalities—which are not of any substantial significance to my purposes—see Herstein (1975).
Perhaps the best way to argue for my view is to look at the actual practice of arguing for the impossibility claim. What does a standard text in higher algebra do to prove the impossibility of the trisection? Algebraically construed, a geometrical construction is a finite sequence of real numbers satisfying certain requirements dictated by the limitations of straightedge and compass; the last member of the sequence represents the constructed thing. As I will shortly explain in slightly more detail, the proof of the impossibility of trisecting 60° consists of showing that, given our axioms, there is no such sequence ending in \(\cos 20°\)—the proof shows that the existence of such a sequence is incompatible with our axioms. To settle the trisection problem we have to go beyond the ancient axioms for plane geometry; nevertheless we are still in the realm of mathematics.

Insofar as we believe in our axioms and we believe in the correctness of the incompatibility claim, we will not try to trisect the 60-degree angle—we believe that any such attempt is doomed to failure. Had the existence of a method for trisecting the 60-degree angle not turned out to be incompatible with our axioms, we would say that we might find a way of trisecting that angle. In cases—an example of which will be given below—where doing so and so is provably compatible with our axioms, we positively say that, in principle, we can do so and so. (Occasionally, but not always, a positive claim about the possibility of a so-and-so consists of actually presenting a method of constructing a so-and-so, not a non-constructive consistency proof.)
The conclusion I want to infer from this case is that the impossibility of the trisection of the 60-degree angle is not an irreducibly modal fact; rather, insofar as its truth condition and its proofs are concerned, it is the non-modal fact that, given our axioms, there is no construction of the specified form. I submit that this is what is meant when a mathematician says that that angle cannot be trisected. More has to be said for this reducibility claim; but I will have to digress a little to clarify the content of the claim first.

Strictly speaking, I am not here interested in the perhaps partly psychological question of what a typical mathematician means when she talks about the impossibility of trisection. My claim is that what I presented works as a plausible rational reconstruction of such a talk: it gives us the truth-conditions of the impossibility claim, and it fits what mathematicians actually do. I think the case study here will reveal that not only can we always offer modality-free formulations of modally formulated mathematical claims, but also that the proofs that mathematicians offer for the modally formulated statements are invariably just proofs for the corresponding non-modal claims. The proof of a possibility claim is the proof of a compatibility claim: to prove It is possible to φ, mathematicians just prove that φing is compatible with what they have accepted as axioms. The same is the case for necessity claims, mutatis mutandis.

To make my case as strong as possible, I have chosen a theorem which even some highly prestigious mathematics texts state in a modal language.
Let me examine two instances of such a modal talk. Van der Waerden concludes his mathematical discussion of the problem by saying (1970: 195, my italics),

we infer that an angle cannot be trisected by means of ruler and compass.

And this is Theorem 5.4.2 in Herstein (1975: 230, my emphasis):

It is impossible, by straightedge and compass alone, to trisect 60°.

What is the significance of ‘cannot’ and ‘impossible’ in these statements? It is just that there is no way of trisecting, I suggest, with no modality involved. Let me go into some details to verify this claim. Herstein provides the standard proof. One first defines a notion of constructibility to capture the intuitive notion of a geometrical construction: a real number α is constructible iff “by the use of straightedge and compass alone we can construct a line segment of length α” (1975: 228). This is not a modal notion, any more than there is modality in saying that an integer n is, by definition, an even number just in case we can find an integer m such that $n = 2m$, which is merely a not very precise way of saying that n is even iff there is an integer m such that $n = 2m$. (In fact, if n is big enough, we cannot “find” the relevant m!)
Let us return to the case of trisection. Herstein observes that if a real number is constructible, then it lies in some extension of the field of rational numbers of degree of a power of 2. Now if there were a way of trisecting 60°, then cos 20° would be constructible, which it is not, because, as simple geometry and previously learnt Galois theory show, this real number satisfies an irreducible polynomial of degree 3 over rationals. (One can check that, like the notion of constructibility of a real number, the notion of reducibility of a polynomial is not a modal notion.)

The proof of the theorem just shows that the assumption of the constructability of cos 20°, which is equivalent to the existence of a way of trisecting the 60-degree angle, leads to contradiction. The use of modal language is either a mere façon de parler (in the case of definitions), or else is reducible to non-modal notions (in the more relevant case of the statement of the theorem and the assertions made in its proof).²

I deliberately made things harder for myself by choosing the trisection problem: this is one of the few cases where we find modal talk in formulations of theorems in standard texts—normally, one does not see

² Also note that in the previous paragraph, which is written in (almost) natural English, there is a subjunctive conditional: if there were a way of trisecting 60°, then cos 20° would be constructible. Despite its grammatical form, the conditional is material (or at least what is argued for is the truth of a material conditional), as it can be seen if one tries to establish the claim.
modalities in such contexts. A Scientific American type of expository paper for
the general audience might say that one cannot find a non-trivial integer
solution to the equation $x^n + y^n = z^n$, or that there cannot be a computer
program doing so and so; technical discussions normally say that Fermat’s
equation does not have non-trivial solutions, or that there is no algorithm doing
so and so. And even when a mathematician uses modal language, her proofs
are proofs for the corresponding non-modal facts.

The fact that the mathematics of geometrical constructions is widely accepted
and has not been subject to great controversies might prevent us from
immediately seeing how theory-dependent the impossibility claim of the
trisection is. Things are not always that simple. Let us look at a modern
eexample.

Consider the question: Is there (or: could there be) a way of cutting a sphere
into a finite number of pieces—say five or nine pieces—so that by rearranging
them (via moving and rotating) one gets two spheres each of the same size as
the original? Assuming the axiom of choice, the answer is Yes (the Banach-
Tarski theorem); moreover, it is known that in the absence of the axiom of
choice the theorem does not follow.\(^3\) So the answer to the question about the

\(^3\) Such a decomposition of a sphere is called a paradoxical decomposition. For technical
details and the history of the problem see Wagon (1985).
possibility of paradoxical decompositions of spheres depends on which mathematics, with or without the axiom of choice, we choose to work with—and, as is well documented by Moore (1982), the axiom of choice was subject to a huge controversy in the last century.

Again, there is no irreducible modality here: the content of the Banach-Tarski theorem is that, assuming the axiom of choice, there is a paradoxical decomposition of the sphere, as can be seen by observing that the proof of the theorem is a proof of this statement, not of any modal statement.

There is an obvious objection to the claim about theory-dependence of the modalities of the above examples. Here it is. As a matter of set-theoretical fact, the axiom of choice is either true or false; if it is true, then the paradoxical decomposition is possible, as the proof of the theorem shows. If the axiom of choice is not a set-theoretical truth, then—the objection continues—we know that there are certain models of real numbers for which the paradoxical decomposition is possible, and there are other models for which it is not. But, insofar as mathematical facts are concerned, set theory with the axiom of choice is not on a par with set theory without the axiom of choice; neither are

---

The counter-intuitiveness of the theorem can be explained by the fact that here not all of the pieces that the sphere is cut into are, in mathematical jargon, measurable, and it does not make sense to talk about the “size” of a non-measurable subset of the space, let alone its preservation under transformations. The existence of non-measurable sets follows from the axiom of choice.
different models of real numbers on a par. Exactly one of these theories or models provides an accurate description of what sets and real numbers really are, and the paradoxical decomposition is either possible or impossible accordingly. Thus the possibility of the paradoxical decomposition of spheres is not really a theory-dependent claim: it depends on facts, on how things really are. Perhaps we do not yet know whether the axiom of choice is true or not, but that doesn’t matter here: the point is that the possibility depends on mathematical facts, not on mathematical theories. (End of the objection.)

Even without the italics, it is clear that this objection is deeply rooted in realism (a.k.a. Platonism) about mathematics. But surely realism in mathematics needs to be argued for: it is not the only available position—nor is it uncontroversially the least problematic one. Setting this aside, even if one accepts its presupposed realism, the objection just shows that, “as a matter of fact”, either there is a paradoxical decomposition of the sphere or there is not; the objection does not touch my main thesis, viz. the claim that modality here is reducible to non-modal notions: the possibility claim is reduced to a non-modal existential claim. Where we have a totally uncontroversial theory (call it the True Mathematics if you will), we might think of modal claims as theory-independent; yet it seems that even in that case the modalities are reducible to non-modal facts.

In general, I think this is true about modal talk in mathematics.
THESIS (MODAL TALK IN MATHEMATICS).

In mathematics, so far as truth-conditions are concerned, to assert a statement of the form necessarily $\varphi$ is to assert that $\varphi$ follows from the background theory. To assert a statement of the form possibly $\varphi$ is to assert that $\varphi$ is compatible with the background theory.

The background theory may vary from case to case: it may be the set of axioms for group theory, plane geometry, set theory with or without the axiom of choice, and so on. (More details about the notion of background theory and some refinements of the idea behind this thesis can be found in the next section on physics.)

So far I have argued that the modality in a mathematical claim, as used by a working mathematician, is reducible to the logical relationship of a non-modal statement to the background theory. Now, one may ask, are there any modalities in background mathematical theories themselves? It is hard to imagine how a mathematician could come to accept a genuinely modal claim as an axiom; but here we may use brute force, instead of speculation, to investigate the question: we may go through different mathematical theories and examine the question for each particular case. It is easily verifiable that in case of modern theories (e.g., measure theory and group theory), axioms are not formulated in a modal language. In the case of other axioms, say an
ancient axiom that says that a line-segment can be extended in both directions indefinitely, one may either rationally reconstruct the modalities in the way I did in this section, or just choose to work with modern axiomatizations of the theory in question, here Hilbert’s axioms for geometry, which contain no modalities.

But there is a shortcut: set theory. It is perhaps ontologically and methodologically fortunate that in modern mathematics one can almost always present a fragment of the standard Zermelo-Fraenkel set theory with the axiom of choice (ZFC), plus a fragment of the first-order predicate calculus, as the background theory.⁴ A quick look at the axioms of ZFC and classical logic shows that none of the axioms is modal—see, for instance, Jech (1978).

To summarize, I have argued that when mathematicians use modalities, their statements can—perhaps should—be understood as relations between some

---

⁴ I say ‘almost always’ to include, perhaps among a few other cases, topics in so-called large cardinals which might have extensions, rather than fragments, of the standard axioms of set theory as their background theory.

Also, it should be noted that here I work with the standard picture of mathematics as set theory. Following a suggestion by Putnam (1967), one may think of another picture, viz. mathematics as modal logic. Although there is a book on the subject (Hellman (1989)), it seems fair to say that this latter picture represents a minority view, and not yet fully developed.
non-modal statements and the theories they accept in the background. The background theories themselves are not formulated in a modal language.

2. **Physics.**

Is it physically possible to move faster than light? I suggest that physicists’ negative answer to this question can be understood as an answer to another, non-modal question: Is moving faster than light compatible with our currently accepted physics? We know that the answer is negative: special relativity rules out faster-than-light speeds—for every object, its moving faster than light is incompatible with our favourite theory.

This section is devoted to presenting a rational reconstruction of physicists’ modal talk, an example of which given in the previous paragraph. As in the case of mathematics discussed in the previous section, I think modal claims in physics can always be understood as dependent on our background theory: in my rational reconstruction, physical possibility is compatibility with the background theory, and physical necessity is provability from the background theory (see the Thesis below). But I have first to clarify the notion of *background theory* which is in use here.

In modern mathematics, we may almost always present portions of standard set theory and logic as our background theory. Things may not always be that clear-cut in the case of physics. In any given context, by the **physical background theory** (or just the **physics**, for short) I mean all the
explicitly formulated physical principles, plus all the assumptions about initial conditions, plus all the needed mathematics and logic, that are used in arguments and derivations.

Thus in a given context, our physical background theory—as I stipulatively define it—is the totality of whatever assumptions we use in that context. This might also include some statements of commonsense intuitions, as well as “a good deal of unformulated general opinion”, as Monton and van Fraassen (2003: 410) put it in their discussion of counterfactuals.

To re-emphasize, the background theory varies from context to context—some initial conditions or particular facts might be part of the background theory in one context and not part of another. To get a clearer image of this notion of background theory, it might be in order to give a non-scientific example. Here is a case of a shift in the context, which is based upon an example given by Russell (1905: 519). Consider these two events: (A) the death of W.V. Quine, 25 December 2000, and (B) the terrorist attacks of 11 September 2001. Was it necessary that A happened before B? I think we ordinarily answer this question in the negative: for all we know about

---

5 To repeat, what I call a ‘background theory’ is a set of sentences. There are philosophical debates on how to understand the notion of a scientific theory (as contested by those who favour the so-called semantic and syntactic views of theories), but such discussions are orthogonal to my concern here, which is understanding modal talk in the sciences, not theorizing about the notion of theory.
Quine’s health conditions and about the terrorists’ plans, there is no important connexion between A and B. Yet in a context in which you and I both know about the times these events occurred, if I tell you that so and so happened before the death of Quine but after 9/11, you could quite reasonably say that *that is impossible because Quine died before 9/11*. The point is that there is a context wherein we take the particular fact into account that A occurred before B — this, together with many other facts such as the linear order of moments of time, is part of the background theory of this little dialogue between you and I.

