

A DEFENSE OF THE SUPPOSITIONALIST VIEW OF HYPOTHETICAL ENTITIES

by  
Jonathon Hricko

A dissertation submitted to Johns Hopkins University in conformity with the requirements  
for the degree of Doctor of Philosophy

Baltimore, MD  
September, 2013

---

## *Abstract*

When scientists put forward hypotheses, they sometimes involve new kinds of entities, which we can call ‘hypothetical entities.’ Hypothetical entities are pervasive in the sciences, and some examples include caloric and, up until very recently, the Higgs boson. Some hypothetical entities are discovered, as was the case with the Higgs boson, while scientists conclude that others, like caloric, do not exist. Hypothetical entities pose a number of important challenges for the philosophy of science, and my goal is to develop and defend what I will call the suppositionalist view of hypothetical entities. In chapter 1, I examine the extant views of hypothetical entities, which I draw from the scientific realism debate. I argue that these views are all committed to the claim that terms for hypothetical entities putatively refer to empirical entities. In chapter 2, I develop the suppositionalist view of hypothetical entities. On this view, terms for hypothetical entities refer to what are called ‘objects of supposition.’ Examples of such objects from other domains include fictional characters like Superman and mathematical objects like the natural numbers. I draw from analogies with fiction and mathematics in order to develop the suppositionalist view in the scientific domain. In chapter 3, I give a history of a hypothetical entity that I will later use as a test case for views of hypothetical entities. In the late-eighteenth century, Antoine Lavoisier hypothesized that muriatic acid is composed of oxygen and a hypothetical entity

called the ‘muriatic radical.’ In the early-nineteenth century, Humphry Davy’s work on muriatic acid showed that it is actually composed of hydrogen and chlorine, and so muriatic acid is hydrochloric acid. Finally, in chapter 4, I use the history of the muriatic radical in order to argue against the extant views, and for the suppositionalist view. I argue that the former are committed to giving a history of the muriatic radical that is either whiggish or incomplete. The latter, however, can give us a non-whiggish history that is more complete, and hence it is preferable to the extant views.

**Dissertation Committee Members:** Robert Rynasiewicz (advisor), Steven Gross, Yitzhak Melamed, Wilda Anderson, Kyle Rawlins

---

## *Acknowledgements*

Two people stand out in my mind as playing a crucial role in my development as a philosopher. They are Rob Rynasiewicz and John Waterman. You are the people most responsible for shaping my views in philosophy, and so I'd like to begin by thanking you for this, and for all of your help and support.

Without my family and friends, I wouldn't have been able to finish this dissertation. An account of all of the ways in which you've helped me would most likely match the length of my dissertation. I thank all of you for your support and encouragement. I owe special thanks to Dan Hricko, Lee Hricko, Matthew Hricko, and Chitra Venkataramani.

I'd like to thank the following people for valuable feedback on the various papers, presentations, and chapter drafts that led to this dissertation: Marianna Bergamaschi-Ganapini, Justin Bledin, Nick Goldberg, Steven Gross, Genco Guralp, Derek Leben, Jeff Maynes, Bryan Miller, Rob Rynasiewicz, Shane Steinert-Threlkeld, Nicholas Tebben, John Waterman, and Karen Yan.

I'd also like to thank the administrative staff of the philosophy department for all of their help and hard work. Alicia Burley and Leslie Bean have made the department a great place to work, and I owe my thanks to Veronica Feldkircher-Reed for her help with all of the challenges involved in scheduling my defense.

Lastly, I'd like to thank my wife, Karen Yan. My time in graduate school would have been worth it even if I ended up leaving without my Ph.D., because that's when I met you. This was by far the best thing to happen to me. Moreover, there's a good chance that I would have left graduate school without my Ph.D. had I not met you. I thank you for your love and encouragement, and I want you to know that you've done more than anyone else to motivate me to finish this dissertation.

*For Dan Hricko and Lee Hricko*

*You don't actually have to read this.*

---

## *Contents*

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>xii</b>
<b>1 Extant Views of Hypothetical Entities</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Hypothetical Entities . . . . .	2
1.2.1 The Rough Guide . . . . .	3
1.2.2 Hypothetical, Theoretical, Unobservable . . . . .	9
1.2.3 Is It Minimal Enough? . . . . .	16
1.3 Scientific Realism and Hypothetical Entities . . . . .	19
1.3.1 The Three “Stances” and the No-Miracles Argument . . . . .	19
1.3.2 Realistic Interpretation . . . . .	21
1.3.3 Realism About Hypothetical Entities . . . . .	25
1.3.4 Some Points of Contact . . . . .	31

1.4	Constructive Empiricism and Hypothetical Entities . . . . .	37
1.4.1	Empirical Adequacy, Observability, and the Semantic View . . . . .	39
1.4.2	Literal Construal and Realistic Interpretation . . . . .	43
1.4.3	Constructive Empiricism About Hypothetical Entities . . . . .	48
1.5	Structural Realism and Hypothetical Entities . . . . .	51
1.5.1	Epistemic Structural Realism . . . . .	56
	Structure and Ramsey Sentences . . . . .	56
	The Semantics of ESR . . . . .	63
	ESR and Hypothetical Entities . . . . .	69
1.5.2	Ontic Structural Realism . . . . .	73
	Structure and the Semantic View . . . . .	73
	The Semantics of OSR . . . . .	80
	OSR and Hypothetical Entities . . . . .	89
1.6	Conclusion . . . . .	92
<b>2</b>	<b>Suppositionalism and Hypothetical Entities</b>	<b>94</b>
2.1	Introduction . . . . .	94
2.2	The Phenomena of Supposition . . . . .	95
2.2.1	Everyday Life . . . . .	97
2.2.2	Mathematics . . . . .	97
2.2.3	Natural Deduction . . . . .	98
2.2.4	Fiction . . . . .	100

---

2.3	Fiction: Imagination or Supposition? . . . . .	101
2.3.1	Imagination . . . . .	101
2.3.2	Imagination, <i>Not</i> Supposition . . . . .	102
2.3.3	‘Supposing*’ as a Technical Term . . . . .	104
2.3.4	The Nature of the Distinction . . . . .	105
	The Subset View . . . . .	106
	A Sharp Distinction . . . . .	108
	A Vague Distinction . . . . .	114
2.3.5	Imaginative Resistance . . . . .	116
2.4	Suppositionalism: A Unified Treatment . . . . .	119
2.4.1	Props, Interpretive Engines, and Suppositions . . . . .	120
2.4.2	Objects of Supposition . . . . .	123
2.4.3	Normal and Conformal Predication . . . . .	131
2.5	Hypothetical Entities in the Sciences . . . . .	136
2.5.1	Suppositionalism and Science . . . . .	136
2.5.2	Hypothetical Entities as Objects of Supposition . . . . .	140
2.5.3	The Appropriation Model . . . . .	143
2.5.4	The Role of Intentions . . . . .	155
<b>3</b>	<b>A History of a Hypothetical Entity:</b>	
	<b>The Case of the Muriatic Radical</b>	<b>162</b>
3.1	Introduction . . . . .	162

---

3.2	The Birth of the Muriatic Radical . . . . .	164
3.2.1	The Oxygen Theory of Acidity . . . . .	165
3.2.2	Two Undecomposed Acids . . . . .	168
3.2.3	The Phlogiston Theory of Acidity . . . . .	174
3.3	From Muriatic Acid to Hydrochloric Acid . . . . .	179
3.3.1	Davy on Oxygen, Phlogiston, and the Acids . . . . .	184
	The 1807 Bakerian Lecture . . . . .	184
	“Electro-Chemical Researches . . .” and the 1808 Bakerian Lecture .	187
	The 1809 Bakerian Lecture . . . . .	195
	“Researches on the oxymuriatic Acid . . .” and the 1810 Bakerian Lecture . . . . .	197
	Davy’s Unpublished Notes and Post-1810 Writings . . . . .	206
3.3.2	Three Puzzling Issues . . . . .	211
	Puzzle 1: Davy’s Relationship to the Phlogistic and Antiphlogistic Theories . . . . .	211
	Puzzle 2: Why Davy? Why Not Gay-Lussac and Thénard? . . . . .	216
	Puzzle 3: Is Chlorine an Element for Davy? . . . . .	222
	Two Wrongs <i>Can</i> Make a Right . . . . .	225
3.4	The Aftermath of Davy’s Work and the Death of the Muriatic Radical . . .	227
<b>4</b>	<b>An Objection to the Extant Views and a Defense of Suppositionalism</b>	<b>232</b>

---

4.1	Introduction . . . . .	232
4.2	The Extant Views and the Muriatic Radical . . . . .	234
4.2.1	The Scope of the Extant Views . . . . .	234
4.2.2	What the Extant Views are Committed To . . . . .	244
4.3	Actual and Counterfactual Histories of the Muriatic Radical . . . . .	246
4.3.1	Actual History: Failure of Reference . . . . .	247
4.3.2	Counterfactual Histories: Successful Reference . . . . .	254
4.4	The Problem with the Extant Views . . . . .	263
4.4.1	‘Muriatic Radical’ as a Successfully Referring Term . . . . .	265
4.4.2	‘Muriatic Radical’ as a Term that ‘Kind Of’ Refers . . . . .	276
4.4.3	Whig History . . . . .	279
4.4.4	‘Muriatic Radical’ as a Term that Fails to Refer . . . . .	283
4.4.5	The Other Extant Views Revisited . . . . .	286
4.5	Suppositionalism and the Muriatic Radical . . . . .	288
4.5.1	The Muriatic Radical as an Object of Supposition . . . . .	289
4.5.2	The Actual and Counterfactual Histories and the Appropriation Model	291
4.5.3	Suppositionalism, Whig History, and Incomplete History . . . . .	297
4.6	Conclusion . . . . .	301
	<b>Bibliography</b>	<b>302</b>
	<b>Curriculum Vita</b>	<b>326</b>

---

*List of Figures*

1.1	A diagram of the relationships among observable, unobservable, theoretical, and hypothetical entities. . . . .	14
2.1	Vindicated and discredited theories, and their relation to the world. . . . .	152
2.2	File $F_1$ : a two-card file. . . . .	158
2.3	File $F_2$ : a two-card file. . . . .	159
2.4	File $F_3$ : a three-card file. . . . .	159
3.1	Phlogisticated acids. . . . .	177
3.2	Dephlogisticated acids. . . . .	177
4.1	A diagram of the actual and counterfactual histories of the muriatic radical. . .	263
4.2	Vindicated and discredited theories, and their relation to the world. . . . .	293

## Chapter 1

---

### *Extant Views of Hypothetical Entities*

#### 1.1 INTRODUCTION

When scientists put forward hypotheses, they are usually loath to propose new kinds of entities that would inflate the ontologies of their theories. But sometimes their hypotheses do introduce such entities, and some of these entities are what I will call ‘hypothetical entities.’

My aim in this chapter is to discuss the extant views of hypothetical entities that are at least implicit in the main positions that occupy the position space of the scientific realism debate. To that end, I will begin by circumscribing the phenomenon of interest, namely, hypothetical entities in the sciences. If, as I claim, there are some extant views of hypothetical entities, there must be some phenomenon that all of these views are views *of*, even if it’s the case that they differ in what they have to say about this phenomenon. I will then go through a selective survey of the scientific realism debate, in which I will focus on three main positions in the philosophy of science, namely, scientific realism, constructive empiricism, and structural realism.<sup>1</sup> I will spell out the view of hypothetical entities that each position is committed to. Philosophers of science haven’t exactly framed their discussions

---

<sup>1</sup>Henceforth, ‘realism’ should be understood as ‘scientific realism.’ Whenever I have in mind another kind of realism, I will specify this.

in terms of the notion of ‘hypothetical entity’ that will be my focus here. In that case, it will be important to be clear about this notion and any points of contact with the literature, and it will take some work to extract the extant views of hypothetical entities from the above-mentioned positions.

The upshot of this selective survey is that one of the ways in which philosophers of science have attempted to distance themselves from logical positivism is by committing themselves to some variant of the realistic interpretation of the language of science. In short, the basic idea is that theoretical terms are to be understood as putatively referring expressions that have putative reference to empirical entities. *A fortiori* for theoretical terms used to introduce hypothetical entities into scientific discourse. In this chapter, I will argue that the commitment to realistic interpretation is widespread. In the remainder of the dissertation, I will argue that this widespread commitment is a mistake, at least when it comes to terms used to introduce hypothetical entities, and I will develop an alternative view of hypothetical entities to put in its place.

## 1.2 HYPOTHETICAL ENTITIES

Hypothetical entities are ubiquitous in science, and perhaps the best way to introduce them is to give a few examples. Towards the end of the eighteenth century, Antoine Lavoisier hypothesized the existence of a simple substance called ‘caloric,’ which he thought to be the matter of heat. In the mid-nineteenth century, John Adams and Urbain Le Verrier hypothesized the existence of an undiscovered planet. It would not be long before astronomers on the continent observed Neptune through their telescopes. As I write this

sentence, the status of the Higgs boson is hypothetical, but it is hoped that the Large Hadron Collider will change this status.<sup>2</sup> Caloric, Adams' and Le Verrier's undiscovered planet, and the Higgs boson are all examples of hypothetical entities. The examples could be multiplied, and I will discuss more examples later, but these should suffice for introducing the phenomenon of interest.

The term 'hypothetical entity' is not new to the philosophy of science. Philosophers mostly use the term as a synonym for 'theoretical entity,' as David Lewis does when he claims that "[t]heoretical entities might better be called (as they sometimes are called) *hypothetical* entities" (Lewis (1970), 428).<sup>3</sup> As the above-mentioned examples may already suggest, I mean something different by 'hypothetical entity'—something that makes it differ in meaning from 'theoretical entity.' In order to make this clear, I will first have to be a bit clearer about what I take a hypothetical entity to be.

### 1.2.1 THE ROUGH GUIDE

I don't propose to give necessary and jointly sufficient conditions for what makes an entity hypothetical, but I do believe that the following rough guide captures the basic idea.

A hypothetical entity is a new (kind of) purported entity that a scientist puts forward as a (kind of) purported empirical entity in advance of decisive empirical reasons to do so.<sup>4</sup>

---

<sup>2</sup>And, indeed, it has.

<sup>3</sup>Emphasis is the author's unless otherwise noted.

<sup>4</sup>This rough guide is roughly in agreement with what Rynasiewicz, Steinert-Threlkeld, and Suri (2010) mean by 'hypothetical entity.'

Terms introduced to name hypothetical entities are what I will call *HE terms*, short for *hypothetical entity terms*.

The rough guide that I've put forward can be used to circumscribe a phenomenon of interest, namely, hypothetical entities in the sciences. But in order for this to be the case, the rough guide must be at least somewhat “theory-neutral,” in the sense that it is minimal enough to be agreed upon by philosophers involved in the realism debate who may disagree about much else.

As we'll see, there's a sense in which the rough guide cannot be accepted as-is by all parties. A philosopher may have to understand it in a specific way so as to make it square with various other beliefs and commitments. There will therefore be a number of different ways of understanding the rough guide. But this is the sense in which the guide is rough—it offers a starting point to circumscribe the phenomenon of interest. The different ways of understanding are the beginnings of offering a specific view of hypothetical entities, but there needn't be any incommensurability among these views if the rough guide can be used to make a specific phenomenon salient. That said, the rough guide is still in need of some clarification, and I will provide some of that right now.

First of all, I use the word ‘entity’ in a broad sense, to cover things, individuals, objects, substances, events, processes, properties, and relations.<sup>5</sup> In this case, the rough guide can be used by philosophers who do not admit one or more of these into their ontologies. My examples of hypothetical entities are perhaps most easily classified as (kinds of) things, individuals, objects, or substances, and this will be my primary focus. But the rough guide

---

<sup>5</sup>For an example of this broad sense of ‘entity,’ see van Fraassen (1980), 15.

itself is more comprehensive.

Secondly, a hypothetical entity is a *purported* entity, since, after all, it's hypothetical. But not all purported entities are hypothetical entities. For example, Jack the Ripper is a purported entity. He is an *entity*, since he is an individual person, and he is a *purported* entity because it's possible that more than one person committed the murders in question. But he is not a hypothetical entity, in the sense that I'm concerned with, because it wasn't scientists who put him forward as a purported empirical entity.

Thirdly, when a scientist "puts forward" a hypothetical entity, she does so in print or in speech. For example, when Lavoisier puts forward caloric as a hypothetical entity in his *Elements of Chemistry*, he is concerned to explain various changes in state that occur at different degrees of heat. He writes:

It is difficult to comprehend these phenomena, without admitting them as the effects of a real and material substance, or a very subtile fluid, which, insinuating itself between the particles of bodies, separates them from each other.

(Lavoisier (1802), 52)

He "allow[s] that the existence of this fluid may be hypothetical," and claims that, "strictly speaking, we are not obliged to suppose this to be a real substance" (Lavoisier (1802), 52, 53). Nonetheless, he his collaborators "have distinguished the cause of heat, or that exquisitely elastic fluid which produces it, by the term *caloric*" (Lavoisier (1802), 53). This example illustrates some more general points about the act of putting forward a hypothetical entity. Scientists usually do this in order to explain some set of phenomena. And Lavoisier's

rather tentative attitude towards caloric shows that putting forward a hypothetical entity does not entail that the scientist believes that that entity exists, though she may certainly believe this. The rough guide therefore emphasizes the act of putting forward, and not the beliefs of the scientist.

Fourthly, I enclose “kind of” in parentheses because there are times when scientists hypothesize kinds (e.g., the electron) and times when they don’t (e.g., Adams and Le Verrier’s planet). I use the term ‘kind’ so as not to commit to the existence of natural kinds—the kinds referenced in the rough guide may be natural kinds, but they needn’t be.

Fifthly, by ‘empirical entity’ I have in mind an entity that exists in the natural world. This rules out abstract entities like numbers and sets, idealizations like ideal gases and frictionless planes, and useful fictions like quantum orbitals. The hope is that it should be possible to confirm that the entity exists in the natural world, or to confirm that it does not, and to do so on grounds that are at least partly empirical. If that is, in principle, impossible, then the hypothetical entity is a *hopeless* hypothetical.<sup>6</sup>

Sixthly, ‘decisive empirical reasons’ should be understood with reference to the consensus of the scientific community—the consensus need not be unanimous, but the reasons in question should be capable of eventually convincing the vast majority of the scientific community.<sup>7</sup> Scientists put forward hypothetical entities for empirical reasons, insofar as they use such entities to explain some (possibly anomalous) phenomenon or phenomena. But an explanation that is crafted to accommodate a known set of phenomena is not a decisive

---

<sup>6</sup>For this terminology, see Rynasiewicz et al. (2010), 11.

<sup>7</sup>To be clear, I take it that reasons in general needn’t be capable of eventually convincing the vast majority of a community, but that decisive reasons must be so capable.

empirical reason that shows that a new (kind of) empirical entity exists. What is needed is some independent support. An explanation that employs a hypothetical entity would have this support if it, for example, ends up explaining or predicting something besides the set of phenomena that the explanation was crafted to accommodate in the first place.<sup>8</sup> The idea is that such an explanation must have further consequences, apart from those consequences that explain the phenomena in question, and that at least some of these consequences are in-principle testable. In particular, it must be possible to find out more about the hypothetical entity in question. For example, when the electron was a hypothetical entity, it was thought to be a charged particle. Physicists working in the late-nineteenth century concluded that it must therefore have other properties, like a charge-to-mass ratio, which they then went on to measure in a number of ingenious ways. It was this work that led physicists to conclude that the electron is not merely hypothetical, but a well-established empirical entity.

Seventhly, here and throughout, I use the word ‘name’ in a slightly technical sense that doesn’t entail that reference to an empirical entity is successful. This is important since some hypothetical entities, like caloric, are no longer thought to exist.

Eighthly, and finally, it should be emphasized that some hypotheses are confirmed, while others are rejected. Such hypotheses are, at the time of confirmation or rejection, no longer ‘live,’ in the sense that scientists no longer have any reason to entertain them. In that case, hypothetical entities have a ‘shelf life.’ It doesn’t follow that Neptune is a hypothetical entity today just because Adams and Le Verrier hypothesized the existence of a new planet. And it doesn’t follow that caloric is a hypothetical entity today just because

---

<sup>8</sup>For some elaboration on this point, see Chalmers (2009), 8–10.

Lavoisier hypothesized the matter of heat. This is because no scientists are entertaining these hypotheses anymore. Neptune is best thought of as an empirical entity, while caloric is what we might call a *dead* hypothetical.<sup>9</sup> We can then leave it up to individual philosophers to decide whether to construe this to mean that ‘caloric’ fails to refer, or refers to a non-existent entity, abstract object, or something else. Likewise, the terms ‘Neptune’ and ‘caloric’ are not HE terms today, since neither Neptune nor caloric are hypothetical entities nowadays.<sup>10</sup>

Philosophers have generally not distinguished a class of hypothetical entities in the sense I’ve just sketched.<sup>11</sup> Distinguishing such a class is important insofar as hypothetical entities have been and continue to be pervasive in the sciences. They have even drawn the interest of those who are not scientists or philosophers of science. In physics, the Higgs boson and the five-sigma standard for the discovery of a particle have recently found their way into newspapers. Likewise, the discoveries of evolutionary biology’s various missing links still make the news today. Students of chemistry could see the blank spots in the periodic table for years before these spots were filled with elements that are no longer hypothetical. Particular examples of hypothetical entities may even be of more direct interest to philosophers working in subfields other than the philosophy of science.

---

<sup>9</sup>I propose to use this terminology in such a way that a dead hypothetical is a hypothetical entity in the same sense in which a counterfeit dollar is a dollar.

<sup>10</sup>This eighth point is somewhat contentious. While my intuition is that caloric is no longer a hypothetical entity, Robert Rynasiewicz has pointed out to me that this intuition may not be widely shared. I’ve indicated the reasons for my view, and I do think that the rough guide should be understood along these lines. But there is certainly room for disagreement here. Since nothing in chapter 1 hinges on this point, readers who do not share my intuition are invited to drop it from the rough guide. But I will make use of this eighth point when I develop the alternative view in chapter 2, and defend that view in chapter 4.

<sup>11</sup>One exception is Ian Hacking, who distinguishes between real and hypothetical entities (Hacking (1982), 76).

To take one example, psychologists, cognitive scientists, and neuroscientists are now in the process of studying consciousness, and so the qualia of philosophers of mind are now paradigm cases of hypothetical entities. The rough guide collects these and other examples under one head, as a phenomenon worth studying.

Hypothetical entities and HE terms also raise specific diachronic issues that haven't been addressed explicitly in the realism debate. How are we to understand the relationship between the electron *qua* hypothetical entity and the electron *qua* empirical entity? Are they the same entity? Is there any way to understand them as distinct entities? How are we to understand the relationship between the HE term 'electron' and the theoretical term that we use today? As different tokens of the same type? Or different types? There is certainly much that goes on in the realism debate that bears on these questions, but on the whole, they haven't been framed explicitly in this way.

It wouldn't be inappropriate for a reader to ask for more clarification of this rough guide, and I propose to provide some by pursuing two tasks. First of all, I will situate this rough guide within some of the literature in the philosophy of science concerning the realism debate. Secondly, I will argue that this rough guide is minimal enough to be agreed upon by philosophers who disagree about much else, and hence can be used to circumscribe a phenomenon about which philosophers can develop opposing views.

### 1.2.2 HYPOTHETICAL, THEORETICAL, UNOBSERVABLE

Philosophers have, for the most part, framed the realism debate in terms of a distinction between theoretical and non-theoretical terms, and a distinction between observable and

unobservable entities. There are a number of ways in which one could characterize this debate, along with a corresponding number of issues that are at stake, and I won't touch on all of these here. For the time being, I will take it that the central question behind the debate is whether we can make sense of science, and of scientific activity, without believing in unobservable entities. A closely related question, which is important for my purposes here, is whether we can make sense of science, and of scientific activity, without believing that terms refer to unobservable entities.<sup>12</sup> We will see later how proponents of various positions answer these questions—logical positivists and constructive empiricists say “yes,” but for different reasons; realists say “no”; and structural realists fall somewhere in the middle.

The observable-unobservable distinction is meant to classify *entities*. Bas van Fraassen offers the following “rough guide to the avoidance of fallacies” when it comes to classifying entities as observable or unobservable: “X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it” (van Fraassen (1980), 16). For van Fraassen, observation is the same as perception, and it is something that we can do without the aid of instruments (van Fraassen (2008), 93). He admits that the distinction between observable and unobservable entities is vague, but claims it to be employable insofar as there are clear cases of observable entities (like the moons of Jupiter) and clear counter-cases (like electrons) (van Fraassen (1980), 16–17). The moons are observable because if an astronaut were close enough, she would be able to see them with the naked eye. This is the case, according to van Fraassen, even though the

---

<sup>12</sup>Cf. Bas van Fraassen, who argues for his constructive empiricism on the grounds that “it makes better sense of science, and of scientific activity, than realism does” (van Fraassen (1980), 73).

moons have not yet been observed, but only detected by means of telescopes and satellites. Electrons, for van Fraassen, can never be observed, since it's impossible to see them with the naked eye.

Van Fraassen's views on the observable-unobservable distinction have been influential, but also controversial. Philosophers have taken issue with some of van Fraassen's views, like the consequences that a look through a telescope does not count as an observation of the moons of Jupiter, and that a look through a microscope does not count as an observation of a one-celled organism.<sup>13</sup> Moreover, physicists, along with some philosophers, do sometimes claim that we observe electrons. Their use of 'observe' is therefore much closer to what van Fraassen means by 'detect' (van Fraassen (1980), 17). To be sure, there is disagreement over exactly how to draw the observable-unobservable distinction. But when it comes to the philosophers engaged in the realism debate, there is much agreement over what entities do and do not count as observable. At the very least, van Fraassen's critics have accepted much of his classification of entities for the sake of argument, if only to attack his other views.

The theoretical-non-theoretical distinction, on the other hand, is meant to classify concepts and bits of language—here, I will focus on *terms* since these are what name entities. A theoretical term is just a term “introduced or adapted for the purposes of theory construction” (van Fraassen (1980), 14), or in short, a “scientific term” (Chakravartty (2011), §1.1). So 'electron' would count as a theoretical term, while 'basketball' would not, since only the former, and not the latter, figures in scientific theories.

---

<sup>13</sup>Regarding the latter point, see Hacking (1985).

The observable-unobservable distinction and the theoretical-non-theoretical distinction do not line up together in a neat way. It is true that there are theoretical terms that name unobservable entities—e.g., ‘electron’ names electrons. And there are non-theoretical terms that name observable entities—e.g., ‘basketball’ names basketballs. But, as Lewis has clearly pointed out, theoretical terms also name observable entities—e.g., ‘H<sub>2</sub>O’ names water. And non-theoretical terms also name unobservable entities—e.g., ‘living creature too small to see’ may name a bacterium (Lewis (1970), 428).<sup>14</sup> In this case, the two distinctions are largely orthogonal to one another.

Philosophers involved in the realism debate have also used the term ‘theoretical entity’ in addition to the above distinctions.<sup>15</sup> Van Fraassen holds that the term amounts to a category mistake, since “[t]erms or concepts are theoretical” while “entities are observable or unobservable” (van Fraassen (1980), 14). Strictly speaking, van Fraassen is correct. Perhaps some of the confusion is due to the fact that, as Anjan Chakravartty has pointed out, “‘theoretical term’, prior to the 1980s, was standardly used to denote terms for unobservables” (Chakravartty (2011), §1.1). In that case, ‘theoretical entity’ and ‘unobservable entity’ may have been viewed as synonymous by many philosophers of science. But we’ve already seen that, based on what Chakravartty has identified as “now the more common usage,” theoretical terms are just scientific terms (Chakravartty (2011), §1.1). If ‘theoretical entity’ is understood as shorthand for ‘entity named by a theoretical term,’ there is no category mistake, and there is no danger of conflating unobservable entities with theoret-

---

<sup>14</sup>For similar points, see Putnam (1975c) and van Fraassen (1980), 54–55.

<sup>15</sup>See, for example, Maxwell (1962) and Lewis (1970).

ical entities, which may or may not be unobservable.<sup>16</sup> This is how I propose to use the terminology.

At this point, we can say where hypothetical entities fit in. Some hypothetical entities are unobservable entities. For example, the electron was, at least at one point, a hypothetical entity, and that hypothetical entity is unobservable. But some hypothetical entities are observable. For example, the planet that Adams and Le Verrier hypothesized is something that, if it were present to us, we could observe it. Some hypothetical entities turn out not to exist, and it may seem counterintuitive to classify non-existent entities as observable or unobservable.<sup>17</sup> Here, I follow van Fraassen in classifying non-existent entities in terms of the observable-unobservable distinction. Van Fraassen claims that “[a] flying horse is observable—that is why we are so sure that there aren’t any—and the number seventeen is not” (van Fraassen (1980), 15). His disbelief in the reality of flying horses is evident in the quotation, and his disbelief in the reality of the number seventeen follows from his nominalism.<sup>18</sup> I claim that Vulcan and caloric can be classified as observable and unobservable, respectively, in the same sense that a flying horse and the number seventeen can. While observability may be an important thing to keep in mind when studying hypothetical entities, it’s important to realize that the issues that hypothetical entities raise are not just the same old issues raised by the observable-unobservable distinction.

---

<sup>16</sup>For recognition of the latter point, see Lewis (1970), 428 and Ladyman and Ross (2007), 253.

<sup>17</sup>Talk of non-existent entities is not meant to imply any specific metaphysical commitments to a being-existence distinction of the kind defended by Alexius Meinong, and later by Terence Parsons (see Parsons (1980)). I take it that those who find talk of non-existent entities like the round square to be contentious can cash that talk out in terms of some less contentious paraphrase, for example, in the way that Bertrand Russell does (see Russell (1905)). I use the language of non-existent entities only for ease of exposition. See the next subsection for more on this issue.

<sup>18</sup>See, for example, Monton and van Fraassen (2003), 412 for a statement of van Fraassen’s “disbelief in the reality of abstract entities.” See also van Fraassen (1975).

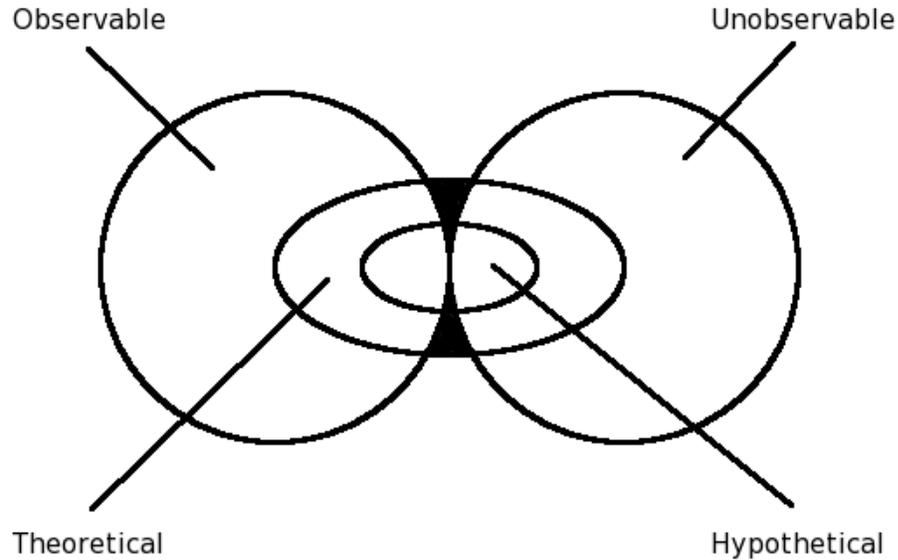


Figure 1.1: A diagram of the relationships among observable, unobservable, theoretical, and hypothetical entities.

We can also see where HE terms fit in. Although hypothetical entities can be named by non-theoretical terms, HE terms, insofar as they are terms introduced to name hypothetical entities, count as theoretical terms. Importantly, then, what philosophers of science have to say about theoretical terms applies *ipso facto* to HE terms. And since HE terms are theoretical terms, hypothetical entities count as theoretical entities. In that case, what philosophers of science have to say about theoretical entities applies *ipso facto* to hypothetical entities, provided that they take it that a theoretical entity is any entity named by a theoretical term. Figure 1.1 provides a diagram of the relationships among observable, unobservable, theoretical, and hypothetical entities.<sup>19</sup> For ease of representation, the figure assumes a sharp distinction between the observable and the unobservable.

Before moving on, it's worth noting an objection to what I've said about the relation-

<sup>19</sup>Thanks to Marianna Bergamaschi Ganapini for suggesting this useful diagram.

ships among these different kinds of entities and terms. One might object to my claims that hypothetical entities are a subset of theoretical entities, and that HE terms are a subset of theoretical terms. After all, the man in the moon and the Loch Ness monster would seem to be hypothetical entities, and the terms ‘man in the moon’ and ‘Loch Ness monster’ would seem to be HE terms. But these terms do not appear in any scientific theory, in which case they are not theoretical terms. And for this reason, these entities are not theoretical entities.<sup>20</sup>

In reply, I wish to emphasize that, according to the terminology that I’ve adopted here, the ‘hypothetical’ in ‘hypothetical entity’ refers back to hypotheses that scientists make in the course of engaging in scientific activity. There is, no doubt, a sense in which individuals who are not scientists also put forward various hypotheses, and these hypotheses may name the man in the moon or the Loch Ness monster. But my use of terminology excludes these from being hypothetical entities, and excludes the corresponding terms from being HE terms. I justify this on the basis that I am concerned with scientific activity here, as opposed to the hypotheses of non-scientists.

Here, then, is how hypothetical entities, in the sense given by the rough guide, are connected to the literature in the philosophy of science at the most general level. We can use the rough guide to identify some hypothetical entities. We look for the terms that scientists use to name those entities. Those terms (HE terms) are a subset of the theoretical terms. So what philosophers of science have to say about theoretical terms tells us something about what they are committed to regarding HE terms. This, in turn, tells us something about

---

<sup>20</sup>I thank Rynasiewicz for suggesting this objection.

what they are committed to regarding the hypothetical entities named by HE terms.

### 1.2.3 IS IT MINIMAL ENOUGH?

Here I'll briefly consider two objections to my claim that this rough guide is sufficiently minimal, in the sense that it can be agreed upon by philosophers involved in the realism debate who may disagree about much else. Even if I successfully address these objections, it won't establish that the rough guide is sufficiently minimal. But my replies should make it the case that this is at least *prima facie* plausible.

**Objection 1:** The rough guide may commit one to a kind of realism about hypothetical entities. To take one example, caloric was, at one point, a hypothetical entity, and by our best lights, caloric does not exist. But I've been writing about caloric as if it were an entity—indeed, I've labeled it as such. In this case, perhaps the rough guide commits us to belief in non-existent entities, abstract entities, possibilia, or a host of other metaphysically contentious entities. Any philosopher with reservations about these kinds of entities, then, would seem to be unable to accept the rough guide. To take one example, van Fraassen's nominalism would preclude him from belief in abstract entities and possibilia, and might therefore also prevent him from accepting the rough guide. He "do[es] not feel a need to believe in the existence of abstract entities" (van Fraassen (1975), 39). And he claims that, "[f]rom an empiricist point of view, there are besides relations among actual matters of fact, only relations among words and ideas. Yet causal and modal locutions appear to introduce relations among possibilities, relations of the actual to the possible" (van Fraassen (1989), 213). For him, then, empiricism implies modal nominalism, which rules out the existence

of possibilia.<sup>21</sup>

**Reply to Objection 1:** The rough guide may seem to commit us to belief in non-existent entities, abstract entities, or possibilia. But philosophers who hold that such entities are metaphysically contentious are in the business of arguing that we needn't believe in such entities, even if the way we talk might, at first blush, seem to commit us to believing in such entities. Let's suppose that philosopher *A* accepts the rough guide, but does not believe in non-existent entities. Philosopher *B*, however, claims that the rough guide commits *A* to a belief in non-existent entities, since caloric was a hypothetical entity, and caloric does not exist. In response, *A* can argue that, though we sometimes talk about caloric as an entity that does not exist, this does not commit a speaker to the belief that there is a non-existent entity called 'caloric.' To be sure, some, like *B*, may interpret such talk as committing the speaker to a belief in non-existent entities. Others, like *A*, will interpret such talk as involving terms and/or descriptions that fail to pick out anything at all. Such interpretations may involve paraphrasing what the speaker says in such a way that the appearance of a commitment to non-existent entities disappears, for example, in the way Russell does in response to Meinong (Russell (1905)). In order for the objection to go through, it would have to be the case that *B*'s interpretation is the only correct one, and this has not been shown to be the case. More generally, philosophers who hold that certain kinds of entities are metaphysically contentious will have something to say about why talk of such entities doesn't commit one to believe in them. In that case, the rough guide can be accepted by parties who may disagree about what entities are and are not metaphysically contentious.

---

<sup>21</sup>But see Ladyman (2000) for some reasons to be confused about what, exactly, van Fraassen's views on modality amount to.

**Objection 2:** Although I've stated that the rough guide needn't commit one to the existence of natural kinds, one may argue that it, in fact, does. I employ the language of *kinds* of empirical entities. I've stated that an empirical entity is an entity that exists in the natural world. For there to be kinds of empirical entities, it would seem that the world must have some kind of objective, mind-independent kind structure. It's not clear what I could mean by claiming that this kind structure is not a *natural* kind structure. And in this case, any philosopher who rejects natural kinds would have to reject the rough guide. So critics of natural kinds like Ian Hacking would have to reject the rough guide.<sup>22</sup>

**Reply to Objection 2:** In order to answer this objection, it is important to distinguish two things. On the one hand, there is the world, which may or may not have any mind-independent natural kind structure. On the other hand, there are our theories, which categorize the entities they talk about into various kinds. If one believes that the world is structured in terms of natural kinds, then our best guess as to what that structure is comes from the kind structure of our best theories. But one needn't believe that the world has such a structure, even if one believes that our theories do categorize entities into such kinds. All that is needed for acceptance of the rough guide is this latter belief.

At this point, we can turn to the specific positions that occupy the position space in the realism debate, and extract the extant views of hypothetical entities from these positions.

---

<sup>22</sup>See, for example, Hacking (2007).

### 1.3 SCIENTIFIC REALISM AND HYPOTHETICAL ENTITIES

The fall of logical positivism went hand-in-hand with the rise of scientific realism. The latter set the stage for the current incarnation of the scientific realism debate, and so the first position I will discuss is realism. I will begin with a brief examination of what realism is, and how realists defend it. I will then turn to the realists' semantic commitments, before extracting their view of hypothetical entities. Finally, I will discuss some examples of specific realist positions that instantiate this view of hypothetical entities.

#### 1.3.1 THE THREE "STANCES" AND THE NO-MIRACLES ARGUMENT

Realists typically characterize their position in terms of three kinds of commitments: metaphysical, semantic, and epistemic. Chakravartty puts this in terms of taxonomies which group these commitments in terms of "three lines of inquiry" (Chakravartty (2007), 9–10) and three "dimensions of realist commitment" (Chakravartty (2011), §1.2). In a similar vein, Stathis Psillos puts this in terms of three kinds of "stances" that realists are committed to:

- 1 The metaphysical stance asserts that the world has a definite and mind-independent natural-kind structure.
- 2 The semantic stance takes scientific theories at face value, seeing them as truth-conditioned descriptions of their intended domain, both observable and unobservable. Hence, they are capable of being true or false. Theoretical assertions are not reducible to claims about the behaviour of

observables, nor are they merely instrumental devices for establishing connections between observables. The theoretical terms featuring in theories have putative factual reference. So, if scientific theories are true, the unobservable entities they posit populate the world.

- 3 The epistemic stance regards mature and predictively successful scientific theories as well-confirmed and approximately true of the world. So, the entities posited by them, or, at any rate, entities very similar to those posited, do inhabit the world. (Psillos (1999), xix)

The strongest argument in favor of realism is the so-called no-miracles argument (also known as the ultimate argument and the miracle argument). The basic idea goes back to Hilary Putnam's claim that "[t]he positive argument for realism is that it is the only philosophy that does not make the success of science a miracle" (Putnam (1975b), 73). The literature on the no-miracles argument is voluminous, and there are a number of formulations that differ subtly from one another.<sup>23</sup> In general, the form of the argument is inference to the best explanation. The explanandum is the explanatory and predictive success of our best scientific theories, and oftentimes this success is explicitly restricted to novel predictive success. The explanans can be put in terms of the three stances adumbrated above. If our best theories are literally (approximately) true descriptions that refer successfully to a mind-independent world, then it should be no surprise that they are as successful as they are. This explanation of the success of science is often contrasted with alternative,

---

<sup>23</sup>To take just a handful of examples, see Putnam (1978), 18–22; Brown (1982); Brown (1985); Musgrave (1988); Musgrave (2006); and Psillos (1999), 78–97.

antirealist explanations, which reject the realist claim that our best theories are literally (approximately) true. Realists argue that, on such antirealist accounts, the success of science is a miracle. And since we should prefer a non-miraculous explanation to a miraculous one, we should prefer the realist explanation as the best. We then infer that realism is true, since it offers the best explanation of the success of science.

### 1.3.2 REALISTIC INTERPRETATION

Although realism depends on metaphysical and epistemic commitments, its semantic commitment is most important for extracting the realist view of hypothetical entities, and so I will devote the most attention to this commitment.

We've already seen this commitment formulated in Psillos' semantic stance. Psillos claims that theories "are capable of being true or false" and that "[t]heoretical terms featuring in theories have putative factual reference" (Psillos (1999), xix). Here he follows Richard Boyd, who puts the commitment in similar terms:

"Theoretical terms" in scientific theories (i.e., non-observational terms) should be thought of as putatively referring expressions; scientific theories should be interpreted "realistically". (Boyd (1983), 45)

Likewise, Chakravartty expresses the same commitment as follows:

A realist semantics implies that theoretical claims about this reality have truth values, and should be construed literally, whether true or false. (Chakravartty (2007), 9)

Semantically, realism is committed to a literal interpretation of scientific claims about the world. In common parlance, realists take theoretical statements at “face value”. According to realism, claims about scientific entities, processes, properties, and relations, whether they be observable or unobservable, should be construed literally as having truth values, whether true or false.

(Chakravartty (2011), §1.2)

As is evident from the quotations above, this commitment goes by a number of different names. We see Psillos’ “semantic stance,” and Chakravartty’s “literal construal” and “literal interpretation.”<sup>24</sup> The term “semantic realism” has also been employed to name the same commitment.<sup>25</sup> While all of these names are apt, for the sake of regimenting vocabulary, I will follow Boyd and use “realistic interpretation.”

The central idea behind realistic interpretation, then, is twofold:

(RI<sub>1</sub>) Theoretical terms are putatively referring expressions that have putative reference to empirical entities.

(RI<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable of being true or false.<sup>26</sup>

The restriction to empirical entities in (RI<sub>1</sub>) is meant to mirror Psillos’ language of “putative factual reference” (Psillos (1999), xix) and Chakravartty’s emphasis on reference to

---

<sup>24</sup>Chakravartty may be borrowing this language from van Fraassen. See van Fraassen (1980), 10, but see my discussion below for some differences between van Fraassen and Chakravartty.

<sup>25</sup>See, for example, Frost-Arnold (2011).

<sup>26</sup>Chakravartty actually holds a view stronger than this, namely, that such assertions are either true or false (Chakravartty (2007), 9; Chakravartty (2011), §1.2). I will devote some discussion to this below, where I discuss van Fraassen’s views on literal construal.

“scientific entities” (Chakravartty (2011), §1.2). Similarly, talk of reality in (RI<sub>2</sub>) is meant to apply to the natural world, as opposed to any world of abstract objects or possibilia.

With regard to (RI<sub>1</sub>), it’s important to emphasize that, though this is a claim about the reference of terms, as Howard Sankey notes, “no commitment to a specific account of reference is required by the realistic interpretation of theoretical discourse” (Sankey (2008), 15, fn. 4). To be sure, realists may be required to commit themselves to one account or another, but there is no one account that all of them must adopt.

With regard to (RI<sub>2</sub>), proponents of realistic interpretation may have some freedom when it comes to theories of truth. Chakravartty notes that some realists advocate a correspondence theory, while others advocate a deflationary theory (Chakravartty (2011), §1.2). But it’s at the very least contentious whether a deflationary theory of truth is consistent with a truth-conditional semantics of the kind entailed by (RI<sub>2</sub>). Claire Horisk (2007) argues that the two cannot be combined, which would tell against the possibility of combining realism with deflationism about truth. But Steven Gross (2011) argues that there is no tension between the two, in which case realists may adopt deflationism. For now, I will assume that realism is consistent with a number of different theories of reference and theories of truth. But in chapter 4, I will argue that realists do have some more specific semantic commitments that they aren’t always explicit about. In any case, nothing I will have to say will depend on the possibility of a combination of realism and deflationism.

Realists are quick to point out that the commitment to realistic interpretation is what distances realism from various forms of instrumentalism and positivism. Chakravartty explicitly contrasts realism with instrumentalism immediately after his discussion of realistic

interpretation (Chakravartty (2011), §1.2). Psillos has the same concern in mind when he claims that “[t]heoretical assertions are not reducible to claims about the behaviour of observables, nor are they merely instrumental devices for establishing connections between observables” (Psillos (1999), xix). Psillos’ terminology of “truth-conditioned descriptions” is also presumably meant to distance realism from instrumentalism and positivism (Psillos (1999), xix), and I follow Psillos by using the terminology of “truth-conditioned assertions” in (RI<sub>2</sub>). One might object to this terminology, on the grounds that it’s questionable whether instrumentalists and/or positivists ever held that scientific theories involve descriptions or assertions that are not truth-conditioned. Indeed, it’s more natural to construe them as believing that scientific theories do not involve descriptions or assertions when it comes to what we cannot observe. However, I take it that what Psillos means is that scientific theories, for the realist, are capable of being true or false, and that what makes them true or false is whether their truth conditions obtain in the natural world. This is all that I intend by using the language of truth-conditioned descriptions and assertions, and this is something that certain sorts of antirealists would deny.