Having such a broad notion of background theory, my notions of provability and compatibility are strictly logical: *p* is **provable** from *T* iff there is a finite proof of *p* from *T* in the standard logical sense, and *p* is **compatible** with *T* iff *p* is consistent with *T* in the standard logical sense.

As in the case of mathematics, a realist might contend that, independent of what the consequences of our current theories are and independent of what is or is not logically compatible with our physical theories, moving faster than light is, *as a matter of fact*, either physically possible or physically impossible.

Now whether or not there are irreducible modal facts about the world is an issue I do not deal with in this chapter. (Chapter 5 is about a related issue.) What I am arguing for here is that what a scientist does when he examines a physical modal claim can be rationally reconstructed as examining the logical
relationship between a corresponding non-modal claim and his background theory. Let me flag the thesis before arguing for it and considering several objections to it.

**Thesis.**

In any given context, physical modal statements can be understood as true or false relative to a background physical theory (in short: they can be understood relative to a physics). With respect to a physics $T$, the assertion of the physical possibility of a statement $\phi$ can be understood — insofar as truth-conditions are concerned — as the assertion of $\phi$’s logical compatibility with $T$; so far as truth-conditions are concerned, the assertion of the physical necessity of $\phi$ can be understood as the assertion of the provability of $\phi$ from $T$.

Normally, the reference to $T$ is dropped when the context is clear. Also, here $\phi$ is assumed to be non-modal; if $\phi$ is itself modal, the truth-conditions are given recursively.

Note that this account satisfies the quite minimal requirement of respecting our basic logical understanding of the relationship between possibility and necessity: on the one hand, the necessity of $\phi$ is commonly taken to be equivalent to the impossibility of not-$\phi$; on the other hand, the provability of $\phi$ with respect to a theory $T$ is logically equivalent to the inconsistency of not-
φ with T. So the thesis at least captures the relationship between necessity and possibility.

One major objective of this chapter is to defend the above understanding of modalities (call it the relativistic understanding) against the realist’s objections. I will concentrate on the case of physical necessity/possibility, though I think the relativistic reading is also a plausible rational reconstruction of modal talk in all sciences.

Note what this thesis is not. It is not a thesis saying that modal talk in physics is nonsense; it is not a non-cognitivist position either: it actually presents truth-conditions for modal physical statements.

Nor does the thesis say that modal physical claims are useless or unrelated to the physical world. We certainly are interested in the logical consequences of our physical theories, because, obviously, if a theory is true (i.e., if it holds in the actual world), then so are its consequences. If special relativity is true, then so is its particular consequence that no object moves faster than light; thus, if we believe in special relativity, then we (should) believe that any attempt to move faster than light will be unsuccessful.

The thesis does not say that physical modality is just a matter of what we believe. For every physics T and for every statement φ, either φ follows from T or not, and either φ is compatible with T or not; and these facts are independent of the way we think of them—they are even independent of the
fact that anyone ever formulated $T$ or thought of $\varphi$. All the thesis says is that modalities can be understood as derivative: their truth-conditions are determined by logical relations between theories and statements. The choice of the background theory is ours; yet, given any background theory, what is possible or necessary relative to that theory is independent of our will. As a matter of mathematical fact, moving faster than light is incompatible with the theory of special relativity; this fact was true even before Einstein, even if no intelligent creature ever lived in the universe.

I suggest that to critically assess a modal physical statement, we fix a background theory, i.e., we specify the physics relative to which the truth-conditions of modalities are given. So, again, when a physicist says that so and so is physically necessary, the thesis rationally reconstructs him as saying that so and so follows from his background theory $T$. A conversational implicature here might be that the physicist’s talk presumes the physicist’s belief that $T$ is true; but this need not be the case—in fact, perhaps every working physicist believes that his background assumptions are less than wholly and completely true. For all I am interested in here, the physicist may believe that $T$ is true, or just that $T$ is empirically adequate—and these are not all the options. Fixing $T$ as the background theory, the physicist may believe that $T$ is false or even empirically inadequate: for some reason, he might be interested to see what the world would look like if $T$ were true or if $T$ were empirically adequate. Or he may just give it as an exercise to his students to
show that so and so follows from, say, a patently false assumption in
Aristotelian physics, hence physically necessary relative to Aristotelian
physics.\textsuperscript{6}

Lastly, the thesis is not offering \textit{synonyms} for ‘it is physically necessary’, ‘it
is physically possible’, and the like: it does not offer analytical meanings for
these terms. Arguably, in English ‘bachelor’ means unmarried adult man and
is synonymous to something like ‘unmarried adult man’; to say that Jack is a
married bachelor is thus perhaps to contradict oneself. On the contrary,
‘Physical principles are not physically necessary’ is not a contradiction in
terms, which would be if the thesis were a correct account of the analytical
meaning of ‘it is physically necessary’. The thesis just says that, in a given
context, \(\varphi\) is \textit{physically necessary} is true if and only if \(\varphi\) follows from the
background theory of the context—it gives \textit{truth-conditions} for modal
statements. The thesis is a rational reconstruction of physicists’ modal talk—
an attempt to make sense of such talk, in a way that fits what they do as
physicists. Here I want to argue that the thesis is a plausible rational
reconstruction.

\textsuperscript{6} Since the background theory is not assumed to be true, sentences that are necessary relative
to a background theory might be false; hence the law \(\Box\varphi \rightarrow \varphi\) fails. For more on this see
Appendix 5.
Before considering some objections, let me examine a case which, I think, supports the relativistic reading of physical modalities. If the way I explain this particular case is sound, then my use of it is rather ironic, because, in a different context (laws of nature) the kind of scenario described in this case is often suggested by a realist argument against an empiricist view.

**THE CASE OF “UN-ACTUALIZED PHYSICAL POSSIBILITIES”.**

In a famous example, David Armstrong (1983: 17 ff) asks us to suppose that nowhere in the universe has there ever been a solid lump of gold with a volume greater than a cubic mile. (Let us call any such huge lump of gold a **Hugold**.) Suppose, moreover, that there will never be any Hugold at any place in the future. Nevertheless, according to Armstrong, the existence of a Hugold is not a physical impossibility, as opposed to the existence of a piece of uranium-235 of the same size (a **Huranium**), which is a physical impossibility, as critical-mass considerations show. There is a manifold of such examples in the literature on laws of nature.\(^7\)

---

\(^7\) Almost the same example is given by Reichenbach (1954: 11 f). A similar example can be found in Hempel (1966: 55). For some other examples of the same spirit see Kneale (1950: 124), Molnar (1969: 81), and Popper (1980: 427f). As used by these authors, such examples are intended to refute the so-called regularity theory of laws of nature—they are meant to show that, for a general statement, being a law of nature is not just a matter being true at every time. Here I am not concerned with the debate over the notion of a law of nature.
So, let us assume that nowhere at any moment in the whole history of the universe—past, present, and future—is there either a Hugold or a Huranium. Still we feel that there is a difference: in principle, if we cared to (and if we had enough gold), we could make a Hugold. On the contrary, no matter how hard we try and how much uranium we possess, there really cannot be a Huranium. What accounts for the difference?

I think that by means of the relativistic reading of physical modalities we can explain this feeling. The existence of a huge lump of $^{235}$U is ruled out by our accepted physics: critical-mass considerations (whatever they are), which are incorporated in our physics, are incompatible with the existence of a Huranium and we know this; hence the corresponding impossibility judgement. On the other hand, so far as we are aware of the consequences of our physics, no such considerations are applicable to a huge lump of gold—the existence of a Hugold is, to the best of our knowledge, compatible with our physical principles; hence the corresponding possibility judgement. If we believe our physical principles to be true, then we have good reason to think that there will be no Huraniums, no matter what; we do not have such a reason in case of Hugolds.

The realist’s intuition is that even if we know that there is no Hugold in the whole history of the universe, there still could be one. However, it is not clear how we can make scientific sense of this ‘could’, if it is not to be understood relativistically. Our reason for asserting the physical impossibility
of the existence of a Huranium is grounded in critical-mass considerations, which are parts of our physics. We cannot hold both (1) critical-mass considerations, and (2) the statement that a Huranium exists at some point in the history of the universe, for we know that they are incompatible. As we have good reason to keep (1), we reject (2). But if this is the way we discover, or argue for, impossibility claims (and there seems to be no other way), it is odd to think that for possibility claims we should seek something over and above compatibility with the background theory. If, so far as truth-conditions are concerned, asserting the impossibility of $\varphi$ is to assert that $\varphi$ is incompatible with our physics, then to assert possibility must be to assert compatibility. And, given that the necessity of $\varphi$ is logico-semantically equivalent to the impossibility of not-$\varphi$, it follows that to assert the necessity of $\varphi$ is, so far as truth-conditions are concerned, to assert the incompatibility of not-$\varphi$ with our physics; hence to assert physical necessity is to assert provability from our physics.

The realist may admit that the way physicists prove or discover physical modalities is via investigating logical relations between statements and theories, but add that this does not show that, e.g., to be physically necessary is to be deducible from the background theory. However, here I am not denying that there might be irreducible physical modalities; my point is that even if there are such things, scientists do not deal with them as such. In rationally
reconstructing modal talk in science we do not need to talk about irreducible modalities, even if there are such things as irreducible modalities.

One move the realist might try at this point is to accept the conclusion of the theory-dependence of physical modalities but insist that some of the principles of our physics (or some laws of nature) are irreducibly modal, and some irreducible modality is thereby inherited by physical statements like the impossibility of the existence of a Huranium. However, again, I do not deny that there might be (irreducibly) modal features of the physical world; but will I argue, in Chapter 5, that even if there are such features, the scientist cannot discover them and include them in his background physics. Hence, if my argument is sound, then irreducible physical modalities are scientifically irrelevant, and irreducible modalities cannot be found in the principles of our physics either. Apart from that, below in my discussion of actual physicists’ use I will explain away the appearance of modality in some of the principles of physics.

**Objection: the open-minded physicist.**

Suppose a physicist says that moving faster than light is physically impossible. According to the relativistic thesis, this is, so far as truth-conditions are concerned, nothing but saying that it is a theorem of this physicist’s favourite physics that no moving object moves faster than light. Let us also assume that he has recently checked the relevant argument again,
and he is absolutely certain about its validity: he knows it for a fact that his physics (in this case a theory which contains the theory of special relativity) rules out speeds greater than the speed of light. Now suppose that today he hears rumours about the success of his colleagues in Berlin in building a faster-than-light rocket. The source having been reliable in the past and the Berliners being world-class scientists and rocket experts, our physicist takes the news seriously—he seriously considers this: moving faster than light is possible. But, according to the thesis (the objection concludes), he just can’t: as we stipulated, he still believes that moving faster than light is incompatible with his physics, hence (according to the thesis) a physical impossibility. So there is more to a possibility claim than just consistency with the background theory.

In reply, I think one should distinguish two cases: our physicist hears that the Berliners have actually observed the phenomenon he considered impossible, or he hears that they just theoretically proved that moving faster than light is possible. As the story has it, in both cases our physicist believes the news; but the implications for the relativistic thesis might be different. For suppose that the Berliners had just claimed that moving faster than light was possible, without claiming that they had observed—or had brought into existence—an instance of it. Then it seems clear what our physicist would do. He would ask for their argument and he would peruse it. Given that he is certain of the correctness of his own proof of the incompatibility between his
background theory and the statement that something moves faster than light, he would enquire into the Berliners’ background theory. Perhaps their physics does not include all of his principles or particular assumptions? If that doesn’t explain the tension, he will try to find mistakes in their argument for the possibility claim. If none of these settle the disagreement, he might think that perhaps a subset of the intersection of his and the Berliners’ physics is inconsistent. Another option, still further from the “edge of the system” (as Quine would put it), is to blame mathematics and logic. But all these are questions of what follows, or does not follow, from principles and extra assumptions. Hence if in the story the Berliners are just said to have theoretically argued for the possibility, then this is no threat to the relativistic thesis. It seems that whatever the ontological status of physical modalities might be, theoretical arguments about impossibilities and non-actual possibilities are just arguments about incompatibility and compatibility with our accepted principles.

So let us assume that our physicist thinks that a statement \( \varphi \) is physically impossible, and he takes it seriously that \( \varphi \) has actually been observed to be the case. What is going on here? As the objection stipulates, the physicist still thinks that, as a matter of (mathematical) fact, \( \varphi \) is incompatible with his background theory \( T \). If \( T \) were a correct description of what is the case in the world, \( \varphi \) would not be the case; now that our physicist has good reason to think that \( \varphi \) is the case, he has good reason to think that part of \( T \) is false—if
one of the logical consequences of $T$ is false, then one of the sentences of $T$

itself must be false. If it really turns out to be the case that $\varphi$ has been

observed to be true, then our rational physicist will say that his $T$ is not true.

However, he still retains his belief that $\varphi$ is ruled out by $T$—despite the

falsification of $T$, this fact remains true, though, of course, it loses much of its

importance. And if, after the rejection of $T$, our physicist now wonders about

the possibility of another statement $\psi$ as part of a research programme and

not just as an exercise in theoretical physics, we have a change of context, a

change of background theory: he is now thinking of the compatibility, or lack

thereof, of $\psi$ with a different, perhaps yet to be developed, physics $T'$.

Perhaps the point of the objection is that just looking through the logical

consequences of theories is not a good way of finding real possibilities: if

things happen to be as in the objector’s story (the second case where the

phenomenon is observed), then moving faster than light is really possible, no

matter what consequences of our physics are. However, it should be clear that

a tu-quoque reply to the realist is available here: the realist himself has no

way other than a compatibility argument to show that a non-actual $\varphi$ is

physically possible. And as for actual $\varphi$s, there is no disagreement about the

fact that they are possible, in whatever sense of possibility.
OBJECTION: PRACTICAL VS. IN-PRINCIPLE.

Consider Armstrong’s example again. It is \textit{in principle} impossible to fabricate a huge lump of uranium-235, as we know from critical-mass considerations. As for a huge lump of gold, however, there is no such in-principle impossibility. But suppose that we know that there is not enough gold in the universe to make a lump of gold of the specified size, and suppose that this particular fact is included in our background physics. Now the relativistic thesis would announce that, with respect to this physics, the existence of a Hugold is a physical impossibility. But surely (the realist objects) there is no “deep” reason for this—the hypothetical impossibility of a Hugold would be just \textit{practical}. There is an intuitive distinction between the in-principle and the practical impossibility; but the relativistic thesis is too coarse to make this distinction—if enough particular facts are included in a background physics, then the practical vs. in-principle distinction is lost.

The realist has an intuition about the difference—which is presumably a difference of kind—between “practical” and “in-principle” impossibilities. How can he demarcate the two? One way may be to say that an in-principle impossibility is one that is ruled out by laws of nature, whatever they are, and a practical impossibility is one which is ruled out by laws of nature plus some other true assumptions (e.g., assumptions about how much gold we have). But if \textit{this} is what distinguishes the two notions of physical impossibility, how can we ever know that a logically possible situation is an in-principle physical
impossibility? Well, the realist might say: if we know the laws of nature, then we know the in-principle impossibilities. However, I reply, even waiving worries about the antecedent of this conditional (including the controversies about the very notion of laws of nature), the problem is that now the realist’s account of impossibility is really not different from the relativistic account. Certainly the set of laws of nature, should there be such a set, is an excellent candidate for a background physics, and, with respect to this background theory, the relativistic physical impossibilities are the same as the realist’s in-principle impossibilities. If we know the laws of nature, then we can measure impossibilities against them and reserve the unqualified ‘impossible’, or the qualified ‘impossible in principle’, for whatever that is excluded by laws of nature, and some of the statements ruled out by other background physical theories we may call ‘(merely) practical impossibilities’. The two accounts are really not different in this case, and the realist has admitted that, with respect to truth conditions of impossibility claims, modality is a matter of a logical relationship of a statement with the background theory — only he thinks that the background theory is something very special. But even in this case there are no irreducible modalities. (Again, the realist might think that some laws of nature are themselves modal. Here the realist owes us some examples. I will consider two putative cases of such laws below in this chapter; in the next chapter I will argue that even if there are modal features of the world,
presumably described by some genuinely modal laws of nature, such features cannot be discovered by the empirical scientist.)

So I think the realist owes us an explanation of his intuition, a theory about in-principle physical impossibility. Not that if he cannot offer a good explanation we have to quine the intuition; but here the intuition, which is perhaps not backed by a good theory, need not be shared by the empiricist—to appeal to an intuition about the difference between the in-principle and the practical is perhaps begging the question against the empiricist.8

But perhaps one can explain the difference between the in-principle and the practical impossibility in an empiricist-friendly way, without appealing to irreducible modalities. I think the difference is not a difference in kind, and I think determining it is a pragmatic issue. Quine (1951) convincingly argues that nothing, not even mathematics or logic, is absolutely immune to revision. With no claim of having a worked-out theory of this, I want to suggest that perhaps the in-principle vs. practical distinction is just a matter of degree: the

8 Another thing that the realist owes us here is an argument to the effect that the practical vs. in-principle distinction, if it can be made, is scientifically significant. If the distinction is supposed to be based on the notion of natural laws, then there are reasons to think that it is scientifically irrelevant—as Nagel puts it (1961: 49), “The label ‘law of nature’ (or similar labels such as ‘scientific law,’ ‘natural law,’ or simply ‘law’) is not a technical term defined in any empirical science; and it is often used, especially in common discourse, with a strong honorific intent but without a precise import.” Chapter 8 of Mumford (2004) presents a defence of Nagel’s view.
more central and the less susceptible to revision a background theory $T$ is, the more of an in-principle character the $T$-impossible statements are. If it is not that hard to revise a background $T$, if the costs of such a revision are not very high, then we think of what is ruled out by $T$ not in-principle impossible. Thus revising, or totally setting aside, special relativity will be a great change in science; so moving faster than light is considered to be an in-principle impossibility. On the contrary, it will not be a big deal if we realize that our estimation of the total amount of gold in the universe was mistaken; hence the merely practical impossibility of a Hugold (given that it is part of our physics that there is not enough gold to build a Hugold).

**Objection: Actual Physicists’ Use.**

Is the relativistic reading a good account of how physicists actually use modal terms? One objection goes like this. If the relativistic account is correct, if physical statements—insofar as their truth-conditions and their assertions by physicists are concerned—have their modal status only in virtue of their relations to physical theories or principles, then it does not make much sense to formulate physical principles themselves in a modal language—it would be a tautology that every principle (properly so called; an axiom, that is to say) is physically necessary. Yet we have numerous cases of modal talk in physics texts, especially in the expositions of principles.
It is undeniable that very often we see modal statements in physics texts. For example, a recent edition of a standard undergraduate text-book states Newton’s first law in a modal language—this is from Halliday et al. (2005: 88, my italics):

If no force acts on a body, the body’s velocity cannot change; that is, the body cannot accelerate.

However, there are—so I claim—always modality-free formulations of a modally formulated scientific proposition. Regarding Newton’s second law for instance, here is the required version, formulated by Isaac Newton (1726: 416, my non-italics):

Every body preserves in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by forces impressed.

The velocity-cannot-change of the first version takes the form of the there-will-be-no-acceleration of the second one. I take it for granted that the authors of the two versions are expressing the same law. (But see also the next chapter, where I consider the objection that perhaps they are not.)
Here is another example, the second law of thermodynamics. In this case the original formulations are modal. I quote from Bailyn (1994: 88), to which I add emphasis. Rudolf Clausius, 1850:

No process is possible whose sole effect is to transfer heat from a cold body to a hot body. By sole effect is meant without the rest of the universe changing, or changing in a cycle of operations.

And William Thomson (Lord Kelvin of Largs), 1851:

It is impossible by means of inanimate material agency to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects.

Now it is well known that the second law of thermodynamics admits of many equivalent formulations. The following is from Baierlein (1999: 29, my italics):

If a system with many molecules is permitted to change, then—with overwhelming probability—the system will evolve to the macrostate of largest multiplicity and will subsequently remain in that macrostate. Stipulation: allow the system to evolve in isolation. (The stipulation includes the injunction, do not transfer energy to or from the system.)
A quick review of the concepts involved here might be in order, to make sure that Baierlein’s formulation is not modal. As one would expect, a macrostate—an abbreviation for ‘macroscopic state of affairs’—is a state described by “a few gross, large-scale properties,” such as pressure, volume, temperature, and total mass (p. 27). A microstate is one described by “specifying in great detail the location and momentum of each molecule and atom” (p. 25). The multiplicity of a macrostate is the number of microstates that correctly describe it. Thus (Baierlein’s analogy, p. 27) suppose there are four balls—call them A, B, C, D—and two bowls. The macrostate all balls are in the left-hand bowl has minimum multiplicity, viz. 1 (the location of balls within a bowl doesn’t matter). The macrostate the balls are evenly distributed in the two bowls has the largest multiplicity, viz. 6: each of the microstates (AB, CD), (AC, BD), (AD, CB), (BC, AD), (BD, AC), and (CD, AB) corresponds to it. Finally, though Baierlein does not explicitly define it, it is clear that his notion of the probability of an event is the familiar, purely combinatorial one—thus (p. 26) if you toss a “fair” coin a million times, the probability that the number of heads is within 1 per cent of 500,000 is $1 - 2.7 \times 10^{-23}$, which is “overwhelmingly” close to 1. There is no modal notion here.

Or, to put the second law more succinctly, let us talk about entropy, which is basically defined as the logarithm of multiplicity. Now, “for all practical purposes, the one-line version of the Second Law is this: An isolated
macroscopic system will evolve to the macrostate of largest entropy and will then remain there” (p. 46). Again, no modality is involved.

It may be instructive to examine a large number of physics texts and try to find non-modal versions of the statements that are often formulated in a modal language. It may also be interesting to see if there is a correlation between the presence of modal discourse, or lack thereof, in a physics text, and the extent to which its author is, on other grounds, considered a rigorous author. This, however, is not what I wanted to do here. I hope I have provided enough empirical data (about what one can find in physics texts) to confirm my a posteriori claim that for each physical statement one finds in a physics text, there is a non-modal version of it. This is of course not

---

9 The case of an elementary text in philosophy might be of some interest. While explaining the difference between necessity and certainty, Elliott Sober gives the example of the second law of thermodynamics: “Finally, in the nineteenth century, physicists working in the area called thermodynamics proved that perpetual motion machines are impossible” (2005: 49, my italics). In his next chapter, Sober argues against the creationists’ confused application of the second law: “They claim that this law makes it impossible for order to arise from disorder by natural process” (p. 62). And what is the second law, again? Now that precision matters more, his formulation is basically the modality-free one I quoted from Baierlein: Sober writes, “What the Second Law actually says is that a closed system will (with high probability) move from states of greater order to states of lesser order” (ibid., with change of emphasis).
unexpected, given that we see no modal operators in the formulas we find in physics texts.

The advocate of genuine, irreducible physical modalities may object, echoing Bressan (1974: 299), that “the use of modalities (possibility concepts) and the use of a modal language are not equivalent”. I agree. Yet, if the formulas of physics texts do not contain boxes and diamonds, then I think it is incumbent on the realist to argue that, nevertheless, physics deals with irreducible modalities. Moreover, if my argument in Chapter 5 goes through and the physicist can aim at discovering non-modal facts only, then this objection will be adequately replied.

Also note that for the Thesis to be a good rational reconstruction of physical modal talk, it need not be the case that for every modally formulated physical statement there is already a text-book non-modal version. The Thesis is not a sociological claim about the way physicists actually talk (though I think I have provided some evidence that it is not alien to actual physicists’ talk); rather, the Thesis presents a way to make sense of their modal talk. Even if I had failed to find Baierlein’s formulation of the second law of thermodynamics, I still could offer a routine reformulation of Clausius’s: there is no process, nor will there be one, whose sole effect is to transfer heat from a cold body to a hot body.
Where are we now? I have argued that in scientific discourse and insofar as truth-conditions are concerned, modal talk can be reduced to talk about logical relationships between background theories and non-modal statements. But are there any irreducible modalities in background theories? In the next chapter I will argue that, to the extent that background theories are results of scientific practice, the answer is No. This will complete the justification of my rational reconstruction of modal talk in science: we can make sense of such a talk without any need to appeal to irreducible modalities.
CHAPTER 5
The epistemic inaccessibility of non-relativistic modalities

Do empirical scientists deal with modalities? Certainly they use modal terms, at least in informal discourses. If the relativistic account of modalities that I developed in the previous chapter works, then whenever a scientist asserts the necessity of a non-modal statement \( \varphi \), we can understand her as asserting that \( \varphi \) follows from the background theory of the context. But what if there are modal statements in the background theory itself? The background theory of a context being, by definition, the set of all the assumptions used in that context, it seems reasonable to say that we do not have any further theory to which we can reduce the modalities of the background theory — there is no natural candidate for a background meta-theory in the context of an empirical science. Hence if there are modalities in the background theory of a scientific context, they must be irreducible to logical relations between statements and theories. So if there are modalities in some background theories in sciences, then the relativistic reading of modalities falls short of being a totally reductive account of scientific modal talk. This would be damaging to my project since the aim of this part of the dissertation, as announced in the introduction to Chapter 4, is to argue for the dispensability of modal talk to science, and the relativistic reading is my main tool.
In this chapter I will try to meet this challenge to the relativistic account by denying the antecedent of the above-mentioned conditional. In the course of doing so, I will take it for granted that if there are modal statements in the background theory of a scientific context, then at least some of these statements are about properties of things or events in the physical world and are not knowable a priori. The major part of this chapter (Section 2), then, is devoted to an investigation of an immediate epistemological question: How can scientists discover, or infer, irreducible modal facts about the world? I will argue that they cannot. On the other hand, what is epistemically inaccessible to scientists cannot be part of science. Therefore, if the argument is sound, there are no modalities in scientific theories—scientifically speaking, irreducible modalities are irrelevant, even if there are modal features of the world.

There is a short section before the main one. In Section 1 I will talk about the prospects of there being genuine, irreducible modal facts about the world. In that section I will raise a question about the very concept of an irreducible modal feature of the world: What do we mean when we say that so and so is physically necessary, if this is not to be understood relativistically? After flagging my worries about the answerability of this question, in the remainder of the chapter I will assume that it does make sense to talk about irreducible physical modalities. (Section 1 does not play a logical rôle in the main argument of the chapter, which is presented in Section 2.)
1. Modal features of the physical world?