Recent work in the history of twentieth century philosophy has emphasized that logical positivism was far from a monolithic enterprise, and has suggested that so-called instrumentalists like Henri Poincaré and Pierre Duhem are actually best read as structural realists.<sup>27</sup> Nonetheless, the realists do have opponents. For example, Philipp Frank, one of the members of the Vienna Circle, explicitly rejects realistic interpretation in the following passage:

---

<sup>27</sup>For a good introduction to recent work on logical positivism, see Uebel (2011) and Creath (2011). John Worrall is the first philosopher to suggest that Poincaré and Duhem are structural realists. See Worrall (1989).

[T]he concept of true existence, e.g., of the action quantum  $h$ , is first defined by the agreements in the whole group of experiences involving  $h$ . ... [T]he concept of a “really existing action quantum” is only an abbreviation for the whole group of experiences with the symbol system belonging to it. It is completely false to say—as is often said—that the agreements of the values of  $h$  are explained most naturally by means of the *hypothesis of the real existence* of an action quantum. (Frank (1941), 85)

Here we see that, according to Frank,  $h$  is not a putatively referring term, but an “abbreviation,” which is at odds with realistic interpretation. Moreover, we see an explicit rejection of what might be a form of the no-miracles argument. A realist might claim that the agreements of the values of  $h$  would be a miracle if there were no “really existing action quantum,” but Frank claims that this is false. It’s plausible, then, that realistic interpretation is one of the key ideas that sets realism apart from at least some forms of instrumentalist positivism.

### 1.3.3 REALISM ABOUT HYPOTHETICAL ENTITIES

Since HE terms are a subset of theoretical terms, one might suspect that a straightforward application of (RI<sub>1</sub>) to the set of HE terms would yield the realist view of hypothetical entities. I take it that this suspicion is on point, and I will call this view ‘realism about hypothetical entities,’ or ‘R<sub>HE</sub>’ for short.

R<sub>HE</sub> HE terms are putatively referring expressions that have putative reference to empirical entities.

What, then, of hypothetical entities? If an HE term actually, as opposed to just putatively, refers, then it refers to an empirical entity. In this case, the empirical entity that it refers to is a hypothetical entity. If an HE term fails to refer, then there is no empirical entity that it refers to, and hence there is no hypothetical entity that it refers to. Once a hypothesis is confirmed, then, scientists refer to the same empirical entity that the HE term referred to. And once a hypothesis is rejected, talk of the hypothetical entity in question must be understood as naming a *purported* entity, which is to say that the HE term failed altogether to refer.

While I take it that this is, in fact, the realist view of hypothetical entities, it will first be necessary to note a caveat about (RI<sub>1</sub>). This has to do with terms like ‘Hamiltonian,’ ‘Lagrangian,’ and ‘wave function.’ According to the account of theoretical terms that I have adopted, such terms are, in fact, theoretical terms, since they are used in the construction of scientific theories. But it would be unfair to saddle the realist with the claim that such terms have putative reference to empirical entities. The caveat, then, is that realists needn’t apply (RI<sub>1</sub>) to all theoretical terms. But in this case, the question is whether (RI<sub>1</sub>) applies to HE terms. The answer is that realists do, in fact, treat HE terms in line with (RI<sub>1</sub>), and some examples should make clear the realist’s commitment to R<sub>HE</sub>.

To begin with, realists do admit cases in which HE terms referred to empirical entities. David Papineau, who has defended realism, exemplifies this attitude.<sup>28</sup> He finds the causal theory of reference problematic because it “seems unable to account for terms, like ‘positron,’ ‘neutrino,’ and ‘quark,’ that are explicitly introduced to refer to hypothetical

---

<sup>28</sup>For his defense of realism, see Papineau (2010).

entities which are conjectured to play certain theoretically specified roles, before any direct experimental manifestation of these entities is available for any dubbing ceremony” (Papineau (1996), 4). On the causal theory that Papineau is considering, reference to such entities can be fixed only in a dubbing ceremony. In these cases, since positrons, neutrinos, and quarks are unobservable, the dubbing ceremony would amount to pointing to some phenomenon caused by the entity in question, and thereby dubbing that entity with a name. The important point here is that, according to Papineau, HE terms like ‘positron,’ ‘neutrino,’ and ‘quark’ did, in fact, refer to empirical entities before there was any decisive reason to believe in the existence of these entities. As I mentioned above, a commitment to realistic interpretation is independent of the details of a specific theory of reference, and other realists might therefore reply to Papineau by showing that the causal theory can be modified so as to account for these examples. But they would not reply by arguing that the terms in question failed to refer.

Realists also exemplify this attitude when responding to alleged counterexamples to the no-miracles argument. Larry Laudan, in arguing against the realist’s no-miracles argument, attempts to undermine the project of explaining the success of theories by appeal to the successful reference of their terms. He points to a number of theories that, according to his realist opponents, referred, but which were nonetheless quite unsuccessful. Among his examples are the chemical atomic theories of the eighteenth century (Laudan (1981), 24). This would not present a counterexample to the realist project unless realists do, in fact, hold that ‘atom’ referred as early as the eighteenth century. Realists have not responded to Laudan’s counterexample by denying that ‘atom’ referred in the eighteenth

century. Instead, they claim that successful reference is not sufficient to explain a theory's success (Brown (1985), 54–55; Musgrave (1988), 236–237).<sup>29</sup><sup>30</sup> The atom is, of course, a paradigm case of a hypothetical entity. Insofar as realists hold that 'atom' referred in the eighteenth century, they admit the existence of a referential relation before there was decisive reason to believe that the entity in question exists. I take it that the earliest one could reasonably date such a decisive reason would be Jean Perrin's work at the beginning of the twentieth century, since it wasn't until after that work that there was a strong consensus.<sup>31</sup> Clyde L. Hardin and Alex Rosenberg have claimed that, in at least some cases, successful reference is possible even before decisive reasons are available. This is because "reference is severed from detailed beliefs of the theorist, and its success is accorded retrospectively in the light of subsequent further approximations to the truth" (Hardin and Rosenberg (1982), 608). The important point is that, if the reference of 'atom' could be fixed in the eighteenth century, it should be the case that other referential relations were fixed before there was decisive reason to believe in the existence of the entities in question. Hence, there are some HE terms that referred to empirical entities.

Realists also admit that there are cases in which HE terms failed to refer to empirical entities. Boyd admits this possibility (Boyd (2010), 216, 220), and so does Psillos,

---

<sup>29</sup>See also Hardin and Rosenberg (1982), 611 for the claim that Dalton referred to atoms in the nineteenth century.

<sup>30</sup>It's worth noting that other realist responses are possible. Alan Chalmers (2008), for example, has argued that the atom did not play a productive role in nineteenth-century chemistry. In this case, it is presumably open to the realist to respond that 'atom' lacked a referent in the nineteenth century, but that 'atom' was not a central theoretical term in chemistry at that time. The success of the theories can be explained by appeal to the reference of terms that were central. This would be to adopt Psillos' strategy with regard to 'caloric' (Psillos (1999), 115–130). The point is just that realists haven't adopted this strategy with regard to 'atom' and that most realists hold that the term referred, at least by the eighteenth century.

<sup>31</sup>See Perrin (1916).

who discusses two examples in a fair amount of detail. He holds that “ ‘phlogiston’ refers to nothing” (Psillos (1999), 291). And his all-things-considered view seems to be that ‘caloric’ is in the same boat (Psillos (1999), 127), though he admits the possibility that it may have referred partially to the internal energy of a substance (Psillos (1999), 128–129).

A realist may want to claim that HE terms that fail to refer to empirical entities actually refer to some kind of abstract entity. For example, one could defend a kind of realist-disjunctivist position, according to which HE terms have putative reference to either empirical entities (in the cases in which the hypothesis will be confirmed) or abstract entities (in the cases in which the hypothesis will be rejected).<sup>32</sup> Instead of non-referring HE terms, then, we would have HE terms that refer to abstract entities. This position would represent something of a departure from what realists have typically held, and the burden of proof would be on realists to develop and defend such a position. That said, the argument against realism that I will offer in chapter 4 is applicable to this kind of position as well. For my purposes, then, this position does not differ significantly from  $R_{HE}$ , and in what follows, I will ignore it.

Finally, realists admit some cases of HE terms that are intermediate between successful reference and failure of reference. Psillos, for example, motivates a notion of “approximate reference,” which he explains as follows:

The current posit is ascribed some (but surely not all) of the *attributes* ascribed to the abandoned putative entity, attributes in virtue of which it was thought to produce its effects. Hence, although there is nothing in the world that possesses

---

<sup>32</sup>I thank Justin Bledin for making this suggestion.

all the attributes ascribed to an abandoned posit  $\alpha$ , there may well be a current posit  $\beta$  which is ascribed some (sometimes most) of the attributes ascribed to  $\alpha$  and is also considered to be causally responsible for the same phenomena as  $\alpha$ . If this situation occurs, (and it's at least arguable that it has occurred in the transition from the luminiferous ether to the electromagnetic field), then we may be willing to say that the term intended to refer to the abandoned posit  $\alpha$  approximately refers to the current posit  $\beta$ . (Psillos (1996), S313)

Psillos also occasionally speaks of terms “referring partially” to empirical entities. One of his examples, as mentioned above, is that ‘caloric’ may have referred partially to the internal energy of a substance (Psillos (1999), 128).

In a similar vein, Boyd admits the existence of “referential relations weaker than full-blown reference” (Boyd (1981), 649). His primary example of such a referential relation is Hartry Field’s notion of partial denotation.<sup>33</sup> Field’s main example of a partially denoting term is Newton’s ‘mass.’<sup>34</sup> According to Field, with the advent of special relativity, two different definitions of mass can be distinguished, which Field labels ‘proper mass’ and ‘relativistic mass.’ Field puts his main point as follows:

Newton’s word ‘mass’ partially denoted proper mass and partially denoted relativistic mass; since it *partially* denoted each of them, it didn’t *fully* (or *determinately*) denote *either*. (Field (1973), 474)

---

<sup>33</sup>See also Boyd’s discussion of partial denotation in Boyd (2010), 216.

<sup>34</sup>Though see Earman and Fine (1977), which shows that Field’s central example is poorly chosen.

Full or determinate denotation would then correspond to Boyd’s “full-blown reference,” while partial denotation would count as a weaker referential relation in Boyd’s sense.<sup>35</sup>

All things considered, the realist is committed to  $R_{HE}$ , which is the view that HE terms are putatively referring expressions that have putative reference to empirical entities. Moreover, the realist is committed to the following seemingly exhaustive trichotomy: either

( $R_{T1}$ ) an HE term, so understood, refers to an empirical entity; or

( $R_{T2}$ ) it altogether fails to refer to an empirical entity; or

( $R_{T3}$ ) it ‘kind of’ refers to that entity.<sup>36</sup>

#### 1.3.4 SOME POINTS OF CONTACT

I’ve emphasized that philosophers of science haven’t framed the issues concerning hypothetical entities and HE terms in quite the same way as I have here. Even so, there are some points of contact with some work done by realist philosophers of science, and we can see some of their positions as fitting the trichotomy discussed above.

One point of contact in the literature is Philip Kitcher’s distinction between “working posits” and “presuppositional posits.” Kitcher draws the distinction in the following passage:

Distinguish two kinds of posits introduced within scientific practice, *working posits* (the putative referents of terms that occur in problem-solving schemata)

---

<sup>35</sup>One might be tempted to object to this identification on the grounds that Field writes of ‘denotation,’ while Boyd writes of ‘reference.’ However, Field intends these to be synonyms (Field (1973), 464).

<sup>36</sup>( $R_{T3}$ ) includes, but is not limited to, views like Psillos’ approximate reference and Field’s partial denotation, which are intermediate between full-blown reference and failure of reference.

and *presuppositional posits* (those entities that apparently have to exist if the instances of the schemata are to be true). (Kitcher (1993), 149)

Kitcher's main example of a presuppositional posit is the ether, since he claims it was "rarely employed in explanation or prediction, [and] never subjected to empirical measurement" until Michelson's famous experiment (Kitcher (1993), 149). He contrasts the ether with "the working posits of theoretical science that are referred to and characterized directly in successful schemata: atoms, molecules, genes, [and] electromagnetic fields" (Kitcher (1993), 149). According to Kitcher, the main lesson that we can draw from the history of theory change in science "is not that theoretical positing in general is untrustworthy, but that presuppositional posits are suspect" (Kitcher (1993), 149).

Kitcher's view fits the general outline of the realist view developed above. It's not clear whether Kitcher identifies posits with theoretical entities, or thinks that the former is a subset of the latter. But his examples of posits would count as hypothetical entities (at least at some point in history) based on the rough guide I set out earlier. His discussion of "the putative referents of terms" is enough to commit him to (RI<sub>1</sub>), and therefore to R<sub>HE</sub>. And his distinction between working posits and presuppositional posits accounts for the difference between (R<sub>T1</sub>) and (R<sub>T2</sub>). Insofar as Kitcher uses this distinction in order to make a claim about what is preserved over theory change, he would claim that working posits would fall under (R<sub>T1</sub>), while presuppositional posits would fall under (R<sub>T2</sub>). As for (R<sub>T3</sub>), Kitcher doesn't employ a notion of approximate reference or partial denotation. But he does develop a theory of reference according to which different tokens of the same term type may differ in reference. So, for example, one token of 'dephlogisticated air' may refer

to oxygen, while another would lack a referent altogether (Kitcher (1993), 100–103). Such HE term types may be cases that fall between full-blown reference and failure of reference, and therefore may be said to ‘kind of’ refer, in the sense intended in (R<sub>T3</sub>).

Chakravartty’s semirealism gives us a second point of contact with the realist literature in the philosophy of science. Chakravrtty begins with the familiar distinction between the observable and the unobservable, but then draws a further distinction at the unobservable level—that between the detectable and the undetectable. Detectables are “unobservables one can detect using instruments,” while undetectables are not detectable, but are merely entities “whose existence one posits for theoretical or explanatory reasons” (Chakravartty (2007), 14–15). As examples of the former, Chakravartty gives the mitochondrion and the neutrino. Examples of the latter include “Newton’s conceptions of position and velocity with respect to absolute space, and causally inefficacious entities such as mathematical objects” (Chakravartty (2007), 15). Chakravartty notes that, with regard to undetectables, “[e]ven if they exist, such things are undetectable” (Chakravartty (2007), 15). This distinction between detectables and undetectables, like the distinction between observables and unobservables, is presumably a *metaphysical* distinction, in the sense that there is a fact of the matter about what we are able to detect, which is independent of our current detection abilities.

Chakravartty goes on to draw an *epistemic* distinction between two different kinds of properties: detection properties and auxiliary properties. To begin with, the properties that particulars in fact possess are what Chakravartty calls “causal properties” (Chakravartty (2007), 41). He goes on to state that

[d]etection properties are causal properties one has managed to detect; they are causally linked to the regular behaviours of our detectors. Auxiliary properties are any other putative properties attributed to particulars by theories. . . . Detection properties are the causal properties one knows, or in other words, the properties in whose existence one most reasonably believes on the basis of our causal contact with the world. The ontological status of auxiliary properties is unknown—they may be causal properties, or fictions. . . . As the sciences move on, some auxiliary properties are retained as auxiliary, some are converted into detection properties, and others are simply discarded. (Chakravartty (2007), 47)

Chakravartty uses Augustin-Jean Fresnel’s famous equations in order to give examples of detection properties and auxiliary properties. Detection properties are properties “that are required to give a *minimal interpretation* of these sorts of equations,” and in the case of Fresnel’s equations, we have “intensities, and directions of propagation” (Chakravartty (2007), 48, 49). He claims that “[a]nything that exceeds a minimal interpretation . . . is auxiliary,” and “[i]n the very limited context of these specific equations, ethers and fields are auxiliary posits” (Chakravartty (2007), 48, 49). The distinction between detection and auxiliary properties is epistemic since it has to do with which properties we know, and which we don’t. As such, the distinction admits to change over time—if we convert auxiliary properties into detection properties, or discard them, we gain some knowledge in the process.

These two distinctions suffice to give the basic idea of Chakravartty’s semirealism: a

realist can “commit to relations of detection properties, and remain agnostic or sceptical about auxiliary properties” (Chakravartty (2007), 48). Semirealism is a sophisticated position that involves more than what I mention here. But this basic idea is sufficient to show that semirealism fits the general outline of the realist view developed above.

Although Chakravartty frames his discussion mainly in terms of properties, it’s not difficult to see how his position extends to entities more generally. He does discuss what might best be classified as objects or substances. For example, he discusses oxygen and dephlogisticated air, which he holds to have different sets of properties (Chakravartty (2007), 55–56). Moreover, he gives the ether as an example of something that is an “auxiliary posit” (Chakravartty (2007), 49). In general, detected entities are detected via their detection properties, while auxiliary posits have no known detection properties, and at most auxiliary properties, which may be discovered to be either detection properties or fictions. An auxiliary posit would count as a hypothetical entity according to the rough guide set out above. If the auxiliary properties are retained as auxiliary, the entity remains hypothetical. If the auxiliary properties are converted into detection properties, or discarded, then the hypothesis is either confirmed or rejected, respectively, and the entity ceases to be hypothetical. Chakravartty, then, can be seen as talking about hypothetical entities in my sense—they are just auxiliary posits. But one important difference here is that Chakravartty’s detection and auxiliary properties apply only at the unobservable level, while it seems clear to me that some hypothetical entities are observable.

Chakravartty’s semirealism falls in line with (R<sub>T1</sub>) and (R<sub>T2</sub>). He claims that since we lack knowledge of the ontological status of auxiliary properties, “they may be causal

properties, or fictions,” but we don’t know which (Chakravartty (2007), 47). Chakravartty would seem to be committed to a similar view with regard to auxiliary posits—either they exist, or they turn out to be fictions, but we don’t know which. And even if an auxiliary posit turns out to be undetectable, Chakravartty claims that undetectables may exist even if we can never detect them (Chakravartty (2007), 15). Chakravartty discusses the example of the neutrino, which was hypothesized in the 1930s, but not detected directly until the 1950s (Chakravartty (2007), 14–15). Though he doesn’t label it as such, the pre-1950s neutrino counts as an auxiliary posit. And by Chakravartty’s own admission, the ether does as well. Since the neutrino has a number of detection properties, semirealists are committed to its existence. The same cannot be said for the ether. So far, my discussion has taken place at the level of metaphysics, but I suggest that these examples fit with the semantic theses of  $(R_{T1})$  and  $(R_{T2})$ . Chakravartty claims that the neutrino that Wolfgang Pauli posited was subsequently detected by Frederick Reines and Clyde Cowan in 1956 (Chakravartty (2007), 14). In this case, Pauli presumably referred to the neutrino before 1956, which fits with  $(R_{T1})$ . The ether is a bit more complicated. Chakravartty admits that undetectables may exist, even if we can’t detect them. If the ether is an undetectable, and it exists, then we can refer to it, and this is another  $(R_{T1})$  case. If it doesn’t exist, then we end up referring to nothing, and this fits with  $(R_{T2})$ . Chakravartty’s view is that we can be “agnostic or sceptical about auxiliary properties” and posits (Chakravartty (2007), 48). And presumably, in order to be “agnostic or sceptical” about such things, one has to admit the possibility that such things exist, and that we may refer to those very things.

$(R_{T3})$  is also left open to Chakravartty, and something like Psillos’ view of approximate

reference might fit with Chakravartty's talk of "graded spectra of commitment" when it comes to "[b]elief in the existence of scientific entities" (Chakravartty (2007), 32–33). Nothing he says commits him to cases of approximate reference or partial denotation, but he doesn't say anything to rule out such cases. In any case, it's enough to show that ( $R_{T1}$ ) and ( $R_{T2}$ ) fit with Chakravartty's semirealism.

Kitcher's and Chakravartty's realist positions, then, fit the general outline of the realist view of hypothetical entities as I have presented it. But realist positions are not the only ones on offer.

#### 1.4 CONSTRUCTIVE EMPIRICISM AND HYPOTHETICAL ENTITIES

For a time, realist philosophers of science seemed to think that their's was the only game in town. They thought that if one rejected realism, one would land in some kind of instrumentalist positivism like that characterized by Frank, among others. And they thought that this kind of position was indefensible. But by the 1980s, van Fraassen had shown how to be an antirealist without being an instrumentalist or a positivist. The thesis of his 1980 book *The Scientific Image* is that we don't have to be realists or positivists, since there is a way to make sense of science that is distinct from both of these positions, namely, a kind of antirealism that van Fraassen labels "constructive empiricism."

Van Fraassen sets up his constructive empiricism in opposition to both realism and positivism. He begins by characterizing realism as follows: "Science aims to give us, in its theories, a literally true story of what the world is like" (van Fraassen (1980), 8). An antirealist position, then would amount to a denial of the claim that literal truth is the aim

of science. But van Fraassen notes that this kind of denial can take two different forms, and these two different forms correspond to two different sorts of antirealism. “The first sort,” according to van Fraassen, “holds that science is or aims to be true, properly (but not literally) construed” (van Fraassen (1980), 10). Such positions would include “conventionalism, logical positivism, and instrumentalism” (van Fraassen (1980), 9). Van Fraassen then moves on to “[t]he second [which] holds that the language of science should be literally construed, but its theories need not be true to be good” (van Fraassen (1980), 10). This is the kind of antirealism that he advocates. His characterization of constructive empiricism mirrors his characterization of realism: “Science aims to give us theories which are empirically adequate” (van Fraassen (1980), 12). Van Fraassen follows this characterization with a “preliminary explication” of empirical adequacy, according to which “a theory is empirically adequate exactly if what it says about the observable things and events in this world, is true—exactly if it ‘saves the phenomena’ ” (van Fraassen (1980), 12).

The two key ideas for van Fraassen, then, are literal construal and empirical adequacy. His commitment to literal construal sets constructive empiricism apart from positivism and instrumentalism. And insofar as he characterizes the aim of science in terms of empirical adequacy rather than truth, his position is distinct from realism. In what follows, I will discuss both of these key ideas, and then extract the constructive empiricist view of hypothetical entities and HE terms from what van Fraassen has to say.

1.4.1 EMPIRICAL ADEQUACY, OBSERVABILITY, AND THE SEMANTIC VIEW

Empirical adequacy is one key component in constructive empiricism, and so my next task is to explain van Fraassen's notion of empirical adequacy.

According to van Fraassen's preliminary explication, empirical adequacy is equated with truth about the observable entities in the world. Hence, constructive empiricism depends on a distinction between the observable and the unobservable. Van Fraassen claims that observation is the same as perception, and that it is something that we can do without the aid of instruments (van Fraassen (2008), 93). As a rough guide, we can say that "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it" (van Fraassen (1980), 16). And "the 'able' in 'observable' refers [to] our limitations, *qua* human beings" (van Fraassen (1980), 17). Van Fraassen holds that what limits there are to observation is an empirical question. He does not draw the distinction in terms of vocabulary, as the positivists did. Indeed, he accepts that "[a]ll our language is thoroughly theory-infected"; instead, he draws the distinction in terms of entities (van Fraassen (1980), 14). He does not view this to be a philosophical matter—it is a theory independent question as to what entities are observable by human beings, and it is the job of science to draw the distinction (van Fraassen (1980), 17, 56–57).