In Chapter 4, after presenting a version of Newton’s first law which is
couched in a modal language, I presented a modality-free version (these were
taken from Halliday, Resnick, and Walker (2005: 88) and Newton (1726: 416),
respectively). The second version says that an object will not accelerate unless
a force acts on it; the first version says that it cannot accelerate unless it is
forced to. I claimed that these are meant to be formulations of one and the
same law, but the realist might refuse to accept that the two versions have the
same truth-conditions. They say different things, and by demodalizing, the
realist might object, we lose some of the content. Thus Newton’s formulation
is just about what will not happen, while the law formulated by Halliday et
al. is stronger: it not only makes a claim about what will not happen, but also
what cannot happen. And surely these are different claims (the objection
continues): to wit, one is about the actual world only, while the other is about
a class of possible worlds, the actual world included.

I think it is fair here to ask for some clarification about the ‘cannot’ of the
modal formulation in the realist’s understanding. It must mean something
other than ‘there will never be’, for there is no modality in the latter. And if
this is a non-relativistic ‘cannot’, then it is not clear how we should
understand it without the notion of a possible world. But what is a
(physically) possible world? To say that a physically possible world is one in
which the laws of nature hold sounds circular if it is said by a realist who
thinks that statements of some laws of nature are modal—and, circular or not, it is at best just a labelling of the problem, not a solution to it. Or are we tempted to try to make sense of possible worlds by means of counterfactuals? It is not entirely clear or uncontroversial how counterfactuals themselves can be understood without the aid of the concept of laws of nature. Moreover, the very attempt to appeal to counterfactuals to explain possible worlds is weird—as Quine puts it (1972: 492-493) in his review of Kripke (1971),

He reassures us regarding his talk of possible worlds: it is not science fiction, but only a vivid way of phrasing our old familiar contrary-to-fact conditionals. Let us recall then that some of us have deemed our contrary-to-fact conditionals themselves wanting in clarity. It is partly in response to this discomfort that the current literature on possible worlds has emerged. It is amusing that some of us same philosophers may be so bewildered by this further concept that we come to welcome the old familiar contrary-to-fact conditionals as a clarification, and are content at last to acquiesce in them.

So, if one thinks that some information is lost in the process of giving non-modal versions of modal physical statements, one owes a clarification of the content of the lost parts—we need to hear more about the meaning, or truth-conditions, of modal physical statements, preferably in such a way that
discovering modal facts will turn out to be a task that scientists are capable of doing.

2. What scientists cannot do.

Let us waive the what-do-you-mean qualm: let us assume that it does make perfect sense to talk about irreducible physical modalities. Let us assume that there really are irreducibly modal aspects of the physical world. The question now is: How is a physicist supposed to discover these modal facts? If other physically possible worlds are real, concrete words (in accordance with Lewis’s (1986) account of possible worlds), then it is axiomatic that other possible worlds are causally disconnected from the world we inhabit; our physicist cannot go there and observe what is going on there—he cannot look at them via a telescope either.

Whatever the ontology of irreducible modal facts may be, it seems that physicists cannot observe such facts—how could they? Wherever they go in space and time to observe things, they are still within the limits of the actual world, and whenever they observe a non-modal \( \varphi \) to be true, it seems that what they observe is compatible with \( \varphi \) not being true in every physically possible world, so far as observation is concerned. For them, there is no outside—they cannot physically escape the actual world. If this is the case,
then the only way a physicist can warrantedly assert modal statements is to infer them.

By talking this way, I am not attributing to the realist any particular view about possible worlds (though I take it for granted that modal talk is translatable into possible-worlds talk). So far as my purposes here are concerned, possible worlds could be maximally consistent sets of sentences, concrete physical objects, or whatever is favoured by the realist’s theory of modality. What I do assume, to give some content to the debate with the realist, is that the actual world is not the only physically possible world. Moreover, I assume that we do not observe what is going on in other possible worlds in the way we observe things in the actual world. If non-actual physically possible worlds are sets of sentences, then perhaps there is not much room for the “outside” metaphor of the above paragraph; still one does not observe that so and so is the case in a non-actual possible world — not in the sense that we observe, about the actual world, that in it unsupported heavy objects fall.

So to assert modal statements we have to infer them. But infer them on what basis? Starting with observing what is the case, how can one infer something that is necessarily the case? What kind of inference could this be?

I will start examining some options in a moment. Before doing so, perhaps a proviso is required to fill a gap in the argument. Granted that we cannot observe necessities, it might not immediately follow that inferring them is the
only way of acquiring modal knowledge. Gödel the Platonist talked about a faculty of mathematical intuition via which “axioms force themselves upon us as being true” (1964: 268). Similarly, one might entertain a faculty of modal intuition which enables us to *intuit* irreducible physical possibilities and necessities. This would be something like direct observation, albeit of a more subjective character.

However, the mere postulation of a faculty of modal intuition does not solve the problem: we have to hear a reasonably detailed account of such a faculty—examples of its function, a theory about how it works, and so on. Absent such an account, talking about modal intuitions or modal-detectors is quite mysterious indeed, a kind of deus ex machina with little or no explanatory value.¹

I conclude that the only way to discover physical modalities is to infer them. Now, to reiterate, our problem is this: Given that we can only observe actualities, how can we infer that something is necessarily the case? In trying to answer this question, the realist must be careful that his account does not

¹ There is a rationalist tradition that tries to explain talk about intuitions regarding necessary truths by theorizing about divine illumination—see Adams (1983: 751). However, it seems that the intuitions in question are all supposed to be about *a priori* propositions. Assuming that at least some scientific propositions are only *a posteriori* knowable, it seems that here realists about physical modalities cannot avail themselves of a readymade rationalist theory of modal intuitions, whatever other merits such a theory may have.
collapse into the relativistic reading of modalities. Here I focus on the ways a
realist could try to make sense of physical *necessities*: as mentioned in the
previous chapter (the case of the open-minded physicist), it is not clear how
one can, even in principle, give an account of irreducible non-actual
*possibilities*, save for saying that non-actual possibilities are situations
regarding which statements describing them are compatible with our
background theory.

I will examine three modes of inference: inductive, abductive, and
deductive.

**A. Inductively.** This idea is discussed in some detail in the opening pages of
Appendix II to Nathan Salmon’s (2005b), and part of my exposition draws on
his. Following Salmon, let me make it explicit that here the position I am
arguing for is *not* sceptical about induction in the actual world. For all that
matters here, I will assume that it is perfectly acceptable to infer that, say, all
elephants are pink, based on the observation that elephants in a good sample
were all observed to be pink. Of course there are many difficult questions
about induction; but our topic here is not theorizing about induction per se—
in particular, for our current purposes we need not know what exactly we
mean by saying that a sample is a good one, or what the difference is between
the predicates ‘is pink’ and ‘is grue’. The question I want to investigate here is
whether we can inductively infer physical necessities based on observed actualities.

How can we inductively infer that so and so is necessarily the case?

Suppose that, by whatever means, our scientist has concluded that a non-modal statement $\phi$ is the case in the actual world. To inductively infer that $\phi$ is necessarily the case, he has to do an inter-world induction: in his sample of possible worlds, each world is such that $\phi$ is true in it; therefore, he concludes by way of induction, so is every possible world. Thus $\phi$ is necessarily true, he concludes.

The problem with this induction is obvious: it is based on exactly one sample — of all the possible worlds, our scientist knows about only one of them, namely the actual world, that in it $\phi$ is true. Now we do have other apparent cases of induction based on a one-member sample — perhaps this is how we infer that those other things beneath hats and clothes are not automata. But cases like other minds deal, as Salmon observes (2005b: 255), with “a manageable number of entities, human bodies, that are similar in a great many important and relevant respects to the single instance on which the inference is based, one’s own human body.” But in the case of an inter-world induction we deal with a presumably huge collection of worlds, some
of which are presumably very different from the one we know. The induction here seems to be a poor one— it is based on highly insufficient evidence.²

The realist might object that the induction here is not that bad. Salmon is talking about metaphysical necessity; thus he had to consider all possible worlds. Perhaps the case would be—as Salmon himself suggests— less bad were we to look only at nomologically possible worlds, since these worlds are supposed to be similar to ours: at least they are governed by the same laws as ours. There are a number of reasons against such a move, though.

Suppose I examine a piece of gold and I observe that it weighs 23 grams. Now even if gold is a natural kind, inferring that all instances of gold weigh 23 grams is not a good inductive inference—although instances of gold share a number of properties, weighing 23 grams is not among them. Thus, obviously, even for members of the same natural kind in the actual world, which are supposed to obey the same laws, we are not always ready to jump to a general claim based on just one observed instance; why should we think that nomologically possible worlds, even if they form a “natural kind” in a much stretched sense of the term, behave in a more orderly way than natural kinds within the actual world?

Presumably, advocates of physical necessity do not want to say that every statement which is true in the actual world is physically necessary. Let ψ be a

² For what appears to be a similar criticism, see part II of Hume’s (1759) Dialogues about one-sample inductive inferences. (I owe the reference to Hume to Sober (2005).)
statement that is true in the actual world but not supposed to be physically necessary, and let \( \varphi \) be a statement, true in the actual world, regarding which scientists are supposed to inductively infer its physical necessity from its actual truth— for instance, we may think of

\[
\psi: \text{Every piece of gold is lighter than } 10^{25} \text{ kg.}
\]

\[
\varphi: \text{No object accelerates unless a force acts on it.}
\]

So far as the inductive base is concerned, \( \psi \) and \( \varphi \) are (by assumption) similar: they are both true in the actual world. To say that induction is justified in one case but not in another, the inductivist needs reason other than truth in the actual world, and in the presence of such reasons induction seems to be idle. (Note that someone who thinks that we can inductively infer physical necessities cannot rest content by saying that inter-world induction is \emph{reliable} and the case of \( \psi \) only shows that inductive inference of physical necessities may not \emph{always} give the right result. The point is that our inductive base makes no distinction at all between physically necessary statements and “accidentally true” statements.)

Can’t it be of any help to note that here we are dealing with \emph{nomologically} possible worlds? Consider the case of intra-world induction again: we observe one piece of gold, and it weighs 23 grams. Why is it a bad induction to infer that, in the actual world, all pieces of gold weigh 23 grams? Perhaps
because we know that there is no important relation between mass and
goldhood—there is no “law of nature” about the masses of pieces of gold.
Now one might think of differentiating the cases of \( \psi \) and \( \varphi \) in a similar
fashion: there is no law of nature about masses of pieces of gold; hence in a
nomologically possible world it may or may not be the case that \( \psi \) is true. On
the other hand (let us assume), there is a law of nature about acceleration and
force; hence in a nomologically possible world, which is governed by the
same laws as our world, \( \varphi \) is true. Nevertheless, whatever the merits of such a
move, note that here \textit{induction} plays no rôle: we have reasons other than
inductive inference to think that \( \varphi \) is physically necessary.

Let me conclude with an observation about the difference between intra-
and inter-world inductions. In the case of induction within the actual world,
there is, in principle, a way of confirming or disconfirming inductive
projections: we may examine the colour of the next elephant, we may wait
and see if the sun will rise tomorrow, or if the farmer will feed the chicken
again. On the contrary, since we cannot observe other possible worlds, we
cannot confirm or disconfirm our inter-world inductive projections in the
same way we do for inductive projections in the actual world. I submit that
this difference is big enough to cast reasonable doubt on the practice of inter-
world induction.
But perhaps all this was a fight against a straw man, as it is not clear if any philosopher thinks of induction as a way of acquiring knowledge of physical necessities. Let me turn to a more promising way.

**B. Abductively.** Our physicist might observe certain regularities in the actual world and ask herself a why-question. Simplifying, let us assume that she has repeatedly observed that an object does not accelerate unless a force acts on it. By whatever means, she concludes that this is generally the case in the actual world, and asks why that is so. She then offers the following as an explanation (assuming that we have a relatively clear notion of “physically possible”):

\[
\text{L: In every physically possible world, no object accelerates unless a force acts on it.}
\]

This, our scientist suggests, explains the observed regularity in the actual world, for, obviously, the actual world is a physically possible world.\(^3\) For whatever reason, she also considers this the best explanation for the truth of the modality-free form of Newton’s first law. She then infers L.

---

\(^3\) This can be considered as a deductive-nomological explanation: here L plays the rôle of a general law, while the actual-world Newton’s first law can be considered a particular fact.
Can this kind of abductive reasoning introduce modality to physicists’ statements? For the sake of the argument, I will here treat abduction as an acceptable rule of inference (though, of course, there are empiricists who are against it; see, in particular, van Fraassen (1989)).

I take it to be plausible to say that, in order to be warrantedly abductively inferred, the best explanation must be a good one. To refer to two friends of abductive reasoning: Cartwright (1983: 111) writes, “There obviously must be certain provisions to ensure that ‘the best [explanation]’ is good enough.” Also, at least in part of his (1993), Lipton offers a way of modifying the normal practice of abduction to ensure that the best explanation is good enough.4 Thus abduction should be understood as a license to infer the best from a collection of acceptable or reasonable explanations; otherwise there will be no real difference between abduction and the practice of inferring a theory because it is the only one in town.

Is L a good explanans? Of course sometimes offering a generalization of the explanandum is a good explanation: I ask why my cat Naomi meows, and you say that all cats meow. However, in such a case the explanans is itself well confirmed or is backed by a good theory—say we have either a good theory about cats and their meowing, or at least have seen many cats meowing. Think of the generalization as a conjunction, say

---

4 Basically what he does is to doctor the set of explanans to make sure that the true theory is among them.
C: $\bigwedge_j (C_j \text{ meows}),$

where $\{C_j\}_j$ is a big set of cats including Naomi. Now C is a good explanation for why Naomi meows only if we have reasons to think that at least some of the conjuncts other than the explanandum are true. Compare this with the case of L and the actual world. L says that, in every physically possible world, no object accelerates unless a force acts on it:

L: $\bigwedge_j (\text{In } W_j, \text{ no acceleration without force}),$

where $\{W_j\}_j$ is the (perhaps infinite) set of all physically possible worlds, including the actual world. Now the problem is that of all the $W_j$s, we have tested the corresponding conjunct for exactly one of them, namely the actual world, and that is exactly the conjunct that L is supposed to offer an explanation for. It seems that, therefore, L is not a good explanation.

Or does L make good predictions? Well, it predicts that in every possible world every object remains unaccelerated unless a force acts on it. Yet, as we cannot test the case of other-worldly objects, the only tangible prediction made by L is the one about the actual world, which is not informative since this is what we started with: we assumed that our scientist had come to the
conclusion that in the actual world there is no acceleration without force, and she was looking for an explanation for *that*.

In general, then, it is not clear how the claim that $\phi$ is physically necessary can explain the actual truth of $\phi$ well. Nor can we infer the necessity of another claim, $\psi$, by saying that *it* explains the actual truth of $\phi$ well: true, $\psi$ might be such that if it holds in every physically possible world, then we have explanations for many actual truths, including $\phi$; but is such a $\psi$ *otherwise* testable?

**C. Deductively.** We observe a statement $\phi$ to be true in the actual world. Based upon this, can we inductively infer that $\phi$ is necessarily true? Salmon, *op. cit.*, answers this negatively only to move to his main concern, viz. the idea of *deductive* inference of the necessity of $\phi$. A paradigmatic example of such an inference is, of course, Kripke’s (1971, 1980) inference of ‘Necessarily, Hesperus is identical to Phosphorus’, based on the observation that, in the actual world, Hesperus is the same as Phosphorus. I take it to be an empirically discovered fact that Hesperus is the same object as Phosphorus.

Also, ‘$(x)(y)(x = y \rightarrow \Box x = y)$’ is a theorem of the popular modal logic S5; hence the necessity of the statement that Hesperus is the same as Phosphorus. Even without formal modal logic, one can see—as Kripke convincingly argues—that one cannot tell a coherent story about a situation wherein
Hesperus and Phosphorus *themselves* exist and are not identical. It seems, then, that we can infer necessities from observed actualities.

This, however, is not a great victory for the realist. Examples of this sort are not examples of the deductive inference of *scientifically significant* modal claims. It seems that science, at least in the realist’s conception, does not deal with particulars—that Hesperus is (necessarily) identical to Phosphorus is not a typical scientific statement. Science deals with substances and general properties: ‘Water is H₂O’, ‘No acceleration without a force’ — *these* seem to be typical scientific statements.

Can we apply the same Kripkean argument to ‘Water is H₂O’ and conclude that water is necessarily H₂O? In Chapters 1-3 I argued at length that we have a transworld identity problem for kinds (like water)—a problem that casts doubt on the coherence of the statement that water is necessarily H₂O. Even if there is a solution to the transworld identity problem, we still have the well-known problem of extending the notion of rigidity, which is crucial to Kripke’s Hesperus-Phosphorus argument, to the case of general terms like ‘water’.⁵ Hence, even if it is coherent to assert that water is necessarily H₂O, it is not clear how the assertion should be established.

The realist may not be moved by the arguments of Chapters 1-3 above about the problem of transworld identity; he may not be bothered by the

---

⁵ See the Introduction for a brief discussion of the problem of rigidity of kind terms.
rigidity problem either. Still, he owes us an example of the deductive
inference of a scientifically relevant modal statement. At least we need a clear
example.
I do not propose to meet the views that I disagree with by controversy, by arguing against those views, but rather by positively setting forth what I believe to be the truth about the matter, and endeavouring all the way through to make the views that I advocate result inevitably from absolutely undeniable data.

--Russell (1918: 36).

Here I cannot act like Russell when he was lecturing on his philosophy of logical atomism, partly because I do not think that the view I advocate results from absolutely undeniable data. In this chapter, then, I will examine some realist views of modalities. I recognize two kinds of views opposed to the relativistic account of modalities (as developed in Chapter 4 above), and I will devote a section to a discussion of a well-known representative of each. Thus Section 1 discusses some of the views suggested by the works of Nancy Cartwright in the philosophy of science, and Section 2 is a re-examination of Sydney Shoemaker’s causal theory of properties, which I understand as a work in metaphysics.

I have presented an account of modalities: in any given context and insofar as truth-conditions are concerned, necessity is a matter of deducibility
from the (non-modal) background theory of the context. I have offered this as a rational reconstruction of modal talk in science. How might one oppose this view?

1. One may draw our attention to a scientific discipline and argue that the relativistic account cannot make sense of the modal discourse in that particular discipline. However, in this chapter I will have nothing to say about this potential line of argument. For in Chapter 5 I argued that irreducibly modal a posteriori facts, if there are such facts, are epistemically inaccessible to scientists: scientists cannot observe such facts; nor—and this was the main thing argued for—can they infer them. Therefore, short of having a mysterious modality-detector faculty, empirical scientists cannot be dealing with irreducible modalities. This constitutes an argument to the effect that no counterexample (of the form specified above) to the relativistic account exists. Thus it would appear that the burden of proof lies with the opponent of the relativistic view: he should present a counterexample, explain how irreducibly modal facts can be known a posteriori, or refute the argument of Chapter 5 in some other way.

2. Whether or not scientists’ modal talk can be rationally reconstructed in a relativistic way, a philosopher of science might argue that in order to (philosophically) account for a scientific practice one has to appeal to irreducible modalities. Thus someone doing philosophy of physics might submit that in order to make sense of physics—in order to understand what
aspects of reality it is about, how its experiments yield knowledge, etc.—one has to talk about irreducible modalities. Nancy Cartwright might be an example of such a philosopher of science.\(^1\) I will examine some of her views in Section 1.

3. The debate might take the form of a debate in metaphysics, rather than in the philosophy of science. One might argue that, at least for a significant class of properties, certainly many of the properties examined in sciences included, the very notion of a property is an irreducibly modal notion. Here a notable example is Sydney Shoemaker’s causal theory of properties.\(^2\) An advocate of such a view might be silent on how we are supposed to know about modalities; yet he insists that many properties have some of their powers essentially, in the metaphysical sense of ‘essential’ defined in Chapter 1. Such a view will be considered in Section 2 below.

Since the relativistic account of modalities does not postulate irreducible modalities, its ontological cost is less than any opposing realistic account. I

---

\(^1\) There is some evidence that Cartwright’s modalities are in fact supposed to be irreducible, and that her view is thus opposed to the relativistic view of modalities. However that may be, so far as defending the relativistic account against a Cartwrightian threat is concerned, we need not investigate the question of irreducibility of her modal notions, because if what I say in Section 1 below is on the right track, then there are modality-free answers to her worries.

\(^2\) Amongst philosophers of science, more recent advocates of Shoemaker’s theory of properties include Mumford (2004) and Bird (2007).
have also argued, in Chapter 5, that there is an epistemic problem concerning irreducible modalities. In this chapter I will argue that Cartwright’s and Shoemaker’s considerations do not rationally force us to accept irreducible modalities. If I am successful here, then, overall, we have good reason to prefer the relativistic view to its rivals.

1. Cartwright.

One major thesis of Nancy Cartwright’s book on capacities is that science cannot be understood without capacities (1989: 1). Capacities are supposed to be closely related to causes, and, presumably, for Cartwright causation is a modal notion. This thesis of Cartwright’s is thus presumably a challenge to my project—if she is right, then one cannot make sense of scientific practice without appealing to the modal notion of capacity.

Below I will examine Cartwright’s view that capacities are indispensable for an understanding of scientific practice; before doing so, however, let me quickly justify the claims I made in the previous paragraph about the relationship between capacities and causation in Cartwright’s view, and her notion of causality. Regarding capacities and causation, she writes “I maintain that the most general causal claims [...] are best rendered as
ascensions of capacities” (1989: 141; cf. 1989: 179).\(^3\) Next, causation is, for Cartwright, closely related to probability, as can be seen via her principle CC (1989: 56, with an amendment on 1989: 96). Roughly speaking, CC says that C causes E if and only if C increases the probability of E. Regarding probability and modality, Cartwright says that the notion of probability she is concerned with might be compared to Humean regularities, but her probabilities are “in an important sense different from Hume’s regularities. For *probabilities are modal or nomological* and Hume’s regularities were not” (1989: 35, my italics).\(^4\)

In her (1989), Cartwright starts using the terms ‘capacity’ and ‘capacities’ in the fourth chapter, in the introduction to which she says that she had already argued, in her Chapter 1 and Section 2.4, that capacities can be measured. Chapter 1 of her (1989) is about inferring causes from probabilities, and the title of Section 2.4 is “How Causal Reasoning Succeeds”. The structure of the book and its arguments is rather complex, and the book is rich in the variety of the disciplines it examines—it contains case studies from sociology, econometrics, and physics, to name a few. To keep the discussion manageable, here I will focus on a later paper of Cartwright’s, “Aristotelian

---


\(^4\) She does not elaborate on this, but, in a footnote, refers to van Fraassen (1980).

Before examining some of Cartwright’s views about natures, an important remark should be made. To understand her views and evaluate her arguments, we naturally need a relatively clear understanding of Cartwright’s idea of the notion of natures of things. She clarifies that her conception of nature is not Aristotelian because, among other things, she does not identify natures with essences. Thus (1992: 47) emitting light is in the nature of an excited atom, but to emit light is not what it is to be an atom in an excited state. Hence Cartwrightian natures are not essences. What are they, then? Here is some further elaboration (1992: 48):

Natures are like powers. To say that it is in their nature [i.e., in the nature of two objects of charge $q_1$ and $q_2$] to experience a force of $q_1 q_2 / r^2$ is to say at least that they can experience this force if only the right conditions occur for the power to exercise itself; for instance, if they have very small masses so that the gravitational effects are negligible. It is also to say that their tendency to experience it persists, even when the conditions are not right; for instance, when gravity becomes important.

---

5 With an additional introduction that relates it to the rest of her more recent works, almost all the material of her (1992) is also presented as Chapter 4 of her (1999).
This is terminologically informative, but it does not define ‘nature’ — and perhaps was not intended to do so in the first place, as may be suggested by its to-say-the-least formulation. So far as I can see, Cartwright does not define the term anywhere in her (1992). Neither does she define the term ‘capacity’ in her (1989), as some philosophers have noticed.

In a recent work she writes (2007: 197),

*Constancy of Tendency* This is the reason I introduced the idea of ‘capacities’ into my discussion of scientific laws in the first place. The outcomes that occur when the gravitational capacity operates are indefinitely various, but there is something fixed. The first mass is always trying to bring other masses closer to it; we say that a mass always attracts other masses no matter how the other masses actually move.

Thus, as we will also gather from my summary of some parts of Cartwright (1992), a kind of persistence is an important characteristic of natures. However, we are not told what exactly she means by natures (or capacities) of things — it is not entirely clear what it is for something to be, e.g., in the nature of electrons.

---

6 One such philosopher is Stathis Psillos, who writes in a section on Cartwright on causation that “no clear picture emerges as to what capacities are” (2002: 192). He then gives a list of the properties Cartwright attributes to capacities.
Absent a clear definition of ‘nature’ or ‘capacity’, it is hard to analyze Cartwright’s theory of natures to see if it works. I will therefore simply talk here about two of the troubles described by her, regarding which appealing to natures is supposed to offer a way out, and I will consider whether one can approach these problems without appealing to modal notions. This is not to pretend that I have complete solutions to her worries; however, due to the absence of a detailed theory of natures, it is not clear if appealing to natures would give us satisfactory solutions either. For a direct criticism of Cartwright’s (1989) theory of capacities—whether it works, and whether we can empirically find out about capacities—see Morrison (1995) and Psillos (2002: 189 ff).

With this long preamble, let us now consider some of Cartwright’s (1992) views. Of her main theses, one is that we cannot make sense of modern experimental method unless we assume that “laws of nature are basically about natures” (1992: 47).

**Coulomb’s Law and What is Not Measured.**

According to Cartwright’s characterization of it, Humean empiricism maintains that laws of nature are not about natures or powers but only about occurrent properties and forces, just about what things actually do. Cartwright disagrees with Humean empiricism: according to her, “fundamental laws are generally not about what things do but what it is in
their nature to do” (1992: 48). To support this claim she looks at a well-known law in physics.