Van Fraassen couches his constructive empiricism, and his more precise account of empirical adequacy, in the semantic view of theories. The semantic view is contrasted with the so-called syntactic or received view, according to which theories are sets of sentences closed under logical consequence. In that case, theories are essentially linguistic entities. On the semantic view, theories are classes of models. Proponents of the semantic view

like Frederick Suppe have claimed that theories are therefore “extralinguistic entities which may be described or characterized by a number of different linguistic formulations” (Suppe (1977), 221). Van Fraassen’s more precise account of empirical adequacy, in terms of the semantic view, is as follows:

To present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the *empirical substructures*) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call *appearances*: the theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model.

(van Fraassen (1980), 64)

Van Fraassen has emphasized more recently that the *phenomena* (the observable entities) are to be distinguished from the *appearances* (the contents of measurement reports) (van Fraassen (2008), 8, 283–286). A theory is empirically adequate, then, if the appearances (1) represent the phenomena, and (2) are embeddable in at least one model of the theory.

Getting clearer on (1) requires saying a bit more about the appearances, the phenomena, and the relation between the two. Van Fraassen offers astronomy as an illustration (van Fraassen (2008), 285). The planetary orbits are observable processes, and so count as phenomena in van Fraassen’s sense. Astronomers record what they observe in the night sky, and the contents of their measurement outcomes are the appearances, i.e., how the planetary orbits appear from a specific viewpoint. Van Fraassen brings out the difference

between phenomena and appearances by means of the example of Mercury: “Mercury’s motion is an observable phenomenon, but Mercury’s *retrograde motion* is an appearance” (van Fraassen (2008), 287). The relation that obtains between the appearances and the phenomena is one of representation, and van Fraassen argues for the claim that

[r]epresentation is a relation between the abstract structure and the phenomena constituted by the user. Nothing represents anything except in the sense of being used or taken to do that job or play that role for us. (van Fraassen (2008), 253).

The abstract structure that van Fraassen mentions is a data model, which is constructed from a number of measurement outcomes, the contents of which are the appearances. His view then, in short, is that representation amounts to a three-place relation that involves a data model, the phenomena, and the user of the data model. So astronomers (the users) may take Mercury’s motion (a phenomenon) to be represented by a data model, which includes the contents of astronomers’ measurements of the apparent motion of Mercury in the night sky (the appearances).

Getting clearer on (2) requires the introduction of some notation, which I will borrow from Otávio Bueno’s useful discussion of van Fraassen’s embeddability account of empirical adequacy (Bueno (1997), 588–589). To begin with, we can think of a scientific theory as a class of *theoretical models*. A *theoretical model*  $T$  is given by the mathematical structure  $\langle D, R_i^n \rangle_{i \in I}$ , where  $D$  is a set of objects and  $R_i^n$  is a family of  $n$ -place relations holding among the elements of  $D$ . An *empirical substructure*  $E$  of some theoretical model  $T$  is given by

the substructure  $E = \langle D', R_i^n \rangle_{i \in I}$ , where,  $D' \subseteq D$  and  $R_i^n$  is the family of  $n$ -place relations holding among the elements of  $D'$ . The *appearances* are given by a mathematical structure  $A = \langle D'', R_i^{m'} \rangle_{i \in I}$ , where  $D'' \subseteq D$  and  $R_i^{m'}$  is a family of  $n$ -place relations holding among the elements of  $D''$ . A theory is *empirically adequate* if the appearances are *embeddable* in at least one of the theory's theoretical models; that is to say, if there is a bijection  $f: D'' \rightarrow D'$  of  $A = \langle D'', R_i^{m'} \rangle_{i \in I}$  onto an empirical substructure  $E = \langle D', R_i^n \rangle_{i \in I}$  of  $T$ , such that for every  $n$ -tuple  $(x_1, \dots, x_n)$  of elements of  $D''$ , and for every relation  $R'' \in R_i^{m'}$ ,  $R''(x_1, \dots, x_n)$  if and only if  $R'(f(x_1), \dots, f(x_n))$  for some  $R' \in R_i^n$ , which is to say that there is an *isomorphism* between the appearances  $A$  and a particular empirical substructure  $E$ . And importantly, the elements of  $D'$  and  $D''$  are observable.

More recently, van Fraassen's thinking has moved in a more structuralist direction (van Fraassen (2006); van Fraassen (2008), 237–261). He now defends what he calls an “empiricist structuralism,” which is essentially a more explicitly structuralist version of constructive empiricism. He begins with the slogan that “all we know is structure,” and offers the following construal of that slogan:

- I Science represents the empirical phenomena as embeddable in certain *abstract structures* (theoretical models).
- II Those abstract structures are describable only up to structural isomorphism. (van Fraassen (2008), 238)

Van Fraassen claims that the sense in which this is a structuralist position amounts to the fact that it accepts “the thesis that all scientific representation is at heart mathematical”

(van Fraassen (2008), 238–239). This doesn't represent a break with his earlier characterizations of constructive empiricism, since he has always emphasized a relationship between mathematical structures and the phenomena. His structuralist empiricism is therefore best read as an attempt to locate constructive empiricism among various structural realist views, which I will discuss in some detail below.<sup>37</sup>

#### 1.4.2 LITERAL CONSTRUCTIONAL AND REALISTIC INTERPRETATION

As I emphasized above, van Fraassen distinguishes himself from the realists by setting up empirical adequacy as the aim of science, while he distinguishes himself from the positivists by employing the literal construal of language. Before I turn to the latter, I will address a preliminary worry. Van Fraassen couches his position in terms of the semantic view of theories. Proponents of the semantic view have emphasized that theories are extralinguistic entities. So, if theories are not linguistic entities, how can there be any room for the literal construal of language, which, as we've seen, van Fraassen is committed to?

This worry might seem to highlight some tension in van Fraassen's view, but upon closer examination, there is really no problem here. This is because the semantic view is a view about *theories*, while literal construal is a view about *language*. In other words, one can consistently claim that theories are extralinguistic entities, while at the same time claiming that scientific language should be construed literally. There is, moreover, a relationship between the two views. As Suppe emphasizes, these “extralinguistic entities . . . may be described or characterized by a number of different linguistic formulations”

---

<sup>37</sup>See also the ‘structural empiricism’ of Bueno (1999) for another attempt to spell out a more explicitly structuralist version of constructive empiricism.

(Suppe (1977), 221), and constructive empiricism requires that all of these linguistic formulations be construed literally.

So what, then, is van Fraassen's view of literal construal? Van Fraassen's readers have, on the whole, agreed that there is no real semantic difference between realism and constructive empiricism, and that the latter is committed to realistic interpretation.<sup>38</sup> But although van Fraassen's literal construal and realistic interpretation as I have characterized it are quite similar, there are, in fact, some differences between the two.

Van Fraassen does agree with the realist insofar as he holds

(RI<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable of being true or false.

He claims that “on a literal construal, the apparent statements of science really are statements, *capable of being true or false*” (van Fraassen (1980), 10). This mirrors what realists like Psillos claim when they say that scientific claims “are capable of being true or false” (Psillos (1999), xix), and commits van Fraassen to (RI<sub>2</sub>). But van Fraassen also claims that “‘literal’ does not mean ‘truth-valued’ ” (van Fraassen (1980), 10). As an example, he discusses P. F. Strawson's treatment of sentences with nonreferring terms or descriptions (e.g., ‘The king of France in 1905 is bald’), which, according to Strawson, lack a truth value (Strawson (1950); van Fraassen (1980), 38). This, of course, goes against views like Bertrand Russell's, according to which such sentences are false (Russell (1905)). On a weak reading, van Fraassen allows for those who employ literal construal to decide the issue themselves, and follow either Strawson or Russell. On a strong reading, van Fraassen

---

<sup>38</sup>See, for example, Boyd (1983), 46; Chakravartty (2007), 10–11; and Frost-Arnold (2011), 1132.

is claiming that literal construal follows Strawson as opposed to Russell. But even on this strong reading, statements like ‘The king of France in 1905 is bald’ are still *capable of* being true or false, since ‘the king of France in 1905’ is a putatively referring expression. It’s just that it lacks a referent, and therefore the sentence lacks a truth value.<sup>39</sup> (RI<sub>2</sub>) also leaves it open whether one follows Strawson or Russell on this issue, and I take it that proponents of realistic interpretation can share (RI<sub>2</sub>) and still disagree on the issue of truth-value gaps. Chakravartty, for example, holds that scientific claims “should be construed literally as having truth values, whether true or false” (Chakravartty (2011), §1.2). So while Chakravartty may accept (RI<sub>2</sub>), he also accepts a stronger claim that rules out the existence of truth-value gaps when it comes to the language of science. In that case, while there is general agreement between realists and constructive empiricists about (RI<sub>2</sub>), it can sometimes be the case that there are semantic differences between proponents of the two positions.

Van Fraassen’s views on literal construal also commit him to something close to (RI<sub>1</sub>) Theoretical terms are putatively referring expressions that have putative reference to empirical entities.

When it comes to theoretical terms that putatively refer to observable entities, he claims that reference to these entities is “unproblematic” (van Fraassen (2008), 3). The real issue, then, arises when it comes to unobservable entities. Constructive empiricism is committed to the

---

<sup>39</sup>Van Fraassen’s development of free logic is premised on Strawson’s view of the matter, and therefore presents some *prima facie* evidence that the strong reading is preferable. I won’t pursue this point any further, but see, for example, (van Fraassen (1966)). The connection between van Fraassen’s free logic and his views on literal construal was suggested to me by Greg Frost-Arnold.

idea that, if an entity is unobservable, it must be possible to be agnostic about its existence (van Fraassen (2001), 151). There would therefore have to be a corresponding agnosticism with regard to the semantics of statements that are purportedly about unobservable entities, which is to say that it must be possible to be agnostic about whether a theoretical term refers to an unobservable entity. But what van Fraassen is agnostic about is *whether* the term refers, not *that* the term is a putatively referring expression. In fact, agnosticism here presupposes that the terms are putatively referring expressions that can have putative reference to empirical entities. Van Fraassen also claims that “[n]ot every philosophical position concerning science which insists on a literal construal of the language of science is a realist position” (van Fraassen (1980), 11), which indicates that he himself thinks that the realist and the constructive empiricist agree on the literal construal of language. In that case, van Fraassen is committed to something like (RI<sub>1</sub>).

However things are slightly more complicated, and van Fraassen wouldn't fully endorse (RI<sub>1</sub>) as is. Immediately after insisting on literal construal, he makes the following comment regarding theoretical terms:

But this [literal construal] does not settle very much. It is often not at all obvious whether a theoretical term refers to a concrete entity or a mathematical entity. Perhaps one tenable interpretation of classical physics is that there are no concrete entities which are forces—that ‘there are forces such that . . .’ can always be understood as a mathematical statement asserting the existence of certain functions. That is debatable. (van Fraassen (1980), 11)

Van Fraassen thereby allows a view more nuanced than (RI<sub>1</sub>), according to which a theoretical term can be construed literally and still purportedly refer, not to a concrete entity ('empirical entity' in my terminology), but to a mathematical entity. This is only an issue for theoretical terms that purportedly refer to unobservable entities.<sup>40</sup> It is possible for a realist to follow van Fraassen here. But this would spell some trouble for the realist's no-miracles argument, since it's not immediately clear how reference to *mathematical* entities would form any part of an explanation of the *empirical* success of science. To put it another way, a realist might attempt to explain the success of science by appeal to the reality of forces. But this explanation would only go through if 'force' refers to empirical entities, as opposed to mathematical ones, which are presumably causally inert. In what follows, I'll assume that the realist will not want to follow van Fraassen in this.<sup>41</sup>

Although there is much agreement between realists and constructive empiricists, given the subtle differences, I introduce the twofold central idea behind van Fraassen's view of literal construal:

(LC<sub>1</sub>) Theoretical terms are putatively referring expressions that have putative reference to either empirical or mathematical entities.

(LC<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable of being true or false, though some may, in fact, lack a truth value.

This is basically a variant of realistic interpretation, and, on the whole, realists and constructive empiricists have much more in common semantically with each other, than either

---

<sup>40</sup>See van Fraassen (1980), 15 for the claim that mathematical entities are unobservable.

<sup>41</sup>If one thinks otherwise, nothing much hinges on this point, since I will object to both van Fraassen and the realist in chapter 4.

position does with, say, logical positivism.

#### 1.4.3 CONSTRUCTIVE EMPIRICISM ABOUT HYPOTHETICAL ENTITIES

The key to extracting the constructive empiricist view of hypothetical entities is analogous to that for the realist view—it comes with the realization that HE terms are a subset of theoretical terms. In that case,  $(LC_1)$  applies to them straightforwardly, which yields what I will call ‘constructive empiricism about hypothetical entities,’ or ‘ $CE_{HE}$ ’ for short.

$CE_{HE}$  HE terms are putatively referring expressions that have putative reference to either empirical or mathematical entities.

From this, it follows that constructive empiricists are committed to something like the realist trichotomy discussed above: either

$(CE_{T1})$  an HE term, so understood, refers to either an empirical or a mathematical entity;

or

$(CE_{T2})$  it altogether fails to refer to an empirical or mathematical entity; or

$(CE_{T3})$  it ‘kind of’ refers to that entity.

$(CE_{T1})$  and  $(CE_{T2})$  follow straightforwardly from  $CE_{HE}$ , and I include  $(CE_{T3})$  because nothing that van Fraassen says rules out the use of partial denotation, approximate reference, and the like.

Van Fraassen doesn’t devote much discussion to the reference of theoretical terms.<sup>42</sup>

This makes sense, insofar as he thinks that reference to observable entities is unproblem-

---

<sup>42</sup>Though he does devote extensive discussion to representation, and occasionally writes of the referent of a representation. See van Fraassen (2008), 245 for one example.

atic. The implication here is that reference to unobservable entities *is* problematic, but constructive empiricists can attempt to justify ignoring such puzzles by appeal to the fact that they can remain agnostic about whether terms refer to unobservable entities.

When it comes to HE terms that name observable hypothetical entities,  $(CE_{T1})$ – $(CE_{T3})$  collapse into  $(R_{T1})$ – $(R_{T3})$ . This is a consequence of van Fraassen’s view that reference to observable entities is unproblematic. It’s only when we deal with HE terms that name unobservable hypothetical entities that reference to mathematical entities becomes a possibility. So if we restrict ourselves to observable hypothetical entities, there is no difference between the realist and the constructive empiricist. More generally, realists and constructive empiricists needn’t have any substantive disagreements over the status of the observable—it’s really only the unobservable about which they disagree.

When it comes to HE terms that name unobservable hypothetical entities, the main difference between the realist and the constructive empiricist, at least as I’ve characterized it, involves the possibility of referring to a mathematical entity. As I claimed earlier, it’s open to the realist to do the same, and to accept  $(CE_{T1})$ – $(CE_{T3})$ . But this may undercut the realist’s use of the no-miracles argument, in which case this issue concerning reference to mathematical entities is a real difference between realism and constructive empiricism.

Strictly speaking, van Fraassen’s nominalism precludes him from claiming that terms *actually* refer to mathematical entities, and so  $CE_{HE}$  may seem to be problematic. In light of these potential difficulties, I will discuss three responses. First of all, van Fraassen can claim that HE terms, in some cases, actually *name* mathematical entities, in the sense that I am using that term, i.e., the sense that doesn’t entail that reference is successful. Sec-

ondly, van Fraassen can claim that one needn't be a nominalist in order to be a constructive empiricist. Since the former is a view about abstract entities, and the latter a view about science, it makes sense that the two positions would be independent. Though to be sure, van Fraassen believes that both follow from a more general commitment to empiricism.<sup>43</sup> Thirdly, it would be odd to tie constructive empiricism to realism about mathematical entities, in which case we should keep in mind the following claim from Bradley Monton and van Fraassen:

Without offering a rival philosophy of mathematics, we may proceed in philosophy of science in the conviction that any satisfactory philosophical account of mathematics must imply that the sorts of applications of mathematics needed in philosophy of science are acceptable, correct, and intelligible. (Monton and van Fraassen (2003), 412)<sup>44</sup>

Presumably, talk of reference to mathematical entities is one of the things to be explained, or else explained away, by an adequate philosophy of mathematics.

To consider one last potential difficulty, according to the rough guide, scientists put forward hypothetical entities as (purported) empirical entities, in which case it may seem problematic that an HE term can refer to a mathematical entity, as  $CE_{HE}$  claims. But once we dwell on some other examples from science, we should be able to see that, at the very least, this isn't a problem that is unique to  $CE_{HE}$ , and at most, it's not a problem at all.

---

<sup>43</sup>Indeed, Monton and van Fraassen have made this response with regard to *modal* nominalism (Monton and van Fraassen (2003), 419–421), in which case it doesn't seem inappropriate to make this response with regard to nominalism more generally.

<sup>44</sup>See also the following quote from van Fraassen, from the discussion in Hilgevoord (1994), 269: "One of my great regrets in life is that I do not have a philosophy of mathematics; I will just assume that any adequate such philosophy will imply that what we do when we use mathematics is all right."

To take one example, the term ‘orbital’ was initially introduced to name the paths of electrons around the nucleus. But now, we take it that ‘orbital’ names a quantum mechanical probability density, and a probability density is a mathematical entity. In this case, it’s not absurd to claim that a term initially used to name an empirical entity ends up naming a mathematical entity.

## 1.5 STRUCTURAL REALISM AND HYPOTHETICAL ENTITIES

For most of the 1980s, the realism debate amounted to a two-sided debate between realists and constructive empiricists. But in his 1989 paper “Structural Realism: The Best of Both Worlds?”, John Worrall (re)introduced a third position, namely, structural realism, into the scientific realism debate. It is to this position that I now turn.

The two “worlds” that Worrall mentions in the title are really two apparently conflicting arguments. The first is the realist’s no-miracles argument (Worrall (1989), 101). As we’ve already seen, the basic idea behind the no-miracles argument is that the explanatory and predictive success of our best scientific theories would be miraculous if those theories weren’t at least approximately true.

The second argument that Worrall discusses is the antirealist’s so-called pessimistic induction (Worrall (1989), 109). The phenomenon of theory change shows that scientific theories are invariably replaced by successor theories. Such successor theories retain and extend the empirical successes of their predecessors, but only at the price of discontinuity at the theoretical level.<sup>45</sup> In other words, some central theoretical entities posited by the

---

<sup>45</sup>Here, and in what follows, I will adopt Worrall’s use of terminology, according to which the theoretical

predecessor theory do not find their way into the ontology of the successor theory. In this case, the predecessor theory cannot be even approximately true. The pessimistic induction, then, is an induction on all of these rejected theories. If these rejected theories are not even approximately true, in light of the fact that their theoretical claims conflict with those of our current best theories, then why think that our current best theories will fare any better? The conclusion of the pessimistic induction is that we should regard all theories, even our best ones, as not even approximately true.<sup>46</sup>

Worrall illustrates these apparently conflicting arguments in terms of an example of theory change from the history of science, namely, the transition from Fresnel's wave theory of light to the "mature form" of James Clerk Maxwell's electromagnetic field theory.<sup>47</sup>

As is well-known, Fresnel's wave theory enjoyed tremendous predictive success. It is a consequence of Fresnel's theory that there should be an illuminated spot at the center of a shadow cast by an opaque disc. Fresnel's peers thought this absurd consequence showed

---

content of a theory is that which purports to describe the unobservable reality, while the empirical content purports to describe the observable reality.

<sup>46</sup>Most philosophers of science characterize the argument in Laudan (1981) as a form of the pessimistic induction. For some examples, see Psillos (1996), Papineau (2010), and Frost-Arnold (2011). As I understand Laudan, this is a mistake. Laudan invokes his laundry list of examples from the history of science in order to show that reference and approximate truth cannot explain the explanatory and predictive success of our scientific theories. Examples like Fresnel's wave theory of light are meant to show that neither reference nor approximate truth is necessary to explain success. And examples like the chemical atomic theories of the eighteenth century are meant to show that reference is not sufficient to explain success. These examples are meant to function as counterexamples to various realist claims, and are meant to undermine the no-miracles argument. But there is no inductive step by which Laudan concludes that our current best theories are false. Indeed, Laudan concludes his paper by claiming that "[n]othing [he has] said here refutes the possibility in principle of a realistic epistemology of science" (Laudan (1981), 48). This claim would be puzzling if Laudan were, in fact, making a pessimistic induction. But it is intelligible if he is merely trying to show that success is no evidence that our theories are approximately true or genuinely referential. Ladyman (2011), 93 seems to be one of the few philosophers to recognize this point.

<sup>47</sup>By the "mature form" of Maxwell's theory, Worrall intends the form that "explicitly denies that these field strengths can in turn be explained via the contortions of some underlying mechanical medium" (Worrall (2007), 127–128). Worrall makes this qualification since Maxwell himself believed that there had to be such a medium.

that Fresnel had to be wrong. But this surprising prediction turned out to come true when tested. The intuition behind the no-miracles argument leads us to entertain the idea that this kind of success would be miraculous unless Fresnel's theory were at least approximately true. But once we look back at the history of science, we realize that Fresnel's theory cannot be approximately true. This is because Fresnel's theory posits a luminiferous ether, i.e., a solid mechanical medium through which light waves propagate. But according to Maxwell's theory (at least in its "mature form"), light is a displacement current in the electromagnetic field, and not a wave in the ether. If Maxwell's theory is true, then Fresnel's theory cannot be even approximately true, since, according to the former, the ether does not exist. This is, of course, the main idea behind the pessimistic induction—in the same way that Maxwell's theory shows that Fresnel's theory is not even approximately true, future theories will show that our current best theories are not even approximately true. The problem, then, is to come up with a non-miraculous explanation for the success of science that doesn't run afoul of the historical phenomenon of theory change, upon which the pessimistic induction relies.

Worrall claims that structural realism gives us the resources to address this problem. Structural realism is a position that Worrall claims to find in the work of Henri Poincaré (1962) and Pierre Duhem (1991).<sup>48</sup> The basic idea is that, when a theory in a mature science is rejected in favor of some successor theory, something more than empirical content is preserved. At the same time, what is preserved is less than the full theoretical content. Worrall's claim is that, at the theoretical level, there is preservation of *structure* over theory

---

<sup>48</sup>Worrall also cites the more Russellian structural realism of Grover Maxwell as an antecedent view. See, for example, Maxwell (1970).

change, but that our theories never get at the *nature* of the unobservable reality. With regard to the Fresnel-Maxwell case, Worrall gives the following rough account:

Roughly speaking, it seems right to say that Fresnel completely misidentified the *nature* of light, but nonetheless it is no miracle that his theory enjoyed the empirical predictive success that it did; it is no miracle because Fresnel's theory, as science later saw it, attributed to light the right *structure*. (Worrall (1989), 117)

According to Worrall, we have no justification for thinking that our best theories ever get at the nature of the unobservable reality (Worrall (1989), 120). We cannot say that light is really made up of waves in a mechanical medium, or displacement currents in a disembodied electromagnetic field, or photons that obey the laws of quantum mechanics. But there is structural continuity over theory change, and this is evident from the fact that, in mature sciences, the equations of a theory are limiting cases of the equations of its successor (Worrall (1989), 120). We see an extreme example of this in the transition from Fresnel's theory to Maxwell's. Fresnel's equations were actually carried over into Maxwell's theory intact. The interpretations of these equations (in terms of the nature of light) are, of course, very different in the two theories. But the equations themselves are the same. Fresnel and Maxwell may have been wrong about the nature of light, but they were "quite right not just about a whole range of optical phenomena but right that these phenomena depend on something or other that undergoes periodic change at right angles to the light" (Worrall (1989), 120). This is the sense in which both theories correctly identify the structure of the

unobservable reality, at least to an approximation.

At this point, we can see how Worrall attempts to accommodate both the no-miracles argument and the pessimistic induction. The pessimistic induction relies on the fact that there is theoretical discontinuity over theory change, and Worrall claims that this discontinuity is one of nature. Worrall would then agree with the proponent of the pessimistic induction, who thinks that we lack good grounds to consider our theories approximately true. But Worrall points to some theoretical continuity over theory change, and this continuity is one of structure. Hence, the success of science is no miracle, so long as our theories have some grasp on the structure of the unobservable reality. Fresnel's theory, in particular, enjoyed the success that it did in light of the fact that Fresnel's equations approximate the structure of light. And this is consistent with the fact that Fresnel misidentified light's nature.

Worrall's paper has sparked a voluminous literature on structural realism, which now has a number of prominent defenders. One of the most prominent is James Ladyman, who was the first to point out that Worrall's position in his 1989 paper is actually ambiguous between an epistemological position and a metaphysical one (Ladyman (1998)). In order to track this distinction, Ladyman has since introduced the terminology of 'epistemic structural realism' (ESR) and 'ontic structural realism' (OSR), both of which, in turn, include their own sub-varieties (Ladyman (2009)). He characterizes the basic idea of each position as follows:

A crude statement of ESR is the claim that all we know is the structure of the relations between things and not the things themselves, and a corresponding crude statement of OSR is the claim that there are no 'things' and that structure

is all there is ... (Ladyman (2009), §4)

In what follows, I will focus on one example of ESR (Worrall's more recent development of structural realism), and one example of OSR (the structural realism that Ladyman and his collaborators have developed). For each position, I will discuss that position's semantic commitments, and use these commitments to extract the view of hypothetical entities and HE terms that that position is committed to.

### 1.5.1 EPISTEMIC STRUCTURAL REALISM

Although Ladyman (1998) makes a convincing argument for the claim that the position sketched in Worrall (1989) is ambiguous between ESR and OSR, Worrall has since developed his structural realism explicitly in the direction of ESR.<sup>49</sup> I will now turn to the details of Worrall's ESR. I'll begin by discussing how Worrall explains structure in terms of the Ramsey sentence of a theory. I will then turn to ESR's semantic commitments. Finally, I'll extract the view of hypothetical entities and HE terms that a proponent of Worrall-style ESR is committed to.

### **Structure and Ramsey Sentences**

Perhaps the best way to understand Worrall's ESR is in terms of his views on Ramsey sentences and reference. Worrall claims that "the full 'cognitive content' of a theory T is captured by its Ramsey sentence R(T)" (Worrall (2011), 168). Ignoring, for the moment,

---

<sup>49</sup>See Worrall (2008) and Worrall (2011), 170 for some evidence that Worrall is now a self-identifying proponent of ESR.

what Worrall means by “cognitive content,” we can formalize T as follows:

$$T = (O_1, \dots, O_n; S_1, \dots, S_m)$$

In the course of formalizing T, Worrall distinguishes between observational predicates (the  $O_i$ ) and theoretical predicates (the  $S_j$ ) (Worrall (2007), 152). Worrall does not attempt to draw a distinction between the two kinds of predicates in any kind of rigorous way. Instead, he operates with an “intuitive distinction that yields gluons, quarks, electrons, spacetime curvatures, and light waves, for example, as theoretical, and planets, people, tracks on cloud chamber photographs, and interference fringes, for example, as observable” (Worrall (2011), 166).<sup>50</sup> The Ramsey sentence of T, then, is as follows:

$$R(T) = \exists \Phi_1 \dots \exists \Phi_m (O_1, \dots, O_n; \Phi_1, \dots, \Phi_m)$$

In the Ramsey sentence, all theoretical predicates are quantified over, while the observational predicates are left alone.

So what does Worrall have in mind when he claims that T and R(T) do not differ in cognitive content? We can begin to see the answer by examining Worrall’s rejection of the causal theory of reference in favor of descriptivism when it comes to theoretical predicates (Worrall (2007), 148–154; Worrall (2011), 169–170). Worrall, as a descriptivist, holds that

---

<sup>50</sup>It should be noted that my use of ‘theoretical term,’ as explained above, differs from Worrall’s use of ‘theoretical predicate.’ As noted above, Worrall tends to equate ‘theoretical’ and ‘unobservable.’ Words like ‘planet’ and ‘interference fringe’ would count as theoretical terms for me, but not theoretical predicates for Worrall. To make matters more complicated, Worrall vacillates between talk of ‘predicates’ and ‘terms,’ which are, of course, logically distinct. In the course of giving an exposition of Worrall’s views, I will follow his usage.

the only way in which our theoretical predicates latch on to reality is by means of some description. Causal theorists, on the other hand, hold that descriptions are neither necessary nor sufficient for our predicates to latch on to reality. Worrall holds that when it comes to theoretical predicates, causal theorists face a dilemma—the causal theory either (1) engages in some kind of fantasy, or else it (2) collapses into descriptivism. Regarding (1), Worrall claims that “[i]t is just a fantasy . . . that we can ‘stand outside’ of our theories and directly compare terms in them with a reality that we can access directly without any theory” (Worrall (2007), 148). Causal theorists who hold that the reference of our theoretical predicates is unmediated by some kind of description are engaging in this kind of fantasy. The way out, for the causal theorist, is to admit that the reference of our theoretical predicates is mediated by some kind of description. But in this case, the causal theorist falls into (2), the second horn of the dilemma. To make the same point in another way, we can take one of Worrall’s examples, namely, the electromagnetic field (Worrall (2007), 148). On the causal theory, the reference of ‘electromagnetic field’ is fixed via some baptismal episode, in which some speaker picks out the electromagnetic field, and dubs it with the theoretical predicate ‘electromagnetic field.’ But how is this baptism supposed to work? A speaker can’t merely ostend the electromagnetic field. So in this case as well, the causal theory must admit that descriptions play some role, or else engage in some fantasy according to which ostension in such cases is possible. For these reasons, Worrall rejects the causal theory of reference and accepts descriptivism when it comes to theoretical predicates.<sup>51</sup>

---

<sup>51</sup>My primary aim is to give an exposition of Worrall’s views. But here I want to briefly turn a critical eye towards his views, in the hopes that doing so will shed some light on those views. Worrall’s talk of ostension and baptism suggest that, when he argues against the causal theory of reference, his primary example is the picture first put forward by Saul Kripke (1980). This suggests that Worrall’s own descriptivism is something

According to Worrall, once we accept descriptivism for theoretical predicates, it follows that T and R(T) do not differ in cognitive content. Worrall explains this as follows:

But if all our knowledge of theoretical entities is descriptive, then it follows that if you are asked what, say, the term ‘gluon’ refers to all you can do is reiterate our current best (total!) theories of gluons: that is, a gluon is a ‘whatever it is’ that structures the phenomena in certain complex ways through specific intricate relationships with the phenomena and with other, similarly characterised, theoretical notions. This characterisation, however, is just an informal statement of the Ramsey sentence for our theory of gluons, in which the theoretical predicates have been replaced by second-order quantifiers. (Worrall (2011), 169)

The basic idea here is that if we accept descriptivism, then the meaning of a theoretical predicate like ‘gluon’ is no different from its Ramseyfied analogue. And, of course, Worrall intends this basic idea to apply to all theoretical predicates. He goes on to elaborate:

The primitive theoretical predicates in the initial un-Ramseyfied theory name (or attempt to name) theoretical entities in the same way that the ambiguous names involved in some systems of predicate logic do—that is, not directly in the way that we think of regular individual constants naming individuals but through the sentences we assert using them. And of course in such systems

---

along the lines of Kripke’s target in *Naming and Necessity*. But Kripke is clear in that work that he is not advancing a theory at all, but rather a “better picture” of how names refer (Kripke (1980), 93). Moreover, he readily admits cases in which reference is fixed by description (Kripke (1980), 79–80). It’s therefore not clear that Worrall’s argument against the so-called ‘causal theory of reference’ is effective against its intended target.

of first order logic, where  $\alpha$  is any ambiguous name,  $P\alpha$  and  $\exists xPx$  are inter-derivable and so ‘cognitively equivalent’. (Worrall (2011), 169)

A theoretical predicate, then, “plays the role of” a second-order ambiguous name (Worrall (2007), 148).<sup>52</sup> This is what Worrall means when he claims that T and R(T) do not differ in cognitive content. Elsewhere, he claims that T “reduces to” R(T), and that T “carries no further epistemically accessible content than” R(T) (Worrall (2011), 170). It seems that he means these to be synonymous with the claim that T and R(T) do not differ in cognitive content.

So where do structure and nature fit in here? Beginning with structure, if R(T) exhausts the epistemically accessible content of T, and structure, according to ESR, is all that we have epistemic access to, then R(T) gives the structure of T. We can illustrate this in terms of the following example of an extremely simplistic theory:<sup>53</sup>

$$V = \forall x[(P_1x \wedge P_2x) \rightarrow \exists yQy]$$

where  $P_1x$ : ‘x is a radium atom,’  $P_2x$ : ‘x decays radioactively,’ and  $Qx$ : ‘x is a click in a suitably placed Geiger counter.’  $P_1x$  and  $P_2x$  are theoretical predicates, while  $Qx$  is an

---

<sup>52</sup>In the same place, Worrall directs the reader to Patrick Suppes (1957) for the terminology ‘ambiguous name.’ According to Suppes, since “we have no definite individual in mind when we use” an ambiguous name, it is “not a genuine proper name” (Suppes (1957), 81). Suppes’ convention is to use an ambiguous name when performing an existential instantiation within a natural deduction proof. Other conventions involve the introduction of a new constant that acts as a “temporary name” (Barwise and Etchemendy (2002), 322–323), and I take it that this is equivalent to what Worrall has in mind by ‘ambiguous name.’

<sup>53</sup>I borrow this example from Frigg and Votsis (2011), 247, who in turn borrow it from Maxwell (1970), 186.

observational predicate. The Ramsey sentence of  $V$  is:

$$R(V) = \exists \Phi_1 \exists \Phi_2 \forall x [(\Phi_1 x \wedge \Phi_2 x) \rightarrow \exists y Qy]$$

For Worrall, then,  $V$  and  $R(V)$  do not differ in cognitive content, and  $R(V)$  gives the structure of  $V$ .<sup>54</sup> More generally, the Ramsey sentence of a theory gives the structure of that theory, and structure amounts to the logical relations that obtain among uninterpreted theoretical predicates and interpreted observational predicates.

At this point, we can turn to nature, or alternatively, non-structural theoretical content. Ioannis Votsis suggests that “[o]ne approach, *implicit* in Worrall’s work, is to reduce talk about natures to talk about theoretical interpretations” (Votsis (2007), 66). Ramseyfying, then, “strips a theory’s theoretical terms of their interpretation and leaves the logical structure and observational interpretation intact” (Votsis (2007), 66). Assuming that Votsis’ gloss on Worrall’s view is accurate, we can say that, when we Ramseyfy some theory, whatever is ‘Ramseyfied out’ is nature. And, indeed, Votsis’ understanding of the matter coheres with what Worrall says. To see this, we must first recall that, according to Worrall, a theoretical predicate is really just a second-order ambiguous name, and this is why  $V$  and  $R(V)$  do not differ in cognitive content. In  $V$ , I’ve assigned the interpretation ‘ $x$  is a radium atom’ to  $P_1x$ . But if Worrall is right that a theoretical predicate is just a second-order ambiguous name, then  $P_1x$  cannot bear this interpretation. More generally, no theoretical predicate can

---

<sup>54</sup>One might think that  $V$  and  $R(V)$  must differ in cognitive content, on the grounds that  $R(V)$ , but not  $V$ , is trivially true. The trivial truth of  $R(V)$  depends on taking  $Q$  to be equivalent to  $(P_1x \wedge P_2x)$ . But, given that  $Q$  is an observational predicate, and it has already been given the interpretation ‘is a click in a suitably placed Geiger counter,’ it cannot be identified with  $(P_1x \wedge P_2x)$ . And in this case, the truth of  $R(V)$  would not be trivial.

bear an interpretation.<sup>55</sup> In that case, we can identify the non-structural theoretical content (i.e., nature) of a theory with the interpretation of a theoretical predicate.

But if  $R(T)$  is all there is to the cognitive content of  $T$ , then is there anything that is, in fact, Ramseyfied out? Votsis notes that Worrall has actually abandoned the view that there is non-structural theoretical content (i.e., nature), though he admits that it may appear that there is. “More accurately,” writes Votsis, “[Worrall] suggests that it is not meaningful to speak of non-structural theoretical content since all theoretical assertions are structural” (Votsis (2004), 83). Worrall views this to be a consequence of accepting the idea that  $T$  and  $R(T)$  do not differ in cognitive content (Votsis (2004), 83). Votsis, on the other hand, argues that Worrall’s ESR needs to appeal to nature in order to accommodate the pessimistic induction—nature, after all, is what is discarded over theory change (Votsis (2004), 83–84).<sup>56</sup> We can see the difference between the two positions in terms of the example of the transition from Fresnel’s theory to Maxwell’s. For Votsis, the claim that light is a periodic vibration in the ether was discarded in the transition to Maxwell’s theory. For Worrall, all that this claim ever meant to begin with is that *something or other* undergoes a periodic vibration. Any appearance to the contrary results from ignoring the fact that ‘ether’ plays the role of a second-order ambiguous name that cannot bear any kind of theoretical interpretation, in which case there was nothing that could have been discarded. More generally, for Votsis, non-structural theoretical content is actual (though presumably not part of the *cognitive* content of the theory), and Ramseyfying is the means by which we strip it from

---

<sup>55</sup>One might wonder, then, how we can distinguish theoretical predicates from observational predicates. This may well be a problem for Worrall, but one response would be to identify the theoretical predicates as the ones that are *not* observational.

<sup>56</sup>It should be noted that Votsis is not a defender of Worrall-style ESR, but defends his own Russellian variant of ESR. See Votsis (2005). Here, I am only concerned with Votsis’ views on Worrall’s ESR.

the theory. For Worrall, it's only apparent, since he holds that theoretical predicates, as second-order ambiguous names, can't hold any interpretation. In what follows, I will attempt to remain as noncommittal as possible regarding this issue, and by 'nature,' I mean either Votsis' *actual* non-structural theoretical content or Worrall's *apparent* non-structural theoretical content.

To sum up, Worrall's ESR is the view that the Ramsey sentence of a theory exhausts the cognitive content of that theory. The Ramsey sentence has structural theoretical content, in addition to empirical content. And it has no non-structural theoretical content (i.e., nature). At this point, we should have a general idea of what Worrall's ESR is supposed to be. In order to see what kind of view of hypothetical entities a proponent of ESR is committed to, we'll have to examine Worrall's views on the semantics of scientific language, and so I now turn to this topic.

### **The Semantics of ESR**

One of the features that distinguishes ESR from the scientific realism discussed above is the difference in semantic commitments. According to Worrall, one reason that ESR is different from realism is that

there is no reason why the way in which a theory mirrors reality should be the usual term-by-term mapping described by traditional semantics. Indeed, . . . if we are talking about an epistemically accessible notion then it cannot be! [ESR] in fact takes it that the mathematical structure of a theory may globally reflect reality without each of its components necessarily referring to a separate

item of that reality. (Worrall (2007), 154)

My goal at this point is to spell out what Worrall has in mind here by examining what he has to say about how ESR differs from realism and antirealism.

To begin with, there is ESR and realism. Although Worrall labels ESR “a version of scientific realism” (Worrall (2011), 171), elsewhere he argues that it is distinct from realism. He borrows Putnam’s characterization of the latter’s semantic commitments, according to which

the realist assert[s] that 1) Theoretical terms in mature science refer; and 2) what our currently accepted theories in mature science say using those terms is at any rate approximately true. (Worrall (2007), 153)

The realist, according to Worrall, admits term-by-term reference, even in the absence of anything close to complete and accurate knowledge of the entity in question. So, to take an example, the term ‘electron’ may refer, for the realist, even if our theory of electrons is at most approximately true. This is to say that our theory of electrons is incomplete and, strictly speaking, false. The realist, then, admits that, while our theories may be false, our terms do not fail to refer. In that case, the realist admits a way in which terms can refer that does not depend on the theories in which those terms occur. This, after all, was the innovation that came with the development of the causal theory of reference, of which Worrall is so critical.<sup>57</sup>

---

<sup>57</sup>It’s worth noting that Worrall seems to saddle the realist with a commitment to the causal theory of reference for theoretical predicates. But this is, at the very least, questionable. Lewis (1970), for example, defends a realist view in terms of a descriptivist theory of reference. Moreover, many realists have been influenced by the causal descriptivism developed in Lewis (1984), and have defended a variety of hybrid views. See, for example, Kitcher (1993) and Psillos (1999).

Roman Frigg and Votsis accuse Worrall of not being very clear about what we should put in the place of the traditional semantics of realism so as to make sense of this global reflection of reality that ESR is committed to (Frigg and Votsis (2011), 254). They claim that this global reflection of reality seems impossible without some kind of term-by-term correspondence, which Worrall claims to reject. But Worrall's rejection of term-by-term reference amounts to the claim that terms do not refer in isolation from the theory in which they occur. Instead, the reference of a term depends on the theory in which it occurs. Since theoretical predicates function as second-order ambiguous names, they refer to whatever it is that satisfies the Ramsey sentence. This is the global reflection of reality that he has in mind, and Worrall's ESR therefore involves a kind of global descriptivism which he contrasts with the causal theory of reference. The replacement for traditional semantics, then, is Worrall's global descriptivism.

Now that we have a better understanding of Worrall's semantic commitments, we can now see how ESR differs from realism. To begin with, recall  $V$  and  $R(V)$ . Worrall holds that  $V$  and  $R(V)$  do not differ in cognitive content. The realist will presumably object to this, on the grounds that the predicate variables in  $R(V)$  may not refer to what  $P_1$  and  $P_2$  refer to, respectively. This is because the realist assigns the above interpretations to  $P_1$  and  $P_2$ , which get assigned independently of the details of  $V$ . This is in accordance with the causal theory of reference. But according to Worrall,  $P_1$  and  $P_2$  are second-order ambiguous names. And so, they refer to whatever properties make the corresponding Ramsey sentence true. The content of  $P_1$  and  $P_2$  is exhausted by the logical relations that they bear to other uninterpreted theoretical predicates and interpreted observational predicates. And this is in

accordance with Worrall's descriptivism.

We can put the same point in terms of natural language, as opposed to the formalism of second-order logic. For the realist, the theoretical predicate 'x is a radium atom' has content apart from the logical relations that it bears to other predicates. Worrall, on the other hand, would treat 'x is a radium atom' as something that plays the role of a second-order ambiguous name. The latter has content that is exhausted by its logical relations to other predicates.<sup>58</sup>

But if Worrall's ESR amounts to the claim that content is exhausted by the Ramsey sentence of a theory, which gives the logical relations that obtain among uninterpreted theoretical predicates and interpreted observational predicates, how is it that ESR is not an antirealist view? In short, Worrall holds that ESR is a form of realism because of its commitments regarding the theoretical level. According to ESR, our theories get at the structure of the unobservable reality, and this goes beyond antirealist positions. Worrall gives at least two reasons why ESR is not an antirealist position, both of which involve his view that the Ramsey sentence gives the full cognitive content of a theory. First of all, Worrall points out that "some sentences expressed in purely observational vocabulary are theoretical" (Worrall (2011), 170). One example is "the assertion that there are unobservable objects—i.e. objects with no (directly) observable property" (Worrall (2011), 167). Worrall thinks that the Ramsey sentence of a theory is another example. Secondly, and relatedly, Worrall points out that the Ramsey sentence may eliminate theoretical terms, but "to stand ready to assert a sentence that quantifies over theoretical terms involves asserting

---

<sup>58</sup>Compare this example with what Worrall has to say about the reference of 'gluon' (Worrall (2011), 169), quoted above.

(not denying) their existence” (Worrall (2011), 170).

Before going further, it will be necessary to mark two terminological issues. First of all, Worrall writes of both theoretical predicates and theoretical terms in a way that suggests that these are the same. Predicates and terms are, of course, distinguished in systems of logic. But it’s clear from Worrall’s usage that he represents theoretical *terms* (understood in the broader, non-logical sense) as theoretical *predicates* (understood in the logical sense). Now that I’ve marked this issue, I’ll frame ESR’s view of semantics in terms of theoretical terms and theoretical assertions, as I have done for realism and constructive empiricism.

Secondly, Worrall’s theoretical predicates are contrasted with observational predicates, whereas my theoretical terms include terms used to name both observable and unobservable entities, and are contrasted with non-theoretical terms. Hence, since Worrall’s theoretical and observational predicates both occur in theories, they are subspecies of theoretical terms, as I understand the latter. I’ll be explicit about this when it matters.

At this point, we can see that ESR does not differ much from the realist idea of realistic interpretation, nor from the constructive empiricist idea of literal construal. Indeed, ESR is committed to realistic interpretation. Recall that the realist holds

(RI<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable of being true or false.

As we’ve seen, the most important theoretical assertion for Worrall is the Ramsey sentence of a theory. We’ve also seen that Worrall claims that the Ramsey sentence of a mature theory globally reflects reality. So the Ramsey sentence is an assertion about reality. And

Worrall puts a lot of work into showing that it can be true for non-trivial reasons. This is because of M. H. A. Newman's objection to Russell's structural realism (Newman (1928), Russell (1927)).<sup>59</sup> This objection entails that the Ramsey sentence of any theory only amounts to a cardinality constraint, and can be made true provided that there are sufficiently many objects in the universe. Worrall has argued that his ESR is immune to Newman's objection because the Ramsey sentence does not quantify over the observational predicates, and this is enough to ensure that it amounts to more than a mere cardinality constraint, in which case it can be true for non-trivial reasons (Worrall and Zahar (2001); Worrall (2007), 147–154; Worrall (2011), 168–169). Moreover, the claim that our best theories approximate the structure of reality just amounts to the claim that their Ramsey sentences are true. Whether or not Worrall's response to the Newman objection is successful, this is enough to show that he is committed to (RI<sub>2</sub>).

Worrall is also committed to the other half of realistic interpretation, namely (RI<sub>1</sub>) Theoretical terms are putatively referring expressions that have putative reference to empirical entities.

If these theoretical terms are observational predicates, then it's easy to see that Worrall accepts (RI<sub>1</sub>), since he interprets such predicates in the same way that realists do. But if these theoretical terms are theoretical predicates, in Worrall's sense, then Worrall differs from the realist. Theoretical predicates for Worrall play the role of second-order ambiguous names, and their content is exhausted by their logical relations to other predicates. But it's clear that these predicates are still understood as putatively referring expressions. Indeed,

---

<sup>59</sup>See Demopoulos and Friedman (1985) for a more recent examination of Newman's objection.

Worrall claims that “a gluon is a ‘whatever it is’ that structures the phenomena in certain complex ways,” and that “primitive theoretical predicates in the initial un-Ramseyfied theory name (or attempt to name) theoretical entities” (Worrall (2011), 169). So while Worrall and the realist may understand the content of a theoretical term like ‘gluon’ in different ways, and may have correspondingly different ways of understanding how a theoretical term like ‘gluon’ can refer, both accept (RI<sub>1</sub>).