Consider Coulomb’s law, which says that the electrostatic force between two objects of charges \( q_1 \) and \( q_2 \) is \( \frac{q_1 q_2}{4\pi \varepsilon_0 r^2} \). Cartwright observes that this is not the net force that the two charged objects experience, for they are also subject to a gravitational force, whose size is given by \( Gm_1 m_2 / r^2 \). The force whose size is given by Coulomb’s law is therefore not an occurrent force, and Coulomb’s law is thus not about occurrent properties. Even if there were no intervening gravitational forces, the number given by Coulomb’s law would still fail to be the size of an occurrent force, for there are other charged particles in the universe that electrostatically act on our two objects.\(^7\)

Of course we may say that Coulomb’s law gives the size of the attractive or repulsive force between the objects which is due to their charges; but, according to Cartwright (1992: 48), this is not a notion the empiricist would like to appeal to because the numerical value given by Coulomb’s law is not the size of the force we measure, for there are other forces acting on the two objects. (We may also think that the due-to talk is not something an empiricist

---

\(^7\) Though she does not mention interfering electrostatic forces, Cartwright is certainly aware of this abstraction. Her (1989) contains a chapter on abstraction, its reverse process of concretization, and idealization. Cartwright admits (1989: 185) that she does not have a complete solution to the problems raised by idealization and related issues. For a criticism of the methodology and arguments of that chapter of Cartwright (1989), see Humphreys (1995).
would likely engage in, because, presumably, the empiricist does not want to say that such and such is responsible for the electrostatic force—there are just regularities, with no causation over and above them, according to him.)

Cartwright also rules out one possible move on behalf of empiricism: she says that the empiricist cannot say that the force whose size is given by Coulomb’s law is the net force that the charged objects would experience if their masses were zero, because the antecedent of such a counterfactual never actually obtains. Cartwright thinks that the best account of the law in question is in terms of natures: “Coulomb’s law tells us not what force charged particles experience but rather what it is in their nature, qua charged, to experience” (ibid.).

Regardless of the merits that talk of natures may have, let us see if a Humean empiricist can deal with the problem Cartwright raises. Suppose that we have two objects with charges \( q_1 \) and \( q_2 \) (expressed by real numbers: positive, negative, or zero), and with masses \( m_1 \) and \( m_2 \) (expressed by positive real numbers). We may say that the size of the force these objects experience is

\[
-q_1 q_2 / 4\pi \varepsilon_0 r^2 + Gm_1 m_2 / r^2.
\]
(If there are more than two objects to be considered, the force acting on each is given by the obvious vector sum.) This is something testable, and to test it we do not have to consider the natures of things.

Of course this is not a law about the force these objects experience “due to their charges”, but Cartwright’s challenge for the empiricist is to present a confirmable law devoid of the expression ‘due to’ — the empiricist’s position is safe insofar as we have a law about the occurring force. Neither is this Coulomb’s law (we may call it the ‘Coulomb-Newton law’); but why should this matter? Pushed by Cartwright, the empiricist may officially forget about Coulomb’s law and just talk about the Coulomb-Newton law.

It is undeniable that for Newton or Coulomb something like natures talk might have played a rôle in the act of discovery. For instance, perhaps it was because of his beliefs about the nature of charged objects (including beliefs about the circumstances under which natures express themselves) that Coulomb decided to set his experiments the way he actually did—perhaps he would not have used his torsion balance or torsion pendulum in his experiments involving electricity if he had had a different attitude towards natures: perhaps if he had had different views about the nature of electric charge he would use different apparatus, and perhaps if he did not believe in natures he would have had “no way of determining which principle is tested by which experiment” (Cartwright (1992: 50-51)), and would therefore never have made any experiments. But this does not seem to be a good reason for
admitting natures—whatever they are—to our ontology. (As an analogy, suppose that Platonism in mathematics was refuted; it might still be a psychological fact that without thinking of mathematical objects as being “out there” in a Platonic heaven no mathematician would be able to discover important theorems.) It does not seem unreasonable to say that at the stage of discovering something, perhaps an overt talk of natures is permissible; afterwards, insofar as the testable law we find is verifiably true, we can think of our heuristic apparatus as a ladder to be thrown away. Explaining the heuristics seems to belong to the psychology of discovery, rather than the philosophy of science.

**INDUCTION.**

“For anyone who believes that induction provides the primary building tool for empirical knowledge,” writes Cartwright at the beginning of her third section (1992: 51), “the methods of modern experimental physics must seem unfathomable.” Here is one problem Cartwright has in mind: How is it that, based upon a very slim inductive base, we infer a general principle? In some cases, like the case of Newton’s critical experiment with prisms that she discusses in a separate section (1992: 62-69), a single experiment is our basis for inferring a universal law. Or we carefully measure the mass of one electron, and then we think we know the mass of all electrons. (Interestingly,
Cartwright observes, typically such a generalization is considered to be much more certain than large-scale social studies backed with a great number of instances.) How is that so? What is the justification for such an inference?

Cartwright’s claim here is that we need “very strong assumptions” (1992: 51), to the effect that fundamental particles do not change their masses or charges—mass and charge of a fundamental particle do not vary, “so long as it continues to be the particle it is”.

Cartwright agrees that, on the face of it, what we need in order to warrant inductive generalizations is just an assumption—which does not seem to rely on natures—to the effect that all systems which are situated like the one we use in our experiment behave in a way similar to the tested system. But she thinks that this assumption is not ultimately helpful. According to Cartwright, the Humean principle “same cause, same effect” does not hold in nature: outside the artificial setting of an experiment, “what happens in one instance is rarely a guide to what will happen in others” (1992: 52). This is an idea Cartwright has advocated for many years. (In her (1999: 49 ff), she uses the term ‘nomological machine’ to refer to a system wherein the principle does hold.) In doing experiments, we look for those very special situations which lend themselves to generalizations. Cartwright’s central thesis here is that what makes such situations special is that they “permit a stable display of the nature of the process under study, or the stable display of the interaction of several different natures” (1992: 52).
Cartwright reminds us that in devising experiments, we do not care about every single aspect of the experiment apparatus; nor do we provide complete descriptions: we only care about salient characteristics and relevant similarities. She admits that in each particular case we may give a natures-free description of what makes a specific combination of factors the right one; but she thinks that without appealing to natures one cannot give a general account of what we really do. Even in a particular case, no natures-free description allows us to see why we make an experimental device in the very particular way we do. Her favourite example is a certain Gravity-Probe-B experiment, “to trace the precession of four gyroscopes in space, to see how they are affected by the space-time curvature relativistically induced by the earth” (1992: 51; see also 1989: 66 ff). One important question in this experiment is what material one should choose for gyroscopes, and the answer is fused quartz, because of its high degree of homogeneity. This is because those who perform experiments with the device want to eliminate sources of non-relativistic precession, such as inhomogeneous materials. This natures-free explanation is not enough, says Cartwright, for we should “take account not only of what else might cause precession but also of what kinds of features would interfere with the relativistic precession, what kinds of factors could inhibit it, and what is necessary to ensure that it will, in the end, exhibit itself in some systematic way” (1992: 53). Cartwright thinks that successfully dealing with all these issues shows “what the nature of
relativistic precession is.” And how should an experiment be set so that one may infer a general law from it? Her answer (ibid.):

I claim that the experiment must succeed at revealing the nature of the process (or some stable consequences of the interaction of natures) and that the design of the experiment requires a robust sense of what will impede and what will facilitate this.

So, on the face of it, one problem that natures are supposed to solve is to justify, or to make sense of, slimly-based inductive inferences in experimental sciences. Nonetheless, my understanding is that justifying induction is not Cartwright’s main project here. Although Cartwright (1992) is not explicit about it, presumably her point is that we do not inductively infer what the mass of the electron is; rather, we find out the mass of the electron via investigating the nature of the electron in a carefully-set experiment. The function of natures is not to justify induction; it is to show how, in a non-inductive way, we infer laws and other general statements from experiments. In fact, in her book on capacities, or natures, she rarely talks about induction—its index cites only one page. Thus I think Margaret Morrison’s summary of one of Cartwright’s (1989) arguments is an accurate report of an argument in Cartwright (1992) as well: “[according to Cartwright,] once you accept that a particular effect is the
result of a capacity then one is at liberty to say that in all contexts (barring interaction) that capacity will be present, and it is that principle, not induction from cases, that motivates the inference” (1995: 165 f).

Once again, it is not quite clear if Cartwright’s appeal to natures would solve the problem of accounting for scientific practice, not least because we do not know what kind of thing natures/capacities are. However, the challenge remains: How can an empiricist justify the (apparently) inductive inference concerning, say, how much the electron weighs or what white light consists of, when he performs a very limited number of experiments?

Whether we observe a single case or hundreds of cases, justifying the projection to unobserved instances is an important problem in the philosophy of science, of course. However, the problem is not peculiar to the natural sciences. In a typical proof in traditional Euclidean geometry we seem to argue about a particular triangle (the one represented by a figure we refer to in the course of reasoning); then, after arguing that—for instance—the sum total of the interior angles of that triangle is 180 degrees, we infer that the internal angles of every triangle add up to 180 degrees. This seems to be something like a single-base induction; what justifies it?

John Macnamara (1991) examines this question. His idea is that while investigating the properties of that particular triangle, we do not use any properties of it except for what is implied by its trianglehood: so far as the proof of the geometrical theorem is concerned, the triangle we are arguing
about is an *arbitrary* one. Macnamara offers a definition of this and a number of connected notions, such as an *inductive sample* (1991: 37 ff). What makes his work relevant to our present topic is that his definitions are not restricted to mathematical discourse.

Like Gupta (1980), Macnamara thinks that to talk about an individual is always to talk about it *qua* being something (and that something is often a natural kind); in any given context, the name for that something is called an **anchoring noun**. Thus while arguing about the sum total of interior angles of triangles, we consider the planar figure *qua* triangle (not *qua* polygon or *qua* closed curve), and ‘triangle’ is the anchoring noun.

Now suppose that we do an experiment that shows that an individual \( i \) has a property \( P \). This individual is, by definition, **arbitrary** iff, “experimental manipulations apart,” the only properties of \( i \) used in showing that it is a \( P \) are that \( i \) belongs to the extension of the relevant anchoring noun, plus all the properties already shown to be shared by all the things in the extension of that anchoring noun (1991: 38).\(^8\)

---

\(^8\) In every experiment, some particular properties of an \( i \) are crucial in showing that it is a \( P \) – e.g., the piece of copper we are experimenting with in exploring the conductivity of copper has the property of being connected to the ammeter. Hence the qualification, ‘experimental manipulations apart’. Macnamara informs us that he has no general theory about what counts as mere experimental manipulation, but also says that he does not think this would undermine the enterprise of his paper (1991: 39).
This is, of course, a very sketchy report of Macnamara (1991). My main point is this: it seems that making sense of the notion of the arbitrariness of samples and having good reasons to think that the samples we experiment with are arbitrary is enough to justify the inductive inferences of the type Cartwright is worried about. In an experiment, we measure the mass of one electron. Why do we think that all electrons share this property with the one we experimented with? Because we have good reason to think that the electron in our experiment was an arbitrary one— to the extent that we are sure that samples are arbitrary, we are certain that the results are universal. There is a set of properties which characterize electrons; in the experiment where we measure the mass of an electron, we somehow make sure that nothing but those properties are used in our experiment.

Talking about *arbitrary* individuals in an empirical context here is not natures-talk in disguise— *an arbitrary N*, where N is an anchor noun, is not an empiricist’s euphemism for something like the *nature of N* or the *essence*...

---

Prior to offering his definitions, Macnamara gives examples of how inductive inference in empirical sciences works— see especially (1991: 28 f). In general, an inductive inference presumes some theories about its subject-matter. This could help us to approach an exact formulation of the notion of mere experimental manipulation— as a first attempt, one may think that i has property Q merely due to experimental manipulation iff our presumed theory does not imply that all members of the extension of the relevant anchoring noun have property Q.
of $N$ in any irreducibly modal sense of these terms like the one provided by
the definition of ‘essence’ suggested in Chapter 1. (Although something like a
definitional essence is needed to make sure that we are talking about the right
kind of thing.) Scientists somehow fix the references of the terms they use—
sometimes, like the case of ‘triangle’, the reference fixing is via one-line
analytical definitions; sometimes, most likely in cases like that of ‘electron’, it
comes with a description of the rôle that is supposed to be played by the
things to be named. Either way, to say that a statement does or does not
follow from our characterization of the extension of the anchoring name (plus
other statements that the scientist has already accepted) does not seem to
require any assumptions about irreducible modalities. To see if the individual
object that the scientist is experimenting with is arbitrary or not, he does not
seem to explore possible worlds.

Admittedly, ensuring the arbitrariness of the samples is occasionally
much more difficult in a physical experiment than in a mathematical case. In
a geometrical proof, we require justification for each step in the argument,
and we can make sure that nothing but the trianglehood of the triangle
(which is explicitly defined), and the already established properties of it, is
used. But while, say, measuring the conductivity of copper, how can we
know that the particular mass or shape of the sample we examined did not
“contribute” to its degree of conductivity? How do we know that
conductivity is independent of shape or mass? To ask the question without
metaphors and modalities: What is our justification for thinking that other pieces of copper with different shapes and masses have the same conductivity as the piece we experimented with?