Perhaps it shouldn’t be all that surprising that ESR is committed to (RI<sub>1</sub>) and (RI<sub>2</sub>), given that Worrall characterizes ESR as a form of realism. In any case, now that we have ESR’s semantic commitments on the table, we can extract its view of hypothetical entities and HE terms.

### **ESR and Hypothetical Entities**

Given that ESR is committed to (RI<sub>1</sub>), it is also committed to R<sub>HE</sub> HE terms are putatively referring expressions that have putative reference to empirical entities.

This follows straightforwardly from the fact that HE terms are a subset of theoretical terms.

One might think that ESR is thereby committed to (R<sub>T1</sub>)–(R<sub>T3</sub>). But matters are a bit more complicated. As it turns out, it’s difficult to see how any theoretical term that occurs in a theory in a mature science could fail to refer if one accepts ESR. I’ll now attempt to reconstruct why this is the case.

To begin with, Worrall restricts his ESR to branches of science that have reached ‘maturity.’ He complains that realists often leave this term undefined, and so he proposes an

account of maturity according to which a science is mature once it has enjoyed what is often called novel predictive success. That is to say, the science that has reached maturity has theories that are “predictive of general *types* of phenomena, without these phenomena having been ‘written into’ the theory” (Worrall (1989), 114). Importantly, this notion of maturity applies to branches of science, as opposed to theories. To take an example, physics was mature in the early years of the twentieth century, even though the general theory of relativity had not yet experienced any novel predictive success. Any such success that physical theories enjoyed in the eighteenth and nineteenth centuries is sufficient to establish the maturity of physics before the twentieth century.

Any science that has reached maturity therefore has theories that enjoy the kind of success that underwrites the no-miracles argument. Novel predictive success is really the kind of success that calls out for explanation. It’s easy enough to explain why a prediction came true if that prediction was written into the theory in the first place. But novel predictions that come true, like the prediction of the illuminated spot in the case of Fresnel’s wave theory, are predictions that call out for explanation. Worrall’s claim is that theories that enjoy this kind of success do so because they approximate the structure of reality, in which case such success is not miraculous, but readily intelligible.

On Worrall’s view the Ramsey sentence of a theory captures the structure of reality. If the science in question is mature, then we can say that the Ramsey sentence of one of its theories is approximately structurally correct. In this case, the equations of this theory will be retained as limiting cases in any successor theory, and will therefore look approximately structurally correct in light of those successor theories (Worrall (2007), 142–143).

Worrall's examples here include the transition from Fresnel's theory to Maxwell's, and also the transition from Newtonian mechanics to relativistic mechanics (Worrall (1989), 109). So what does it mean for the Ramsey sentence to capture the structure of reality? Since Worrall rejects approximate truth (Worrall (2007), 154), the only thing that seems left is that the Ramsey sentence is true.

In any case, we can see that theoretical terms must refer, even without being totally clear on what it means for the Ramsey sentence to globally reflect reality. When it comes to the observational predicates, this is relatively straightforward, since Worrall takes it that these refer. The interesting case is that of theoretical predicates. We can see the point in terms of the transition from Fresnel's theory to Maxwell's. Worrall claims that both were "quite right not just about a whole range of optical phenomena but right that these phenomena depend on something or other that undergoes periodic change at right angles to the light" (Worrall (1989), 120). Hence they both refer to "something or other" that undergoes this periodic change. To be sure, both theories misidentify the nature of light. But they refer to something in virtue of getting the structure right.<sup>60</sup> Indeed, these theories could not have gotten the structure right without referring to the "something or other" that undergoes the periodic vibration, as the above quotation from Worrall makes clear. It just so happens that we cannot know whether that "something or other" is the ether, or a field strength, or something else.

---

<sup>60</sup>Worrall is clear that, on his view, Fresnel was not referring to the electromagnetic field (Worrall (1989), 116–117). I am not claiming otherwise—I merely claim that, on Worrall's view, Fresnel referred to *something*.

Worrall discusses two other examples. As we've seen, he claims that "a gluon is a 'whatever it is' that structures the phenomena in certain complex ways through specific intricate relationships with the phenomena and with other, similarly characterised, theoretical notions" (Worrall (2011), 169). He also claims that "electrons (if they exist at all) are *whatever it is* that satisfy our current relevant theories" (Worrall (2007), 148). In these examples, 'electron' and 'gluon' function as second-order ambiguous names. One way in which these names might fail to refer is if the relevant Ramsey sentences fail to reflect the structure of reality. But Worrall has to admit that they do reflect the structure of reality, or else he cannot accommodate the no-miracles argument. And if they do, then 'electron' and 'gluon' automatically refer to whatever it is that makes the relevant Ramsey sentences true. The referents here may not be the electron and gluon as the realist understands these entities, since the realist will accept the theoretical interpretations of the terms that the proponent of ESR will strip away by means of Ramseyfication. Presumably, this is what Worrall has in mind when he questions whether electrons "exist at all." The electron of the realist may or may not exist, but the electron of the proponent of ESR must exist, so long as the theory in question is (approximately) structurally correct.

Since HE terms are a subset of theoretical terms, and if we restrict ourselves to the mature sciences, the proponent of ESR holds

(R<sub>T1</sub>) an HE term, so understood, refers to an empirical entity

When the term in question is an observational predicate, the proponent of ESR is in complete agreement with the realist. But when the term in question is a theoretical predicate,

the proponent of ESR and the realist differ. In that case, the only way in which we know that entity is in terms of the structure that the Ramsey sentence gives us. We know the entity structurally, but we do not know the nature of the entity. Hence, the proponent of ESR is committed to  $(R_{T1})$ , and importantly, not to either  $(R_{T2})$  or  $(R_{T3})$ .

### 1.5.2 ONTIC STRUCTURAL REALISM

As mentioned above, Ladyman (1998) was the first to note that ESR is not the only position available to the structural realist, and that ontic structural realism (OSR) is another possibility, which he goes on to defend. OSR, crudely put, is “the claim that there are no ‘things’ and that structure is all there is” (Ladyman (2009), §4). The crudeness of this statement of OSR is revealed by Ladyman and Don Ross’ qualification that “there are objects in our metaphysics but they have been purged of their intrinsic natures, identity, and individuality, and they are not metaphysically fundamental” (Ladyman and Ross (2007), 131). I will now turn to explaining the OSR that Ladyman and his collaborators have defended. I’ll begin with an account of how they construe structure in terms of the semantic view of theories. I’ll then turn to their semantic commitments, and I’ll conclude by extracting their view of hypothetical entities and HE terms.

#### **Structure and the Semantic View**

While Worrall develops his ESR in terms of Ramsey sentences, and therefore within the so-called syntactic view of theories, Ladyman et al. follow van Fraassen, among others, and situate their OSR within the semantic view of theories. The notion of structure that

they offer is therefore best understood in terms of that view. To begin with, recall that, according to the semantic view, a theory is a class of models. We can identify a theoretical model  $T$  with a mathematical structure  $\langle D, R_i^n \rangle_{i \in I}$ , where  $D$  is a set of objects and  $R_i^n$  is a family of  $n$ -place relations holding among the elements of  $D$ . Ladyman et al. go on to extend the semantic view in terms of what they call *partial structures*. Partial structures are mathematical structures in which the relations in question are not necessarily defined for all  $n$ -tuples. Such relations are ‘partial relations,’ which Bueno et al. explain as follows:

each partial [ $n$ -place] relation  $R$  can be viewed as an ordered triple  $\langle R_1, R_2, R_3 \rangle$ , where  $R_1, R_2$ , and  $R_3$  are mutually disjoint sets, with  $R_1 \cup R_2 \cup R_3 = D^n$ , and such that:  $R_1$  is the set of  $n$ -tuples that belong to  $R$ ,  $R_2$  is the set of  $n$ -tuples that do not belong to  $R$ , and  $R_3$  is the set of  $n$ -tuples for which it is not defined whether or not they belong to  $R$ . (Bueno, French, and Ladyman (2002), 498)

The partial structures approach also involves the application of partial morphisms, for example, isomorphisms and homomorphisms.<sup>61</sup> Steven French and Ladyman claim that “the partial structures form of the semantic approach offers a general account of theoretical structure . . . and thus represents an appropriate formal framework for [OSR]” (French and Ladyman (2003), 34).

What, then, does structure amount to for the proponent of OSR? In order to answer this question, French and Ladyman draw from the work of Ernst Cassirer (1956):

What is an electron then? Not, Cassirer insists, an individual object (*ibid.*,

---

<sup>61</sup>For more on this approach, see Bueno (1997), Bueno (1999), and French and Ladyman (1999).

180). If we want to continue to talk, in everyday language, about electrons as objects—because we lack the logico-linguistic resources to do otherwise (we shall return to this point below)—then we can do so ‘only indirectly’, ‘... not insofar as they themselves, as individuals, are given, but so far as they are describable as “points of intersection” of certain relations’ (ibid.). (French and Ladyman (2003), 39)

Ordinarily, individual objects are taken as fundamental, and relations are constructed out of ordered  $n$ -tuples of those objects. French and Ladyman, echoing Cassirer, suggest an inversion of this idea. Relations, for them, are taken as fundamental, and their relata, namely, the objects, are to be thought of as constructed out of those relations. This is one way to understand the idea that “structure is all there is.” Another way to understand the idea (which French and Ladyman also accept, and is not in conflict with the first way), “is that the relata of a given relation always turn out to be relational structures themselves on further analysis” (Ladyman and Ross (2007), 155). The basic idea, then, is that the structure is the relational structure of a theoretical model. The objects instantiating this structure are not fundamental, and there is nothing to them over-and-above the structure that they instantiate. Hence, OSR differs from ESR. For the latter, we can only know an entity structurally—we do not know the nature of the entity. For the former, there is nothing non-structural to know.

To be sure, the idea of relations as fundamental and relata as secondary is at the very least difficult to understand. And some commentators have claimed that the idea simply

does not make sense.<sup>62</sup> Indeed, one may not be able to help but think that, if all *relata* turn out to be relations themselves, that OSR leads to an infinite regress. In response, Ladyman and Ross speculate that human psychology may require us to think of relations as involving *relata* that are not themselves relations, but they deny that this entails that such *relata* are more fundamental than relations (Ladyman and Ross (2007), 154–155). Moreover, French and Ladyman suggest that this impression of paradox is a result of the fact that the mathematical apparatus they use to characterize structure, namely, set theory and group theory, are themselves built up from individual objects. French and Ladyman list it as a goal to develop a more structuralist mathematical apparatus, so as to avoid this impression (French and Ladyman (2003), 52).

Up to this point, I have discussed structure in terms of what Ladyman and Ross call “discourse in the formal mode”—such discourse “is always to be understood relative to a background of representational conventions” (Ladyman and Ross (2007), 119). This is the discourse of partial structures. “Discourse in the material mode,” according to Ladyman and Ross, “purports to refer directly to properties of the world” (Ladyman and Ross (2007), 119). We’ve seen what structure amounts to in the formal mode. What remains to be seen is what it amounts to in the material mode.

In the material mode, the claim that ‘structure is all there is’ amounts to the claim that “[t]o be is to be a real pattern” (Ladyman and Ross (2007), 233). Ladyman and Ross borrow the idea of a real pattern from Dennett (1991), and develop it further so as to base their theory of ontology on it. Their precise definition of ‘real pattern’ is quite technical,

---

<sup>62</sup>Van Fraassen, for example, has this to say: “This point is not just striking, it is paradoxical. Indeed, as far as my exposition here goes, it comes to us in barely coherent guise” (van Fraassen (2006), 293).

and takes some work to explain. But the details needn't concern us here, and the following suffices to give the general idea. To begin with, a pattern, for Ladyman and Ross, "is just any relations among data" (Ladyman and Ross (2007), 228). They move away from the traditional conceptions of things, events, and processes, and claim that these are all, in fact, real patterns (Ladyman and Ross (2007), 121). The examples of real patterns that they give include "[p]rices, neurons, peptides, gold, and Napoleon," in addition to "quarks, bosons, and the weak force" (Ladyman and Ross (2007), 300). To take another example, "particles or spacetime points are just patterns that behave like particles or spacetime points respectively" (Ladyman and Ross (2007), 178). Real patterns, to be sure, are relations among data; but they aren't *just any* relations among data. Ladyman and Ross distinguish real patterns from what they call "mere patterns" (Ladyman and Ross (2007), 228). As a rough gloss on this distinction, we can say that real patterns are particularly significant relations among data, in the sense that they carry information that other patterns do not. Mere patterns, on the other hand, are the relations among data that do not carry this information, or else carry it only because they include the relations among data that constitute a real pattern. In other words, mere patterns are "informationally redundant" (Ladyman and Ross (2007), 279). The important point is that Ladyman and Ross intend to use this distinction to separate the entities that exist from those that do not.

Putting all of this together, Ladyman and Ross claim that the structure of a theoretical model in the formal mode represents real patterns in the material mode (Ladyman and Ross (2007), 120). But they also claim that it is possible "to refer to particular real patterns before specifying the structures that represent them" (Ladyman and Ross (2007), 121), and

in order to do this, they introduce the terminology of “locators,” which they explain as follows:

One picks out a real pattern independently of its structural description by an ostensive operation—that is, by ‘pointing at it’. This is intended as evocative talk for operations of fixing, stabilizing, and maintaining salience of some data from one measurement operation to another. (So, think of ‘pointing’ as meaning ‘directing a measurement apparatus’.) In the fully generalized sense, this means that one indicates the real pattern’s *location* in some coordinate system with high enough dimensionality to permit its disambiguation from other real patterns. . . . We will thus speak of all of the traditional kinds of objects of reference—objects, events, processes—by mentioning their *locators*. A locator is to be understood as an act of ‘tagging’ against an established address system. Such address systems can be formal representations (for example, a map or a spatio-temporal coordinate system) . . . (Ladyman and Ross (2007), 121)

Hence, it’s possible to refer to real patterns even if one hasn’t specified the structures that represent them. One does so by means of a locator, and the locator picks out the real pattern by specifying the location at which measurements of that pattern should be taken. The locator, then, “denotes . . . an element of some already partially elaborated structure plus measurement instructions” (Ladyman and Ross (2007), 122). So, to take an example, Ladyman and Ross use ‘aardvark(L)’ to symbolize the locator for aardvarks, and they em-

phasize that “this should not be read as purporting to refer to collections of substantival, pattern-independent individual aardvarks,” but instead to a real pattern (Ladyman and Ross (2007), 122). And in general, “[a]s soon as [some mathematical structure] *S* is minimally physically interpreted by a function that maps some of its elements onto real patterns then those elements of *S* are locators” (Ladyman and Ross (2007), 221). Ladyman and Ross claim that, in their view, real patterns play the traditional metaphysical role of individuals, and locators play the traditional logical role of sortals—just as some sortals successfully pick out individuals, “[s]ome locators successfully pick out real patterns, and these are the objects of genuine existential quantification” (Ladyman and Ross (2007), 239).

Ladyman and Ross also claim that their OSR “ought to be understood as modal structural empiricism” (Ladyman and Ross (2007), 99), and so before turning to the semantics of OSR, I will discuss what they mean by this. To begin with, by “structural empiricism,” they have in mind van Fraassen’s empiricist structuralism as discussed above. Their position goes beyond van Fraassen’s insofar as they commit themselves to the objective reality of modal structure, which they explain as follows:<sup>63</sup>

By modal structure we mean the relationships among phenomena (tracked or located . . . as things, properties, events, and processes) that pertain to necessity, possibility, potentiality, and probability. (Ladyman and Ross (2007), 153–154)

The logical/mathematical modality of a theoretical model in the formal mode, then, represents this objective modal structure in the material mode (Ladyman and Ross (2007), 120).

---

<sup>63</sup>To be sure, as discussed above, Monton and van Fraassen admit that a constructive empiricist *can* be a modal realist (Monton and van Fraassen (2003), 419–421). The point here is just that a proponent of OSR *must* be a modal realist.

And importantly, Ladyman and Ross draw a distinction between structure and non-structure by appealing to modality, as when they claim that “[m]erely listing relations among locators does not state anything with modal force. Therefore, it doesn’t specify structure in our sense and it isn’t yet scientific theory as we’ve defended it” (Ladyman and Ross (2007), 158). Hence structure must have modal force, something that van Fraassen denies. The similarities and differences between OSR and van Fraassen’s empiricist structuralism will be important when it comes to explaining the semantics of OSR, and it is to this topic that I now turn.

### **The Semantics of OSR**

As we’ve seen, the semantic commitments of realism, constructive empiricism, and ESR all have to do with the relationship between linguistic entities (i.e., terms and assertions) and the world. In this section, I will discuss three closely related reasons for thinking that OSR’s semantic commitments, though similar in some ways, differ importantly from what we have seen so far. These three reasons are Ladyman et al.’s rejection of standard referential semantics, commitment to the semantic view of theories, and commitment to an ontology of structure.

To begin with, there is Ladyman et al.’s rejection of standard referential semantics. Ladyman claims that his OSR “depart[s] from standard scientific realism in rejecting term by term reference of theories, and hence standard referential semantics, and any account of approximate truth based on it” (Ladyman (2011), 97). We can see the same rejection of standard referential semantics in Ladyman and Ross, who also indicate a positive proposal

for what to put in its place:

OSR allows for a global relation between models and the world, which can support the predictive success of theories, but which does not supervene on the successful reference of theoretical terms to individual entities, or the truth of sentences involving them. (Ladyman and Ross (2007), 132)

As Ladyman recognizes (Ladyman (2011), 97), OSR is similar to Worrall's ESR in the sense that both admit a global relation between a theory and the world, while rejecting some aspects of standard referential semantics. I suggested above that Worrall only rejects certain aspects of standard semantics, and one reason for this is that, insofar as he operates with Ramsey sentences, he is still operating with an essentially linguistic notion of what a theory is. But Ladyman et al.'s rejection of standard semantics may be more thoroughgoing, insofar as he operates with the semantic view of theories, according to which theories are "extralinguistic entities" (Suppe (1977), 221).

We can therefore get clear on why Ladyman et al. reject standard referential semantics by examining their commitment to the semantic view of theories. As we have seen, van Fraassen's commitment to the semantic view does not preclude him from construing the language of science literally, and thereby adopting standard semantics. And, insofar as Ladyman and Ross identify OSR with modal structural empiricism, one might think that they thereby commit themselves to van Fraassen's notion of literal construal as discussed above. Moreover, they applaud van Fraassen for not falling into the error of employing a verificationist theory of meaning, as the positivists did (Ladyman and Ross (2007), 63).

But Ladyman and Ross go on to contrast van Fraassen's views with those of Ronald Giere (1985):

How should we understand questions about the relationship between theoretical objects and the world in terms of the semantic conception of theories? Giere addresses this issue in his writings on the semantic approach. Unlike van Fraassen, he seems to accept that the semantic approach transforms the terms of the scientific realism debate. In particular, Giere (1985) argues that it is unreasonable to expect all theoretical representation in science to fit the mould in which philosophers cast linguistic representation; Tarskian semantics is not appropriate for a consideration of the representative role of the tensor calculus and differential manifolds or Hilbert space. (Ladyman and Ross (2007), 117).

Ladyman and Ross go on to write: "Giere claims, and we agree, that once the semantic approach is adopted the crucial issue is whether or not theoretical models tell us about modalities" (Ladyman and Ross (2007), 118). If they identify with Giere, as opposed to van Fraassen, on all of these points, this would render their rejection of standard semantics intelligible. The primary issue for Giere, and for Ladyman and Ross, is the relationship between a theoretical model and the world, and according to them, standard semantics is not the way to understand this relationship.

This passage leaves it open that Ladyman and Ross may accept standard semantics for terms and sentences, but not for models. However, their view is that we should reject standard semantics even for terms and sentences, and this is revealed by their commitment

to an ontology of structure, as the following quotation illustrates:

We share Horgan’s and Potrc’s suspicion of the idea that there is a direct correspondence between sentences such as ‘the book is on the table’ and the world, if what is meant by that is the claim that there are individuals that act as the referents of the singular terms ‘book’ and ‘table’. So with them we conclude that we need a different semantics from that of direct reference and correspondence to explain how everyday utterances can be true despite there being no self-subsistent individuals. (Ladyman and Ross (2007), 254).

Moreover, they claim to “endorse a version of instrumentalism about all propositions referring to self-subsistent individual objects, chairs and electrons alike” (Ladyman and Ross (2007), 198). This instrumentalism is a consequence of their view that there are no self-subsistent individuals—only real patterns. The important point for now is just that Ladyman and Ross reject standard semantics across the board—for terms, sentences, and theoretical models.

But even if Ladyman et al.’s rejection of standard semantics is more intelligible at this point, it’s still not entirely clear what their positive proposal is. I take it that the positive proposal is, to some extent, a work in progress and an issue for further research. Nonetheless, Ladyman et al. do give us enough to go on to understand the basics of their semantic commitments.

To begin with, Ladyman et al. are committed to

(RI<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable

of being true or false.

Ladyman and Ross indicate that they endorse an account of realism due to Andrew Melnyk, according to which “realists . . . regard (certain) current scientific hypotheses” as “true or false in virtue of the way the mind-independent world is” (Melnyk (2003), 229; quoted in Ladyman and Ross (2007), 305). The reason for the restriction to certain current hypotheses has to do with further conditions that Melnyk places on what it means to be a realist, and the excerpt I’ve quoted is best read as applying to *all* scientific hypotheses. In that case, Ladyman and Ross accept that scientific hypotheses are truth-conditioned assertions about reality that are capable of being true or false. To be sure, scientific hypotheses arguably form a proper subset of theoretical assertions. But it would be odd for Ladyman and Ross to take some other view of these other assertions. Moreover, they claim that “[t]heories, like Newtonian mechanics, can be literally false,” in which case they can presumably be true as well (Ladyman and Ross (2007), 118). This latter claim is enough to extend what Melnyk says about hypotheses to theoretical assertions more generally.

One might object that Ladyman et al. cannot accept (RI<sub>2</sub>), in light of the fact that assertions are linguistic entities, and their concerns are with the relationships between theoretical models (i.e., extralinguistic entities) and the world. But this objection does not go through. Ladyman et al. operate with the idea that theoretical models are mathematical structures. And within this framework, it makes sense to speak of a sentence  $\alpha$  being true in a structure  $T$ . Indeed, within the partial structures approach, French and Ladyman employ a notion of pragmatic truth that depends on the idea that a sentence can be true in a total (i.e., non-partial) structure (French and Ladyman (1999), 105). A sentence  $\alpha$  is pragmatically true in

a partial structure  $S$  if  $\alpha$  is true in a total structure  $R$ , where  $R$  extends the partial relations of  $S$  into normal relations, and hence extends  $S$  into a normal structure.<sup>64</sup> To be sure, French and Ladyman still claim that, on the partial structures version of the semantic view, theories are extralinguistic entities (French and Ladyman (1999), 118), and I won't take issue with this here. The point for now is just that the present objection does not go through, because we can take the sentences that are true (or pragmatically true) in some structure (or partial structure) as the theoretical assertions referenced in (RI<sub>2</sub>). Moreover, all of this is consistent with Ladyman et al.'s view that the crucial issue in the realism debate has to do with the relationship between models and the world. This is because they can presumably still claim that the semantic relation between a structure  $T$  and the world needn't supervene on the *actual* truth of  $\alpha$  (i.e., not just its truth *in the structure*). And it needn't supervene on the successful reference of theoretical terms within  $\alpha$  to individual entities.

On the issue of the semantics of theoretical terms, we've already seen that proponents of realism, constructive empiricism, and ESR all have something to say, and that all three positions are roughly in agreement. Examination of this issue shows the extent to which OSR differs from these other positions with regard to its semantic commitments. Recall that, for Ladyman et al., the crucial issue in the realism debate has to do with the relationship between models and the world. Specifically, instead of the relation between sortals and individuals, they are concerned with the relation between locators and real patterns. We've seen that locators are elements of a partially elaborated structure, plus measurement instructions that give information about where to direct an apparatus to measure a pur-

---

<sup>64</sup>For the details of the definition of pragmatic truth, see da Costa and French (2003), 18–19.

ported real pattern. As such, locators are extralinguistic entities, and presumably shouldn't be identified with theoretical terms. Moreover, Ladyman explicitly rejects "term by term reference of theories" (Ladyman (2011), 97). Hence, the proponent of OSR cannot accept (RI<sub>1</sub>) or (LC<sub>1</sub>), since both deal with the semantics of terms. I take it that once could potentially bridge the gap between the semantics of locators (e.g., 'aardvark(L)') and terms (e.g., 'aardvark') by claiming that the term refers iff the locator picks out a real pattern. One could then develop a new semantics on the model of standard semantics, where terms refer to real patterns, as opposed to individuals. But Ladyman et al. indicate that their break with standard semantics is more drastic. In this case, an accurate characterization of their semantic commitments requires us to frame these commitments in terms of locators, as opposed to theoretical terms.

The proponent of OSR is therefore committed to

(OSR<sub>1</sub>) Locators are putatively referring expressions that have putative reference to real patterns.

Three things are worth commenting on here. First of all, one might take issue with the language of reference in (OSR<sub>1</sub>), based on Ladyman et al.'s rejection of standard referential semantics. But Ladyman and Ross initially introduce locators "[i]n order to be able to *refer* to particular real patterns" (Ladyman and Ross (2007), 121; my emphasis). Hence, this language is appropriate. And it is consistent with Ladyman et al.'s rejection of standard referential semantics since, even if they reject *standard* referential semantics, the semantics that they put in its place would seem to have to be referential in some sense.

Secondly, it's worth clarifying why it is that Ladyman et al. are committed to the claim that locators are putatively referring expressions, since there may seem to be a potential problem with attributing this claim to them. As we've seen, Ladyman and Ross claim to endorse a kind of instrumentalism about propositions that purportedly refer to self-subsistent individuals. They don't go into the details of what this instrumentalism amounts to, but presumably it amounts to some kind of translation from a proposition that purportedly refers to individuals, to a proposition that purportedly refers to real patterns. This would form the basic idea behind a "different semantics" that would "explain how everyday utterances can be true despite there being no self-subsistent individuals" (Ladyman and Ross (2007), 254). The fact that Ladyman and Ross label their view a form of instrumentalism may lead us to believe that the gulf between OSR and the other positions we've examined is very wide indeed, since all of the proponents of the positions discussed so far attempt to distance themselves from instrumentalism and positivism.

Although OSR is different from the other positions, the gulf is not as wide as one might initially think. And it's certainly not so wide that we'd have to group OSR with the positivism of Frank, who understands a theoretical term as "an abbreviation for the whole group of experiences with the symbol system belonging to it" (Frank (1941), 85). After all, when Ladyman uses, for example, the term 'spacetime points,' he's referring to patterns that behave like spacetime points (Ladyman and Ross (2007), 178). When van Fraassen employs literal construal to distinguish his view from positivism and so-called instrumentalism, he doesn't have in mind the kind of instrumentalism that Ladyman and Ross endorse. According to van Fraassen, "although a literal construal can elaborate, it cannot change logical

relationships,” and “[i]t is possible to elaborate, for instance, by identifying what the terms designate” (van Fraassen (1980), 10). Ladyman and Ross are elaborating, and not changing logical relationships, and they are using putatively referring expressions. In this sense, they agree with the constructive empiricist, and with the proponents of the other positions I’ve discussed. The difference here is that, in OSR, the expressions are locators, while in the other positions, they are terms.

Thirdly, while proponents of the previously examined positions are committed to putative reference to *entities*, proponents of OSR are committed to putative reference to *real patterns*. It’s worth clarifying how this view fits with the previously examined ones. Recall that I use ‘entity’ in a broad sense, which covers individuals, objects, substances, events, processes, properties, and relations. In this case, the semantic commitments of OSR may seem more restrictive than those of the other positions, since OSR only admits structure as ontologically fundamental. But, of course, the proponent of OSR will have something to say about what are traditionally taken to be objects, substances, events, and so on—they are just real patterns that behave like objects, substances, and events. Upon closer examination, then, OSR’s semantic commitments are not more restrictive than those of the other positions discussed. But there is an important sense in which OSR differs from the other positions, and this has to do with the influence of OSR’s metaphysical commitments on its semantic commitments. Ladyman and Ross refuse to say whether structure is physical or mathematical—all that they will say is that “[t]he ‘world-structure’ just is and exists independently of us and we represent it mathematico-physically via our theories” (Ladyman and Ross (2007), 158). In support of their view, they acknowledge van Fraassen’s view that

it's not always clear whether our terms refer to concrete entities or mathematical entities (Ladyman and Ross (2007), 160). But they, in fact, go a step beyond van Fraassen, since he applies the distinction, while Ladyman and Ross do not.

Since OSR differs somewhat from the other positions we've examined, here, in summary, is the two-fold idea behind its semantic commitments:

(OSR<sub>1</sub>) Locators are putatively referring expressions that have putative reference to real patterns.

(OSR<sub>2</sub>) Theoretical assertions are truth-conditioned assertions about reality that are capable of being true or false.

### **OSR and Hypothetical Entities**

Now that we have OSR's semantic commitments in place, we can determine OSR's views of hypothetical entities and HE terms. But first, it will be necessary to introduce some terminology. Instead of hypothetical entities, proponents of OSR will have to speak of hypothetical patterns. Proponents of OSR will then understand the rough guide in the following way:

A hypothetical pattern is a new purported pattern that a scientist puts forward as a real pattern in advance of decisive empirical reasons to do so.

And since we've framed OSR's semantic commitments in terms of locators, as opposed to terms, instead of HE terms, we have *HP locators*, short for *hypothetical pattern locators*.

The first task, then, is to show that there is room in OSR for hypothetical patterns and HP locators. Although Ladyman and Ross do not use the terminology that I've introduced, it's clear that the ideas are already present in their characterization of OSR. To begin with, they explicitly mention the basic idea when they claim that “[s]pecial sciences are free to hypothesize any real patterns consistent with the measurements they accumulate so long as these do not contradict what physics agrees on” (Ladyman and Ross (2007), 252). Now recall that Ladyman and Ross initially introduce locators “[i]n order to be able to refer to particular real patterns before specifying the structures that represent them” (Ladyman and Ross (2007), 121). They go on to claim that “[n]o observer ever has access to the complete extent of a real pattern” (Ladyman and Ross (2007), 240). In that case, the structures that represent such real patterns have not been fully specified, and so it would seem that locators are always necessary in order to refer to real patterns. Moreover, they claim that

it [is] a necessary condition on a pattern's reality that it not be informationally redundant in the complete account of the universe. When actually formulating scientific theories short of this limit, we cannot know that our posited patterns would satisfy this criterion, but we care whether they do, and so we make our existence claims at least implicitly provisional. Furthermore, as science progresses we adjust our ontology in accordance with our concern for ontological parsimony. (Ladyman and Ross (2007), 279)

Without access to a fully specified structure, and to the complete account of the universe, it might seem that *all* patterns are hypothetical patterns.

But I take it that Ladyman and Ross' point here is basically the same as the point that our conclusions in the sciences are defeasible, and elsewhere, they admit some kind of distinction between hypothetical and real patterns. They show this when they “distinguish between the question of whether our concepts and intuitions track anything real, and the question of whether the real patterns they track have all the properties that their intuitively familiar surrogates appear to have” (Ladyman and Ross (2007), 254–255). The first question bears on hypothetical patterns, while the second bears on real patterns. The important point is that this distinction shows that Ladyman and Ross admit both kinds of case. OSR's understanding of the rough guide, then, would be that a pattern is hypothetical if we lack decisive empirical reasons to claim that the pattern is not informationally redundant, where such reasons are understood as being defeasible. Otherwise, we're dealing with a real pattern, though we may not have an accurate specification of its properties.

An HP locator, then, is just a locator used to name a hypothetical pattern. And since HP locators are a subset of locators, it follows, based on (OSR<sub>1</sub>), that proponents of OSR are committed to ‘ontic structural realism about hypothetical patterns,’ or OSR<sub>HP</sub> for short.

OSR<sub>HP</sub> HP locators are putatively referring expressions that have putative reference to real patterns.

And from this, it follows that proponents of OSR are committed to something like the realist trichotomy discussed above: either

(OSR<sub>T1</sub>) an HP locator, so understood, refers to a real pattern; or

(OSR<sub>T2</sub>) it altogether fails to refer to a real pattern; or

(OSR<sub>T3</sub>) it ‘kind of’ refers to that real pattern.

(OSR<sub>T1</sub>) and (OSR<sub>T2</sub>) follow straightforwardly from OSR<sub>HE</sub>. However, (OSR<sub>T2</sub>) is worth commenting on in more detail. Based on what Ladyman and Ross say, it’s unclear what happens when a locator fails to refer to a real pattern. One option is to treat such a locator as lacking a referent, and this makes sense, insofar as real patterns “are the objects of genuine existential quantification” (Ladyman and Ross (2007), 239). Another option is to treat such locators as referring to *mere* patterns, since, presumably, such locators pick out some kind of relations among the data, even if such relations lack the modal force requisite for structure in Ladyman and Ross’ sense. Because of this difficulty, (OSR<sub>T2</sub>) should be understood as encompassing both of these options. I include (OSR<sub>T3</sub>) because nothing that Ladyman et al. say rules it out. Moreover, there may be room within the partial structures approach to develop a notion of partial reference, just as we have a notion of pragmatic truth, and this would be an example of ‘kind of’ referring. For example, we might say that a locator  $x(L)$  in a partial structure  $T$  *partially refers* if there is a structure  $S$  which extends  $T$  into a total structure, and  $x(L)$  refers in  $S$ . But, to be sure, a proponent of OSR needn’t be committed to this, or to (OSR<sub>T3</sub>) in general, and the way in which I’ve set up this trichotomy makes that clear.

## 1.6 CONCLUSION

Based on my survey of the scientific realism debate, which has been far from exhaustive, we should have a clear grasp of the similarities and differences among the various

positions that I've discussed. One similarity is particularly important in what follows. This is the idea that terms used to introduce hypothetical entities into scientific discourse are putatively referring expressions that have putative reference to empirical entities. We've seen that, in one way or another, proponents of scientific realism, constructive empiricism, ESR, and OSR all hold to this idea. And importantly, this semantic commitment is one way in which post-positivist philosophy of science has attempted to distance itself from logical positivism. In my view, we needn't commit ourselves to this, and we can gain a better understanding of science if we reject this idea. My goal in the remainder of the dissertation is to argue for this claim.

## Chapter 2

---

### *Suppositionalism and Hypothetical Entities*

#### 2.1 INTRODUCTION

Supposition is a widespread phenomenon; language users suppose for a variety of reasons, and across many different domains of discourse. For example, a language user may suppose that  $\sqrt{2}$  is a rational number for the sake of a proof, or she may suppose that Superman has heat vision for the sake of engaging with a work of fiction.

The main argument of this dissertation is to establish that supposition gives us a better framework for understanding scientific discourse involving hypothetical entities than do the extant views discussed in chapter 1. My aim in this chapter is to begin that argument by motivating a suppositionalist view of hypothetical entities, according to which the latter are understood as objects of supposition. In order to do this, I will draw on Robert Rynasiewicz's work towards developing a unified theory of supposition in the arts and sciences. I will begin by introducing the phenomena of supposition in terms of some paradigm cases. I will then go on to clarify exactly what I mean by 'supposition' by situating my own view within the literature on the distinction between imagination and supposition. With this clarified notion of supposition in hand, I will turn to the details of Rynasiewicz's suppositionalism. After explaining the details as they apply to some of the paradigm cases

mentioned above, I will show how suppositionalism can be applied to scientific discourse concerning hypothetical entities, thereby leading to a view of such discourse that differs from the extant views.

The primary claim I wish to establish in this chapter is that we needn't accept the extant views of hypothetical entities because there is an alternative, namely, the suppositionalist view.<sup>1</sup> In the remaining chapters of the dissertation, I will draw on some episodes from late-eighteenth- and early-nineteenth-century chemistry in order to argue for two general claims. First of all, the extant views of hypothetical entities have trouble accommodating these episodes. Secondly, the suppositionalist view succeeds where the extant views fail.

## 2.2 THE PHENOMENA OF SUPPOSITION

The literature on supposition is far from voluminous, but philosophers have not been altogether silent on the topic either.<sup>2</sup> They characterize the sense of 'suppose' that they are concerned with in a number of ways. In general, philosophers treat supposing as a propositional attitude, and when we suppose, we temporarily put forward a proposition

- for the sake of argument (Fisher (1989), 402; Green (2000), 376),
- for the sake of drawing out its implications or consequences (Fisher (1989), 402; Weinberg and Meskin (2006), 193),
- for the sake of entertaining a thought (Denham (2000), 203),

---

<sup>1</sup>Hence, my argumentative strategy in this chapter mirrors that of van Fraassen (1980), which argues that scientific realism is not compulsory, and *not* that constructive empiricism is the only rational position.

<sup>2</sup>For some comments on the paucity of the literature on supposition, see Fisher (1989), 401 and Green (2000), 377.

- without asserting it (Fisher (1989), 403),
- without committing to a belief in that proposition (Fisher (1989), 403; Stalnaker (1998), 16; Weinberg and Meskin (2006), 191–192), and/or
- without committing to the truth of that proposition (White (1903), 502; Fisher (1989), 403; Stalnaker (1998), 16; Weinberg and Meskin (2006), 191).

As Mitchell Green has noted, there are senses of the words ‘assume,’ ‘say,’ ‘pretend,’ ‘hypothesize,’ and ‘imagine’ that play much the same role as ‘suppose’ (Green (2000), 377). Green also points out that our term for the act of supposition is actually ambiguous between a speech act and a mental act/state. In other cases, we distinguish between a speech act (e.g., assertion) and a mental act/state (e.g., belief), but with supposition, this is not the case (Green (2000), 376).

In many cases, philosophers discuss supposition not (or not just) for its own sake, but for the sake of getting clear on some other issue. Green, for example, discusses it in order to argue for a kind of attitude externalism, which amounts to the claim that “whether one is in [an intentional] state does not depend solely upon how things are within one’s skin” (Green (2000), 376). Other philosophers have debated over whether suppositions can be reduced to conditionals, or vice versa.<sup>3</sup> My argument does not depend on the outcome of either of these debates. Finally, many philosophers are concerned with whether supposition and imagination are distinct, and if so, what distinguishes them. I will have more to say on this topic below.

---

<sup>3</sup>See, for example, Mackie (1973), 92–108 and Fisher (1989), 410–412.

Philosophers have discussed a number of paradigm cases of supposition. They include examples from everyday life, mathematics, and the various systems of natural deduction.<sup>4</sup> In order to further circumscribe the phenomena of interest, I consider some examples of each.

### 2.2.1 EVERYDAY LIFE

Supposition is something that we engage in quite often. While watching an NFL game, I might say the following to my friend: “Suppose the Ravens beat the Patriots. Will it still be possible for the Steelers to make the playoffs?” Or my friend (who happens to be much more knowledgeable about the NFL than I am) might say: “Suppose the Ravens beat the Patriots. Then the Ravens will get a first-round bye in the playoffs. In that case, the Ravens will play the winner of the Steelers-Colts game in the second round.”

### 2.2.2 MATHEMATICS

In the course of constructing a proof in mathematics, one may make a supposition for a variety of reasons, including (but not limited to) the use of conditional proof, proof by cases, and *reductio ad absurdum*. The proof for the irrationality of  $\sqrt{2}$  is a good example of the latter. First, suppose that  $p$  and  $q$  are integers with no common divisors such that  $p/q = \sqrt{2}$ . Then  $p^2/q^2 = 2$ . Hence,  $p^2 = 2q^2$ . It follows that  $p$  is even, and so  $p = 2r$  for some integer  $r$ . Substituting, we get  $(2r)^2/q^2 = 2$ , which is equivalent to  $2r^2 = q^2$ . Hence,  $q$  is also even. But then  $p$  and  $q$  have a common divisor, namely, 2. This contradicts our

---

<sup>4</sup>See, for example, the cases discussed in White (1903), Fisher (1989), and Green (2000).

supposition that they have no common divisor. Hence,  $\sqrt{2}$  is irrational.<sup>5</sup>

### 2.2.3 NATURAL DEDUCTION

In the various natural deduction systems of formal logic, one makes suppositions for a variety of reasons, including (but not limited to) the use of formal analogues of conditional proof and *reductio ad absurdum*. There are a number of ways to indicate that a step is a supposition, or is under the scope of a supposition. One such way involves the Fitch-style

---

<sup>5</sup>Fisher (1989), 402 and Rynasiewicz et al. (2010), 5 both make mention of this example as a paradigm case of supposition.

proofs used in Barwise and Etchemendy (2002), an example of which is the following:

1		
2	$\neg(A \vee \neg A)$	
3	$A$	
4	$A \vee \neg A$	$\vee$ Intro: 3
5	$\perp$	$\perp$ Intro: 2, 4
6	$\neg A$	$\neg$ Intro: 3–5
7	$\neg A$	
8	$A \vee \neg A$	$\vee$ Intro: 7
9	$\perp$	$\perp$ Intro: 2, 8
10	$\neg\neg A$	$\neg$ Intro: 7–9
11	$\perp$	$\perp$ Intro: 6, 10
12	$\neg\neg(A \vee \neg A)$	$\neg$ Intro: 2–11
13	$A \vee \neg A$	$\neg$ Elim: 12

In this proof, the horizontal lines indicate what we have supposed, while the vertical lines indicate the scope of a supposition. We've supposed steps 2, 3 and 7 for the formal analogue of *reductio ad absurdum*, namely,  $\neg$  Intro.

#### 2.2.4 FICTION

We've seen some paradigm cases of supposition, and the examples could be multiplied. I will now move on to another domain in which supposition is operative, namely, our engagement with works of fiction. Examples of such works include novels, short stories, comic books, television programs, films, poems, drawings, paintings, sculptures, and thought experiments. We engage with a work of fiction, not just when we read or view it, but also when we talk about it with others, or when we write or lecture about it. This list of examples is by no means exhaustive, and there may be examples of each category that we do not engage with by means of supposition. For example, some poems and some paintings may come to mind as counterexamples, though it might be argued that supposed counterexamples are works of art that do not count as works of fiction. It will suffice for my purposes if there are examples of each that we engage with by means of supposition.

The claim that supposition is operative in our engagement with works of fiction is a contentious one, and in the next section, I will offer a defense of this claim. But before doing so, it's worth emphasizing that, when we put forward a set of propositions as a fiction, we do so without asserting them, without committing to a belief in them, and without committing to their truth. At the very least, then, fiction has much in common with the other cases I've discussed.

### 2.3 FICTION: IMAGINATION OR SUPPOSITION?

In this section, I will clarify and defend my claim that we can best understand our engagement with works of fiction in terms of supposition. As mentioned above, this is a contentious claim, primarily because some philosophers argue that imagination, which they view to be distinct from supposition, is the proper activity when engaging with a work of fiction, in the sense that our engagement involves and/or should involve imagining, but not supposing. My primary goal in entering into this debate with these other philosophers is to clarify what I mean by ‘supposition.’ This clarified notion will form the basis of the suppositionalist view that I will begin sketching in the next section. Moreover, in the course of developing the suppositionalist view of science, I will draw some analogies with the case of works of fiction, in which case getting clear on the latter is of some importance.

#### 2.3.1 IMAGINATION

Before going any further, it will be necessary to say a word about imagination. Imagination, unlike supposition, has sparked a large literature in philosophy. But as Tamar Szabó Gendler notes, imagination presents some serious taxonomic challenges, and there is no real consensus over the correct taxonomy (Gendler (2011), §1). These taxonomic challenges arise from the fact that there are a number of paradigm cases of imaginings, and it’s difficult to see what all of these cases have in common. Daydreaming, make-believing, and reading a novel may all involve the imagination, but it’s not clear, at least to writers like Kendall Walton, what all such examples share (Walton (1990), 19).

Although there is no agreed-upon taxonomy, examining some proposals can give us

a sense for the phenomena in question. Gendler discusses a number of such proposals (Gendler (2011), §1). To take one example, Gregory Currie and Ian Ravenscroft distinguish among *creative imagination*, *recreative imagination*, and *perception-like imagination*. Creative imagination involves “put[ting] together ideas in a way that defies expectation or convention: the kind of imaginative ‘leap’ that leads to the creation of something valuable in art, science, or practical life” (Currie and Ravenscroft (2002), 9). Recreative imagination is “the imaginative capacity ... that underpins perspective-shifting” (Currie and Ravenscroft (2002), 9). Perception-like imagination involves various kinds of imagery, e.g., visual and auditory imagery (Currie and Ravenscroft (2002), 11). To take another example, Stephen Yablo distinguishes between *propositional imagining* (e.g., “imagining that there is a tiger behind the curtain”) and *objectual imagining* (e.g., “imagining the tiger itself”) (Yablo (1993), 27). Like Yablo, Walton distinguishes between “imagining a proposition [and] imagining a thing,” but adds another category: “imagining doing something” (Walton (1990), 13).

### 2.3.2 IMAGINATION, *Not* SUPPOSITION

Based on the proposed taxonomies mentioned above, one might suspect that imagination is distinct from supposition. After all, the kind of activity involved in recreative imagination seems foreign to that involved in supposing something for the sake of argument. Moreover, while one may be able to *imagine* seeing a tiger, it’s not clear that one can *suppose* seeing a tiger.

In fact, a number of philosophers have argued that imagination and supposition are dis-

tinct. Gendler discusses such views as attempts to distinguish between *mere supposition* (i.e., “what is involved in simple cases of hypothetical reasoning”) and *engaged* or *vivid imagination* (i.e., “what is involved in aesthetic participation, engaged pretense, or absorbing games of make-believe”) (Gendler (2011), §2.4). Gendler distinguishes the two by appeal to the puzzle of imaginative resistance. Take some moral proposition *M* that you do not believe to hold. Gendler claims that you may resist imagining that *M* holds, but you have no trouble supposing that *M* holds (Gendler (2000), 80–81). Christopher Peacocke distinguishes “supposing falsely” from “auditorily imagining” on the grounds that they are “phenomenologically distinctive state[s]” (Peacocke (1985), 20). Brian Weatherson claims that “[s]upposing can be coarse in a way that imagining cannot,” since “[w]e can suppose that Jack sold a chair without supposing that he sold an armchair or a dining chair or any particular kind of chair at all” (Weatherson (2004), 20). As I will discuss in some detail later, Jonathan Weinberg and Aaron Meskin draw the distinction at the level of cognitive architecture (Weinberg and Meskin (2006), 191–199). And to take one last example, Tyler Doggett and Andy Egan claim that imagination sometimes motivates us to act in certain ways, but that supposition never does this. For example, when we suppose, for the sake of a *reductio*, that we are elephants, we don’t seem to be motivated to act in the ways that children do when they imagine that they are elephants (Doggett and Egan (2007), 2).