Of course Cartwright has a similar problem to solve. In reply to Morrison (1995), she states the problem and briefly offers a solution (1995: 180, my square brackets):

How do we distinguish cases of impeded capacities from nonsense associations? Sometimes I sit on the left-hand side of the bus, sometimes I don’t. Do I have a tendency [i.e., is it in my nature, do I have a capacity,] to sit here? I have claimed that in the central uses of the concept, we assume that within the specified domain tendencies when properly triggered always “contribute” their characteristic behaviours unless there is a reason why not. To judge whether or not this is the case in a given situation requires a lot of knowledge, both about what other tendencies could be affecting where I sit and also about how this tendency naturally operates and how its power to do so is transmitted, what could distort it, what could enhance it, what could damp it and in what ways. We bootstrap it from other such knowledge, and so forth.

I think the same kind of solution is at least equally adequate to a natures-free understanding of scientific practice: we appeal to our previously established laws about the relationship, or lack thereof, between conductivity and other
properties. We then decide whether or not a particular piece of copper is arbitrary.

Overall, even if appealing to natures turns out to be helpful and devoid of obscurities, we have not seen a reason why their intended functions cannot be accommodated by more empiricist-friendly notions.

2. Shoemaker’s theory revisited.

According to Shoemaker (1980, 1998), properties are individuated, inter- and intra-worldly, by their nomic profiles: every property has its causal powers essentially, and in every possible world different properties have different causal powers. Thus one might think that the very notion of a property is a modal notion. Moreover, the modality here is not reducible to any non-modal notions: Shoemaker is concerned with truth in every metaphysically possible world (see below), not what follows from one’s background theory. Hence if science is supposed to investigate properties of things, it thereby has something to do with irreducible modalities, one might think.

In this section I will argue that Shoemaker’s view does not undermine the adequacy of the relativistic account of modalities as a rational reconstruction of modal talk in the sciences. Although I will briefly explain a worry about one of the distinctions made by Shoemaker in the course of his argument, the
main idea of the section is that the empiricist can coherently refuse to accept
the conclusion of Shoemaker’s argument.\footnote{For a summary of Shoemaker’s argument see Chapter 3 above. Shoemaker’s (1998) version
of the causal theory of properties has differences in comparison to his (1980) version; for
present purposes, however, the differences do not matter. In this section I will also set aside
some subtleties of the account—I will focus on Shoemaker’s main idea, as I understand it.}

The theory has two principal parts: (1) properties have their causal features
essentially, and (2) in no possible world do two different properties share all
their causal features. Perhaps after listening to Shoemaker’s arguments for
(2), even a radical empiricist would be happy to submit to the identity of
causally indiscernible properties; here what matters for our discussion is (1).
Shoemaker makes it explicit—e.g., in (1980: 211n8) and (1998: 408)—that here
the necessity in question is \textit{metaphysical} necessity and not just nomological
necessity, assuming the latter to be a weaker notion. Also, ‘possible world’
here means metaphysically possible world.

By ‘property’ Shoemaker means “real”, “intrinsic”, or “genuine” property,
where “A property is genuine if and only if its acquisition or loss by a thing
constitutes a genuine change in that thing” (1980: 207). It seems that, for
Shoemaker’s purposes, it is also crucial for a real property that by means of
referring to it, one can explain certain events. “For example,” writes
Shoemaker (1980: 207) about one of his examples of non-real properties,
my typewriter has the property of being over one hundred miles from the current heavyweight boxing champion of the world. It is not easy to think of a way in which its having this property could help to explain why an event involving it has a certain effect [...].

Shoemaker’s theory is not intended to be true of “mere-Cambridge” properties—the causal theory of properties is not about properties like the one mentioned above. Likewise, predicates like ‘is grue’ or ‘is such that Chomsky was born in 1928’ do not correspond to properties to which Shoemaker’s account is meant to apply. Shoemaker does not present a thorough discussion of the real vs. mere-Cambridge distinction for properties, but perhaps this is not a serious defect of his argument—perhaps a mere intuitive difference suffices.

However, one may doubt whether there is such a distinction to be made. More specifically, it seems that for at least some of Shoemaker’s paradigmatic examples of mere-Cambridge properties, one can in fact tell coherent stories about the real changes they might inflict on things that acquire them, and about some facts that could be explained by appealing to such properties. This does not seem to be difficult for the property of being a bed having been slept in by George Washington (1980: 208): whether or not a bed has this property seems to be empirically testable in principle, and it means a lot to an
antique collector and it may be of much importance to a detective investigating a two-century old murder case. As for the property of being over one hundred miles away from the current heavyweight boxing champion of the world, consider this story. A week before the great match, the challenger of the title realizes that he cannot beat the current champion. In a moment of fury, he decides to take the drastic measure of hitting the champion by a rocket which will destroy everything in the vicinity of one hundred miles from its target. (At the moment of launching the rocket, the challenger is in Chicago, the champion in Philadelphia, more than a hundred miles away from the challenger.) At a certain time, if Shoemaker’s typewriter fails to have the property that Shoemaker is talking about, it will be destroyed—not an unimportant effect.

Perhaps with stronger imaginative power one could even tell stories about how the grueness of an object can cause real changes in things, and changes that are not caused by blueness or greenness. I will not try. Having flagged this worry about the emptiness of the notion of mere-Cambridge properties, I will now assume that the distinction is valid and that there are in fact examples of mere-Cambridge properties.

In Chapter 3 I summarized an argument for part (1) of Shoemaker’s theory. Very briefly, the idea is that since, intuitively, a real property cannot change its causal powers through time and become a property with new causal powers, there is no possible world in which a real property has causal powers
different from its actual powers. Hence for every real property it is necessary that it has the causal powers it actually has.

Is this not a good argument for the claim that genuine properties have their causal powers essentially? I think it is. In fact, in my argument against one way of solving the transworld identity problem for kinds (the stipulative solution, Chapter 3), I argued that every good essentialist should accept Shoemaker’s theory. Where is the disagreement, then?

The disagreement is about the prerequisites of the argument. By this I mean something more basic than disagreement about the premises of the argument; I mean something like disagreement about its presuppositions. If one agrees to engage in possible-worlds talk and accepts that it does make sense to talk about properties in different possible worlds, then, I think, one cannot but accept that real properties have their causal powers essentially. Not so if one has qualms concerning the coherence of possible-worlds talk about properties. In Chapters 1-3 I tried to cast doubt on the presumption that it is clear what we are talking about when we talk about a property or a kind in a non-actual possible world: it is not clear, for example, what the property of greenness in a non-actual possible world is. If one refuses to accept that the notions involved in Shoemaker’s argument are in fact coherent, then one can refuse to accept Shoemaker’s conclusion.

Even if Shoemaker wants to reduce real properties to their causal profiles (something from which he distances himself in his later work (1998: 413)), this
would not solve the problem, for although we would not have a transworld identity problem directly for the property of, say, greenness, we still have a transworld identity problem for causal powers that would define greenness: we do not know what those properties in a non-actual possible world are.

However, in order to see that one can consistently refuse to accept Shoemaker’s conclusion with its “strikingly non-Humean consequences” (1980: 107), one need not buy my arguments about transworld identity. In Chapter 4 I argued that one can make sense of scientific practice without committing oneself to the existence of irreducible modalities. If my account is coherent, then it is coherent to refuse to engage in possible-worlds talk, at least while trying to account for scientific practice.

If (e.g., because of a transworld identity problem) possible-worlds talk of properties is problematic or incoherent, then we have an excellent reason for not accepting Shoemaker’s thesis. On the other hand, even if possible-worlds talk of properties is coherent, the question of this chapter remains: Do we have to engage in such a talk to understand scientific practice? He who thinks that the answer is negative has good reason to avoid such a talk—apart from ontological and epistemological worries, he does not see anything in scientific practice that he cannot explain. It is not a good argument against him to say that if he chooses to talk about possible worlds then a good deal of properties will turn out to have their powers essentially.
If we refuse to engage in possible-worlds talk (and I have argued that at least we *can* do so, even if—despite my arguments in Chapters 1-3—we do not *have to*), then Shoemaker’s argument is a non-starter. And this is fortunate for the empiricist, for I think Shoemaker’s conclusion is inevitable if the argument starts.
Appendices

1. Modality de re and de dicto.

A standard text on modal logic, Hughes and Cresswell (1996: 250), defines a formula to be de re iff it contains an occurrence of a free variable within the scope of a modal operator, and defines a formula to be de dicto iff it is not de re. According to this definition, ‘□(x)Qx’ is de dicto, while ‘(x)□Qx’ is de re. The definition, which might be good enough for some mathematical, modal logical purposes (e.g., some eliminability theorems), is uncritically repeated in some technical papers—Kaminski (1997) is an example. The definition seems to capture what is suggested by the Latin phrases: a de re formula, defined in this way, attributes some modal properties to an object (a res), while a de dicto formula attributes modal properties, if any, only to a sentence or a proposition (a dictum). According to this definition, if ‘a’ is a constant symbol of the language, then ‘□Qa’ is de dicto—think of ‘necessarily, Socrates is mortal’, where the domain of ‘necessarily’ is the sentence ‘Socrates is mortal’.\footnote{I am not here concerned with the extent to which the definitions are compatible with the mediaeval sense(s) of these terms; consulting Kneale (1962) might be useful in this regard.}

However, if the definition is motivated by the idea that de re formulae attribute modal properties to objects, it is be odd to consider ‘□Qa’ as a de

\footnote{I am not here concerned with the extent to which the definitions are compatible with the mediaeval sense(s) of these terms; consulting Kneale (1962) might be useful in this regard.}
dicto formula, for it attributes the property of necessary $Q$-ness to the object $a$.

Moreover, if the opponent of de re modality challenges the meaningfulness of attributing a modal property to an object (rather than to a proposition or a sentence), it should not matter to her whether or not the object in question has a name in her language. Thus, following Forbes (1985: 48 f), I use ‘de re’ in what is called by Fine (1978: 143) the ‘strict sense’:

**DEFINITION.** A formula is **de re** iff it contains a free variable or a constant symbol within the scope of a modal operator; otherwise it is **de dicto**.

2. **Haecceitism.**

In a celebrated, much-quoted passage (1975: 722 f), David Kaplan baptized a view about transworld identity:

The doctrine that holds that it does make sense to ask – without reference to common attributes and behavior – whether *this* is the same individual in another possible world, that individuals can be extended in logical space (i.e., through possible worlds) in much the way we commonly regard them as being extended in physical space and time, and that a common “thisness” may underline extreme dissimilarity or distinct thisness may underlie great resemblance, I call **Haecceitism**. (I prefer the pronunciation Hex’-ee-i-tis-m.) It
would be more exact to speak of Haecceitism with respect to a given kind of entity, but for present purposes we may assume that only individuals are in question and that our individuals are themselves some well-defined kind of entity, perhaps animals. [...]

Haecceitism holds that we can meaningfully speak of a thing itself—without reference either explicit, implicit, vague, or precise to individuating concepts (other than being this thing), defining qualities, essential attributes, or any other of the paraphernalia that enable us to distinguish one thing from another.

Anti-haecceitism for Kaplan (ibid.) is the view that “for entities of distinct possible worlds there is no notion of trans-world being.”

Note that if an individual \(x\) has an essence \(e\) in the sense defined in Chapter 1, then it is certainly legitimate to talk about that individual by means of reference to one of its properties, namely \(e\): by definition, in every possible world an object is \(x\) if and only if it has the property \(e\). But, as noted by Kaplan (1975: 723), haecceitism need not be committed to the view that objects have essential properties. Also note that the very notion of an essential property of a kind of thing presumes haecceitism about that kind of thing; one cannot talk about an essential property of an individual unless it makes sense to talk about one and the same individual in different possible worlds.
NOMENCLATURE. Intuitively, qualitative properties (a.k.a. purely qualitative properties) are those that can be expressed without any use of proper names or indexicals in an otherwise extremely rich language. If need be, we may even assume that the language permits infinitary disjunctions; this expressive power is occasionally needed for arguments—an example is Adams’s (1979:21 f) result, to be explained in Appendix 3 below. For more detailed characterizations of the notion and some methodological remarks, see Lewis (1986: 221) and Adams (1979: 7 f).

A qualitative property is also called a suchness. A thisness (or a haecceity) is the property of being identical to a given individual. Thus the property of being Noam Chomsky is a thisness; the predicate ‘is identical to an individual’ does not express a thisness.

Following Lewis (1986: 221), let us say that two possible worlds are haecceitistically different from each other iff there is no qualitative difference between them and, moreover, they differ with respect to their representation of some individuals. (An example, discussed in Chisholm (1967), is presented in Section 3 of Chapter 3.) For Lewis, the core idea of anti-haecceitism is that there are no haecceitistically different possible worlds, while haecceitism is the view that there are some instances of haecceitic difference. In Bird’s formulation (2005: 444), anti-haecceitism is the view that
“the transworld identity of particulars does not supervene on their qualitative features”.

But this doctrine seems to be conceptually different from the view Kaplan calls by the same name, for, at least prima facie, it could be the case that while it is meaningful to talk about transworld identities of objects without any reference to their attributes, it is nevertheless a matter of fact that there are no haecceitistically different possible worlds—just in the same way that though it is meaningful and legitimate to talk about the set of all sets, it is demonstrable that no such set exists. A good piece of evidence for my claim that Lewis’s formulation of haecceitism is different from, and actually stronger than, what is meant by Kaplan is that Kripke, who is one of the main advocates of haecceitism à la Kaplan, officially abstains from taking a position on the existence of haecceitistically different possible worlds—see Kripke (1980: 18n17).2


Regarding Lewis’s theory of possible worlds as developed in his *On the Plurality of Worlds*, Robert Black argues that (2000: 92)

while rejecting haecceitism for individuals, Lewis assumes almost without argument that there is no problem about identifying qualities across possible worlds. […]

---

2 This point is also made by Salmon (1996: 205), who also mentions Kripke’s (1980: 18n17).
Lewis explicitly allows, for example [(1986: 162)], a world isomorphic with ours, but where one of the quark colours has traded places with one of the flavours. To describe such a world, we have to assume as primitive the identity of qualities across worlds. Nothing *constitutes* the fact that a certain quality playing a certain nomological role in that world is identical with a certain quality playing a different role in ours; they are just the same quality, and that’s all that can be said.

Black coins the term *quidditism* for the view that there is “primitive identity between fundamental qualities across possible worlds” (ibid.)—here a *quiddity* or a *quidditas* is the thisness of a quality. One of Black’s arguments is that quidditism implies that no cardinality can be assigned to the totality of possible worlds. (Apparently, there is no similar argument against haecceitism.) For more on quidditism see Mumford (2004: Ch. 9).

3. Adams’s observation.

Robert Merrihew Adams argues (1979: 21 f) that the transworld identity of individuals of a certain kind is primitive if and only if the corresponding thisnesses of those individuals are non-qualitative. Here I will reconstruct the argument for the ‘only if’ part in a more detailed way than the original presentation by Adams.

In this reconstruction, ‘x’ and ‘y’ stand for *individuals*, not properties: the result cannot be trivially extended to the case of thisnesses of properties,
because, unlike individuals, (the extensions of) properties are not members of the domains of possible worlds—the extensions of properties are subsets of the domains.

So the claim is that if the transworld identity of \( x \) is primitive, then the thisness of \( x \) is not qualitative. The argument goes by assuming that the thisness of \( x \) is qualitative, and inferring that the transworld identity of \( x \) is not primitive. The argument is also based on the further assumption that no two possible worlds share all their qualitative properties. (In Lewis’s terminology, the assumption is that there are no haecceitistically different possible worlds. As we saw in Appendix 2, Kripke thinks that this assumption is compatible with his views. The result thus may shed some light on Kripke’s view on stipulating the identities of individuals.)

Suppose that the thisness of \( x \) is qualitative. Consider the collection \( W \) of all possible worlds in which \( x \) exists. For each \( w \) in \( W \), let \( S(w) \) be the totality of suchnesses of \( w \), and let \( s(w) \) be the totality of all suchnesses of \( x \) in \( w \). For every \( w \) in \( W \), let \( h_w \) be the property described as having all and only the suchnesses \( s(w) \) in at least one world which has all and only the suchnesses \( S(w) \). It is important not to consider this expression of \( h_w \) as a (doubly) quantified one: it has to be understood as a (perhaps infinite) disjunction of (perhaps infinite) conjunctions. In defining \( h_w \) we do not refer to possible
worlds; we just put together some suchnesses, using ‘and’ and ‘or’. The property $h_w$ is thus itself a suchness.

Finally, let $h$ be the (perhaps infinite) disjunction of these properties—that is to say, $h$ has every $h_w$ as a disjunct, and nothing else, where $w$ ranges over $W$. The property $h$, being a disjunction of suchnesses, is itself a suchness, a qualitative property.

Now $x$ has the suchness $h$ in every possible world where $x$ exists. The reasoning is straightforward; but to master the construction, let us see why. Let $v$ be a possible world in which $x$ exists. Then $v$ is in $W$, and it has, by definition of $S(v)$, all and only the suchnesses $S(v)$. Therefore $x$ satisfies the relevant disjunct of $h$, viz. having all the suchness $s(v)$ in at least one world [namely $v$] which has all and only the suchnesses $S(v)$.

Moreover, no other entity in any possible world has the suchness $h$. For suppose that $y$ has $h$. By construction, there is a possible world $u$ such that $u$ has all and only suchnesses of $S(w)$ for some $w$ in which $x$ exists, and, besides, $y$ has all and only the suchnesses of $s(w)$. Since $u$ and $w$ share all their suchnesses, they are one and the same: $u = w$. Now $y$ and $x$ share all their suchnesses ($x$ has all $s(w)$ by definition, and $y$ has all $s(u)$—that is, all $s(w)$—by

---

3 Talking about $w$ might seem to be redundant: it might seem that right away we may assume that $u$ itself is a possible world in which $x$ exists. But note that to make $h$ a suchness, in constructing $h$ we did not refer to particular possible worlds; rather, we worked with a bunch of suchnesses.
our current assumption); the thisness of $x$ being purely qualitative by the assumption made at the beginning of the argument, this implies that $y$ is identical to $x$. Therefore $x$ is the only thing which is an $h$.

Putting together the results of the last two paragraphs we see that $x$ is the unique object across possible worlds (wherein $x$ exists) that has the suchness $h$.

Thus far I have argued that if the thisness of $x$ is purely qualitative, then this thisness is (equivalent to) a suchness. But, looking at this suchness, we see that being $x$ is not primitive in the sense I defined in Chapter 3: we have a non-trivial necessary and sufficient condition for being $x$.

If one thinks, like Adams, of primitiveness in terms of explanations, here is why the transworld identity of an entity with a thisness equivalent to $h$ is not primitive—this is almost a quotation from Adams (1979: 21 f), with some notational changes:

Suppose that the thisness of every individual $x$ is purely qualitative. Then the thisness of an individual $x$ can be constructed as a disjunction of suchnesses, each suchness representing one possible career of $x$, as seen in the definition of $h$. It seems quite possible that in every case the grouping of disjuncts as alternative careers of $x$ could be explained by general principles about transworld identity of one or another kind of individuals, and the transworld
identity of \( x \) could be analyzed as consisting in the satisfaction of the general principles by the relevant disjuncts.

This shows that the thisnesses here are not primitive in my sense either.

In his (1981) Adams argues that “there are no thisnesses of individuals that never actually exist—although of course there could have been other individuals than those there are, and if they had existed they would have had thisnesses. I think this is a necessary truth; in no possible world would there be thisnesses of individuals that do not exist in that world” (1981: 4f).

4. The history (or lack thereof) of modal relativism.

The main idea behind the relativistic account of modalities—viz. that of reducing necessity to provability—seems to be quite well known, and to have been in the air for some time. This is evidenced by the fact that discussions of it do not typically attribute the idea to any particular philosopher or logician. Thus van Fraassen ascribes to “The nominalists, and subsequently the empiricists,” the view that “What is physically necessary is the same […] as what is logically implied by some tacit antecedent—say, the laws of physics” (1977: 71). He mentions no specific nominalist or empiricist with respect to this particular idea. Also, with a particular interest in the Peano Arithmetic as the background theory, logician George Boolos wrote a book (1993) on the interpretation of ‘necessarily’ as provability from a background theory. Like
van Fraassen, who devotes a whole section to “the logical history of the idea” (1977: 84), Boolos is scrupulous about giving credit for ideas; however, he does not tell us where the main idea came from. It seems safe to say that the idea is just part of the logico-philosophical folklore.

The idea of reducing necessity to provability seems also to be an obvious one. “Clearly,” writes Bob Hale (1997: 488) in a survey article on modality, whenever we have a more or less definite body of propositions constituting a discipline D, there can be introduced a relative notion of necessity — expressible by ‘It is D-ly necessary that’ — according to which a proposition will be D-ly necessary just in case it is true and a consequence of D.

In Appendix 5 below I will comment on Hale’s requirement for the truth of D-ly necessary propositions; apart from this extra condition, the idea here is just that of the relativistic reading of modalities. Like van Fraassen and Boolos, Hale gives credit to no particular writer.

The last source that I will mention in tracing back the idea is a paper by Richard Montague, whose main ideas “are based on the following unoriginal considerations”:

Let $\Phi$ be a sentence. Then *it is logically necessary that* $\Phi$ is true if and only if $\Phi$ is a theorem of logic; *it is physically necessary that* $\Phi$ is true if and only if $\Phi$ is
deducible from a certain class of physical laws which is specified in advance; it is 

**obligatory that** $\Phi$ is true if and only if $\Phi$ is deducible from a certain class of ethical 
laws which is again specified in advance.\(^4\)

The moral of this short historical review is that there is no originality in the 

mere idea of interpreting necessity as provability— the contributions of 

Anderson, Boolos, Montague, and van Fraassen are intended to solve some 
technical logical problems in *implementing* the basic idea, not to defend the 

idea that necessity can be understood as provability from a background 
theory as such. Likewise, if there is any novelty in my Chapter 4 above, it 

consists in working out this familiar idea in the realm of the philosophy of 
science, and defending it against some realistic intuitions. Also, as explicitly 

mentioned by Montague (1960: 71), his meta-theory contains no modality; 

while this can simply be *assumed* for a constructed formal language, in a 

scientific context one has to argue that there is no modality in the background 
theory, as I did in Chapter 5 above.

\(^4\) Montague (1960: 71). [The author uses Quine corners; I quote with some changes in 

accordance with the convention mentioned in the Introduction.] Montague’s paper, first 
presented in 1955, is one of the anticipators of the teenage Kripke’s (1959) possible-worlds 
semantics for modal logics. The particular case of ethics is also mentioned by van Fraassen 
(1977: 71), for which he refers to an article by Alan Ross Anderson (1958).
5. The failure of \( \Box \phi \rightarrow \phi \).

According to the relativistic account of modalities presented in Chapter 4 above, \( \phi \) is necessary is true just in case \( \phi \) logically follows from the background theory of the context. Given that background theories are not assumed to be true, it follows that in my account, a sentence which is necessary relative to a background theory \( T \) need not be true—in fact, every false sentence of \( T \) is necessary relative to \( T \). Propositions which are necessary relative to a theory \( T \) need not even be accepted, since there is no acceptance requirement on \( T \) either, as explained in Chapter 4. (But, of course, a necessary sentence is true if the background theory is true; also, if we know that \( \phi \) is necessary and we accept the background theory, rationality requires that we accept \( \phi \) as well.) Hence if we read the box relativistically, the famous law \( \Box \phi \rightarrow \phi \) does not hold, and so there is a difference between my account and Hale’s (1997) version mentioned in Appendix 4 above, wherein a sentence is said to be \( T \)-ly necessary iff it is true and it logically follows from \( T \).

Why does Hale say that relatively necessary propositions must be true? Perhaps part of the reason lies in a certain feeling of uneasiness in saying that something is necessary but not true—to say that a false \( \phi \) is \( T \)-ly necessary might sound awkward or odd. Perhaps such a feeling is a consequence of presuming that the notion of necessity is an “absolute” one; but relative to a set of sentences, none of which need be true, it does not seem problematic to say
that a false proposition is necessary— at least something like this is not a contradiction in terms: ‘Special-relativity-ly, it is necessary that no object moves faster than light; but perhaps we are mistaken: perhaps there are actually objects that move faster than light’.

At any rate, including a truth requirement in the definition of relative necessity comes at a price. Suppose we decide to say that \( \varphi \) is \( T \)-ly necessary just in case \( \varphi \) is true and is a logical consequence of \( T \). But then, given that as a matter of semantics \( \varphi \) is possible if and only if not-\( \varphi \) is not necessary, we are committed to say that \( \varphi \) is possible just in case not-\( \varphi \) is false or not-\( \varphi \) is a not consequence of \( T \). Therefore, if we follow Hale’s definition of \( T \)-ly necessity, we have to say that \( \varphi \) is \( T \)-ly possible just in case \( \varphi \) is true or logically consistent with \( T \). Hence every true sentence which is inconsistent with \( T \) would be \( T \)-ly possible. For instance, let \( T \) be a formulation of Ptolemaic astronomy. The true statement ‘The earth is not the centre of the universe’ is inconsistent with \( T \), and Hale’s definition implies that relative to Ptolemaic astronomy it is possible that the earth is not the centre of the universe. This is perhaps at least as awkward as saying that a \( T \)-ly necessary proposition is not true.

Also, it should be noted that in the modal logic GL (named after Gödel and M.H. Löb) wherein necessity is interpreted as provability, \( \Box \varphi \to \varphi \) is not an axiom. In fact, as demonstrated in Boolos (1993: 11), no sentence of this form is provable in GL unless its conclusion is a tautology. As GL is a
respectable modal logic with its own Kripke-style completeness theorem, the fact that $T$-ly necessary propositions need not be true cannot be criticized on purely logical bases. (More on the “missing law” of $\Box \varphi \rightarrow \varphi$ can be found in the last section of Montague (1960).)
References.

Numbers in brackets are meant to show, to the best of my knowledge, the publication year of the latest edition of the referred work as composed by its author(s); citations in the dissertation are from the works as published with the bibliographical details given below. Thus when refereeing to Putnam’s “Is semantics possible?”, I use the code ‘Putnam (1970)’, because Putnam published it in 1970 and, to my knowledge, did not change it afterwards. But when I quote from that essay, I quote from the 1997 reprint of volume 2 of Putnam’s philosophical papers, originally published in 1975.


Cook, Monte (1980). If ’cat’ is a rigid designator, what does it designate?. *Philosophical Studies*, 37: 61-64.


*Mind, 112*: 475-492.


*Philosophical Studies, 109*: 265-277.

Shoemaker, Sydney (1980). Causality and properties. Reprinted in Shoemaker, 