Moreover, some of these very same philosophers hold that imagination, but not supposition, is the proper activity when it comes to engaging with a work of fiction. This is Gendler’s view, as she writes of supposing “for the sake of argument,” as opposed to engagement with a work of fiction (Gendler (2000), 80). Weatherson claims that his view

in general is indebted to Gendler's, who he sees as "conclud[ing] that what we do in fiction, where we try to imagine the fictional world, is very different from what we do, say, in philosophical argumentation, where we often suppose that things are different from the way they actually are" (Weatherson (2004), 20). Similarly, Weinberg and Meskin claim that "[s]upposition is almost always used for epistemic purposes," while imagination "can be used epistemically, but is deployed in a much larger range of activities, including the construction of and engagement with fictions, role-playing games, and daydreaming" (Weinberg and Meskin (2006), 193).

If these philosophers are correct, then my claim that we can understand engagement with fiction in terms of supposition would seem to be false. In that case, my claim is in need of some clarification, and some defense, and it is to this task that I now turn.

### 2.3.3 'SUPPOSING\*' AS A TECHNICAL TERM

It's not obvious, however, that such philosophers are correct, especially given the fact that other philosophers have argued that supposition is the proper activity when it comes to engaging with works of fiction. Jonathan Ichikawa and Benjamin Jarvis explicitly state that "the propositional attitude we principally come to have when engaging with fictions is *supposing*" (Ichikawa and Jarvis (2012), 129).<sup>6</sup> W. Jeffrey White wrote over a century ago about "artistic suppositions," which are relevant to our engagement with works of fiction (White (1903), 506). Finally, Rynasiewicz and his collaborators have argued that we can, in fact, account for fiction in terms of supposition (Rynasiewicz et al. (2010), 6–7). While

---

<sup>6</sup>Emphasis is the author's unless otherwise noted.

it's not clear that these philosophers have the same notion of supposition in mind, their work shows that there is no consensus that imagination, and not supposition, is the proper attitude to take towards a work of fiction.

I propose that we use 'supposing\*' and 'supposition\*' as technical terms that encompass both mere supposition and vivid imagination, as discussed above. My claim that supposition\* is the proper activity when it comes to engaging with a work of fiction therefore amounts to the view that such engagement involves *both* mere supposition and vivid imagination, though I will often drop the 'mere' and the 'vivid' for the sake of brevity, when it is otherwise clear. I will refer to this as 'my view' in what follows. Hence, I disagree with those philosophers who think that *only one* of these activities is proper when it comes to engaging with fictions.

My defense of this claim will take two forms. The first has to do with the nature of the distinction between vivid imagination and mere supposition. The second has to do with the puzzle of imaginative resistance.

#### 2.3.4 THE NATURE OF THE DISTINCTION

As discussed above, if one wishes to argue that imagination, but not supposition, is the proper attitude regarding fiction, in the sense that our engagement involves and/or should involve imagining but not supposing, one must first distinguish between the two. I will now argue that the nature of this distinction is not a problem for my view. If one draws a distinction according to which one activity is a subset of the other, then there will be cases in which both supposition and imagination are involved when engaging with works

of fiction. If one succeeds in drawing a sharp distinction between the two, according to which the two are mutually exclusive activities, it becomes clear that both imagination and supposition are involved when we engage with works of fiction. And if one succeeds in drawing a vague distinction between the two, then it can often be unclear whether one, and not the other, is involved in our engagement with works of fiction. The upshot is that we benefit from introducing a technical term, namely, ‘supposition\*,’ that encompasses both mere supposition and vivid imagination.

### **The Subset View**

Some philosophers argue that one activity is a subset of another. One possibility is that supposition is a subset of imagination. Currie and Ravenscroft, for example, distinguish between belief-like imagining and desire-like imagining, and claim that “supposition is belief-like imagining that is isolated from, or at least not substantially affected by, desire-like imagining” (Currie and Ravenscroft (2002), 35). In a similar vein, Alvin Goldman distinguishes between two different kinds of imagination, one of which he terms “S-imagination” (short for suppositional imagination). The basic idea here is that when we imagine *that* something is the case, this is equivalent to *supposing* that it’s the case (Goldman (2006), 47–48). In order for such a view to present a problem for my view, one would have to claim that non-suppositional imagination is the only proper activity when it comes to engaging with works of fiction. But if we take either Currie and Ravenscroft’s account or Goldman’s account as our basis, it’s not clear that this is a plausible view. If we take Currie and Ravenscroft’s account as our basis, one would have to argue that, when we

engage with a work of fiction, all of our imaginings are substantially affected by desire-like imaginings. And if we take Goldman's account as our basis, one would have to argue that we never imagine *that* something is the case when we engage with fiction. Neither of these views is plausible. If one wants to advocate this version of the subset view, and argue that imagination is the proper activity when it comes to fiction, the burden of proof is on such an advocate to develop this view in a plausible way.

Another possibility is that imagination is a subset of supposition. To take one example of such a view, Alison Denham holds that "S imagines that  $p$  if and only if he entertains the thought that  $p$  but does not judge and would not be disposed to judge that  $p$ , if put the question" (Denham (2000), 203). She goes on to equate 'entertains the thought' with 'supposes,' in which case imagining is a kind of supposing. Denham's view is, of course opposed to those views that treat imagination and supposition as distinct. But, assuming for the sake of argument that Denham is correct, her view presents no problem for my view. If one were to claim that imagination is the proper activity when it comes to fiction, one would thereby imply that supposition is as well.

A third possibility is that the two are coextensive. Ichikawa and Jarvis come closest to this view, since they use 'imagine' and 'suppose' interchangeably (Ichikawa and Jarvis (2012), 129). But they also emphasize that their language use should be treated as stipulative. In any case, such a view is not at odds with my view, for the very same reason that Denham's view is not at odds with my view. If the distinction is understood as some variant of the subset view, then, this presents no problem for my view.

### **A Sharp Distinction**

Another way in which to understand the distinction between imagination and supposition is as a sharp distinction, according to which the two are mutually exclusive activities. This is one way to understand Peacocke, Gendler, Weatherson, and Doggett and Egan, whose views I discussed above. However, I will argue that such views do not raise any problems for my view. If the distinction is understood as a sharp one, there is good reason to think that our engagement with fiction isn't limited to the imagination. Moreover, there is also good reason to think that the above-mentioned philosophers fail to draw a sharp distinction, and at most, succeed in drawing a vague one. And for reasons I will discuss below, a vague distinction is not a problem for my view.

To begin with, Peacocke's view is that we can distinguish imagination and supposition on the basis of phenomenology. He gives the example of the contrast between "supposing falsely that the orchestra . . . is playing Mozart's Linz Symphony" and "auditorily imagining a performance of that Symphony" (Peacocke (1985), 20).

Against this, Weinberg and Meskin argue that one cannot appeal to phenomenology to draw the distinction. On the one hand, "imaginative states need not involve phenomenology"—to take one example, "in the context of very long novels it is unlikely that each proposition that you imagine has associated phenomenology"; and on the other hand, "there is no reason to think that suppositions cannot be phenomenologically rich"—to take another example, "suppose that there was a house that looked exactly like yours except it was left-right mirror-inverted" (Weinberg and Meskin (2006), 192).

However, it's not clear that such a response shows that Peacocke has failed to draw the

distinction. Against the first point, it might be claimed that what Weinberg and Meskin refer to as imagination without phenomenology, is really just supposition, in which case we do have a sharp distinction. But Weinberg and Meskin's alleged counterexample would then count as an example of suppositional engagement with fiction, which is in accordance with my view.

Regarding the second point, Peacocke needn't commit himself to drawing the distinction in terms of phenomenological *richness*—all that he needs is some *difference* in phenomenology. But, as Weinberg and Meskin point out, we can easily confuse imagination with supposition—“[a] supposition may slide into a daydream via free association if one is not paying close attention, for example, and it may be impossible to say where the one activity leaves off and the other commences” (Weinberg and Meskin (2006), 193). If this is, in fact, impossible in some cases (as seems likely), and if one is committed to a phenomenological difference between imagination and supposition, it may be a phenomenological difference of which we are unaware. In this case, phenomenological differences may tell in favor of a vague distinction between imagination and supposition, as opposed to a sharp one. There may be clear cases in which we can point to phenomenological differences in order to distinguish the two. But in light of examples like the one that Weinberg and Meskin discuss, there also seems to be a range of borderline cases in which phenomenological considerations are not sufficient to classify an activity as one of either imagination or supposition.<sup>7</sup>

---

<sup>7</sup>One might suspect that the cognitive phenomenology debate is relevant to the dispute between Peacocke and critics like Weinberg and Meskin. The main point at issue in this debate is whether cognitive states, of which supposing is an example, have any kind of phenomenology at all. However, apart from the fact that a resolution to this debate seems a long way off, there is reason to think that the debate does not, in fact,

In short, Peacocke may succeed in drawing a sharp distinction, but it's not one that's problematic for my view. Alternatively, he may succeed in drawing a vague distinction. But for reasons I will outline below, this is not a problem for my view. Gendler, Weatherson, and Doggett and Egan, on the other hand, do not succeed in drawing any kind of distinction—sharp or vague.

Gendler's way of drawing the distinction depends on her claim that "imagination requires a sort of participation that mere hypothetical reasoning does not" (Gendler (2000), 80). According to Gendler, this is why we have imaginative resistance, but no suppositional resistance. As Weinberg and Meskin make clear, such participation involves "self-directed judgments"—"we 'take ourselves', when imagining that M holds, to be 'implicated in the way of thinking that M presupposes'" (Weinberg and Meskin (2006), 192).<sup>8</sup> To take an example, if the author of a novel invites us to imagine that there is a duty to commit adultery, that author is, in Gendler's words, "providing us with a way of looking at *this* world which we prefer not to embrace" (Gendler (2000), 79). In such a case, imagining that there is such a duty in some sense involves our assent to the claim that it is, indeed, a duty—not just in some fictional world, but in our world. This way of looking at our world is the kind

---

hold any promise of deciding what is at issue between Peacocke, and Weinberg and Meskin. Following Tim Bayne and Michelle Montague, we can characterize the debate in terms of "conservatives," who hold that cognitive states lack phenomenology, and "liberals," who hold that they possess phenomenology (Bayne and Montague (2011), 2–3). As Bayne and Montague emphasize, both conservatives and liberals hold that sensory and imagistic states possess phenomenology. Hence, both will admit that the kind of imagining that Peacocke has in mind here has some kind of phenomenology. On the conservative view, supposing and imagining may or may not be phenomenologically distinct. This is because cognitive states can be associated with imagistic and sensory states, and so cognitive states can be associated with imagistic and sensory phenomenology. Likewise, on the liberal view, supposing and imagining may or may not be phenomenologically distinct. On this view, both have some phenomenology, and it is an open question whether such phenomenology differs. Hence, it's not clear, at least at this point, that the cognitive phenomenology debate has any implications for Peacocke's view.

<sup>8</sup>They quote from Gendler (2000), 80.

of participation that Gendler has in mind. Against this, Weinberg and Meskin claim that “there is no reason to think that imagining is *essentially* participatory (in this sense) rather than merely *typically* participatory” (Weinberg and Meskin (2006), 192). Weinberg and Meskin do not give any examples of non-participatory imaginings. But one such example may be a case in which the author of a novel invites us to imagine that there is a law that requires people to commit adultery. There is no inclination to map this law onto our own world, as there is with the case of the duty to commit adultery. And in this case, we presumably do not have the kind of participation that Gendler has in mind. If she does have a weaker sense of participation in mind, it may be the case that supposing involves this kind of participation as well. It seems clear that, when we suppose some claim for the sake of argument, and examine what follows from that claim, we are, in some minimal sense, “implicated in the way of thinking that [the claim] presupposes” (Gendler (2000), 80). And if that is all that is required for participation, then supposing is participatory as well. It may be true that imagining that there is a duty to commit adultery is participatory in Gendler’s sense, while supposing such a duty is not. But participation cannot be used to distinguish all cases of imagining from all cases of supposing.

Against Weatherson’s proposal that imagination requires a fineness of grain that supposition does not, Weinberg and Meskin point out some examples of imaginings that fail to be fine-grained in Weatherson’s sense, namely, “the itsy-bitsy spider’s waterspout ascent” and “London Bridge’s dilapidation” (Weinberg and Meskin (2006), 192). Moreover, some variant of the speckled hen problem may be enough to tell against Weatherson’s proposal, since one may imagine a speckled hen without thereby imagining, of some specific number,

that the hen has that many speckles.<sup>9</sup>

Doggett and Egan’s way of drawing the distinction does not fare any better. They give the example of children at play, and point out that “[w]hen they imagine that they are cats or elephants or cops or robbers, this can give rise to all sorts of behavior” (Doggett and Egan (2007), 2). Their claim is that imagining is capable of motivating our behavior, whereas supposing is not. But, to modify one of their examples (Doggett and Egan (2007), 2), it’s not clear that *supposing* that John McCain is elected president is motivationally inert. Drawing out the consequences of this supposition may very well motivate us to act in one way or another. Moreover, while Doggett and Egan can admit that sometimes imagining *doesn’t* motivate behavior, they commit themselves to the claim that it is always *capable* of motivating behavior. But it’s not clear that this is the case. For example, if I were to imagine that it didn’t rain this morning, it’s not clear what, if any, behavior could possibly follow.

Participation, fineness of grain, and motivation, then, cannot be appealed to in order to sharply separate imagination from supposition. All three of these features admit of degrees, and so one might be tempted to explore the possibility of using one or more of these features as a way to draw a vague distinction. After all, an activity can be more or less participatory, more or less motivating, and finer or coarser in grain. But based on the above discussion, this won’t suffice for even a vague distinction. If an act of imagination can involve the same degree of participation, motivation, or grain, as an act of supposition, then one cannot distinguish between the two based on such features at all. Gendler, Weatherson,

---

<sup>9</sup>I owe this suggestion to Steven Gross.

and Doggett and Egan may have succeeded in identifying some features that are typical of imagination. But insofar as they fail to draw a distinction between imagination and supposition, their views pose no problem for mine.

This may motivate one to look elsewhere for a way to sharply distinguish between imagination and supposition. Here, I'll consider one possible way to do so. On this proposal, imagination is inherently imagistic, while supposition is not. The basic idea can be illustrated by Descartes' distinction between imagination and pure understanding (Descartes (1988), 111). Descartes considers two geometrical figures—one with three sides, i.e., a triangle; and one with 1,000 sides, i.e., a chiliagon. Just as we can understand a triangle as a figure with three sides, we can understand a chiliagon as a figure with 1,000 sides. But while we can imagine a triangle, we cannot imagine a chiliagon. For Descartes, this is because imagining requires an application of “the mind's eye” to the sides of the figure and the area enclosed by those sides. Hence, imagination requires something like a mental image, while understanding does not. We might then use this idea to distinguish between imagination and supposition. Imagination requires some kind of mental image, while supposition involves entertaining a proposition without any kind of imagery.

It's not obvious that this is a satisfactory way of drawing a distinction between imagination and supposition. But assuming that it is, it does yield a sharp distinction. And, more importantly, it yields a distinction which presents no problem for my view. It's clear in this case that both imagination and supposition will be required for our engagement with a work of fiction. Any work of fiction that involves a description of a chiliagon is sufficient to show that this is the case.

The foregoing discussion provides some support for the claim that attempts to draw a sharp distinction between imagination and supposition either fail, or yield a distinction according to which it's clear that both imagination and supposition are involved in our engagement with works of fiction. In this case, the burden of proof is on others to show that one can successfully draw a sharp distinction that can cause problems for my view.

### **A Vague Distinction**

I now turn to vague distinctions between imagination and supposition. Weinberg and Meskin have a proposal for drawing the distinction that is explicitly vague. They draw the distinction in terms of cognitive architecture, and begin their account by distinguishing between two “functionally distinct pieces of cognitive architecture,” which they label the “belief box” and the “imagination box,” respectively (Weinberg and Meskin (2006), 179). As a first approximation, then, we can say that their ‘two-box’ view amounts to the claim that beliefs and imaginings are distinct kinds of cognitions. Things become more complicated once we consider how supposition fits into this view. Weinberg and Meskin hold that supposition, like imagination, is an imagination-box-involving cognition, and that some cases “fall in a gray area between supposition and imagination” (Weinberg and Meskin (2006), 197). By this, they have in mind “cases in which we engage in hypothetical reasoning to determine what sorts of outcomes are possible (i.e. to determine not what *will happen*, but what the *range* of things that *might happen* is)” (Weinberg and Meskin (2006), 197). For example, determining what might happen if one fails to pay rent on time may involve drawing out consequences, based on the lease; or picturing one’s landlord in

Hawaii, far away from his mailbox. This kind of activity seems to involve some combination of mere supposition and vivid imagination, but it can be difficult to separate the two in a clean way. On a related note, Weinberg and Meskin also point out that we may slide imperceptibly from a supposition, to a daydream (an imaginative activity), and back to a supposition again, and it may be impossible to determine when one activity commences and the other begins (Weinberg and Meskin (2006), 193). In order to accommodate this gray area, they suggest that the term ‘considering’ be used “as a technical term neutral between all these different ways of engaging with a proposition in the [imagination box]” (Weinberg and Meskin (2006), 197). Under this proposal, then, there is no sharp boundary between imagining and supposing, and both are subspecies of considering.

There are two important lessons to learn from Weinberg and Meskin’s proposal. The first is that a vague distinction between imagination and supposition is unproblematic for my view that both imagination and supposition are involved in our engagement with works of fiction. In order to cause a problem for my view, one would have to show that only one of these activities is operative in our engagement with works of fiction. One way in which to do this is to draw a sharp distinction between the two. And as I’ve already argued, there is reason to think that, if the distinction is sharp, both activities will be involved in our engagement with works of fiction. A vague distinction, like Weinberg and Meskin’s, is premised on the possibility that it can be very difficult, if not impossible, to tell when one activity ends and another begins. In such a case, one cannot clearly show that only one of these activities is operative in our engagement with works of fiction. Given the existence of activities that fall within a gray area between imagination and supposition, prudence

dictates that we find a place for both activities when it comes to engaging with works of fiction.

The second lesson is closely related to the first. Given the existence of such a gray area, we can follow Weinberg and Meskin's fruitful suggestion, and propose a technical term to encompass cases of imagination, supposition, and the gray area between them. Nothing much hinges on the choice of this technical term. But 'supposing' seems to me to be preferable to 'considering.' When we engage with a work of fiction, we may suppose that something happens. But 'considering' can give the impression that we consider *whether* something has happened, or consider something for the sake of some practical decision, and we don't necessarily do these things when we engage with fiction. Moreover, while we often suppose and imagine what is not the case, we oftentimes consider what is the case.<sup>10</sup> Hence, in the spirit of Ichikawa and Jarvis (2012), (129), I will use 'supposing\*' as my technical term.

### 2.3.5 IMAGINATIVE RESISTANCE

The second form of my defense of the claim that both vivid imagination and mere supposition are involved in our engagement with works of fiction has to do with the puzzle of imagination resistance.<sup>11</sup> As mentioned above, Gendler (2000) initially raises the puzzle in terms of moral propositions. If a work of fiction invites us to imagine that some moral proposition *M* is true, and *M* is a moral proposition that we don't believe to hold, then we encounter some imaginative resistance, in the sense that it is difficult, if not impossi-

---

<sup>10</sup>I owe this observation to Nick Goldberg.

<sup>11</sup>For a nice, brief introduction to the puzzle, which I have drawn from in my discussion, see Gendler (2011), §5.2.

ble, to imagine that *M* holds. Weatherson argues that the puzzle is not restricted to moral propositions—we can also raise it with propositions that involve attributions of mental states or attributions of content, and also with propositions that contradict our basic metaphysical beliefs about, for example, constitution (Weatherson (2004), 4–5). To take just one concrete example, Yablo gives us the following, which generates imaginative resistance in terms of shape predicates:

You open a children’s book and read as follows: ‘They flopped down beneath the giant maple. One more item to find, and yet the game seemed lost. Hang on, Sally said. It’s staring us in the face. This is a *maple* tree we’re under. She grabbed a five-fingered leaf. Here was the oval they needed! They ran off to claim their prize.’ (Yablo (2002), 485)

In short, we encounter some resistance to imagining at least some propositions that contradict certain kinds of beliefs.

Gendler divides responses to the puzzle into two groups: *can’t* theories and *won’t* theories. She claims that *can’t* theories, like that developed in Weatherson (2004), “often embrace some sort of *impossibility hypothesis*, suggesting that propositions that evoke imaginative resistance are impossible in the context of the stories where they appear, and that this explains why readers fail to imagine them as true in the fiction” (Gendler (2011), §5.2). *Won’t* theories, such as that developed in Gendler (2000), “maintain that readers are unwilling to follow the author’s lead because doing so might lead them to look at the (actual) world in a way that they prefer to avoid” (Gendler (2011), §5.2).

At this point, we can see why the puzzle of imaginative resistance supports my view. Works of fiction can, and sometimes do, invite us to engage with propositions to which we may encounter some imaginative resistance. Now suppose that imagination is the only proper activity when it comes to engaging with a work of fiction. In that case, imaginative resistance precludes us from engaging, as fully as we could, with certain works of fiction, either because we *can't* imagine certain propositions, or because we *won't* imagine those propositions. We could engage more fully with such works if supposition is admitted to be a proper activity when it comes to engaging with a work of fiction. This is because, as many philosophers have noted, we can suppose anything, even logical impossibilities.<sup>12</sup> So in the example above, I am certainly able to suppose that a maple leaf is shaped like an oval, even if I can't imagine it. And in making this supposition, I don't seem to be doing anything illicit when it comes to my engagement with the work. Hence, the puzzle of imaginative resistance supports my use of 'supposition.' It's probably true that some readers may refuse to suppose when they encounter imaginative resistance. But my view is in trouble only if all readers *do* refuse (an empirical claim that is most likely false), or if they *should* refuse (a normative claim that I see no reason to hold).

At this point, I've given my defense of the claim that both mere supposition and vivid imagination are proper activities when it comes to engaging with a work of fiction. And I've introduced 'supposition\*' as a technical term which covers both mere supposition and vivid imagination. Hence, supposition\* is the activity we engage in, not only when it comes to fiction, but also mathematics, natural deduction, and some parts of everyday life,

---

<sup>12</sup>See, for example, Gendler (2000), 80; Weatherson (2004), 20; Weinberg and Meskin (2006), 193; Doggett and Egan (2007), 1; and Ichikawa and Jarvis (2012), 129.

as discussed above.

## 2.4 SUPPOSITIONALISM: A UNIFIED TREATMENT

My ultimate goal in this chapter is to introduce suppositionalism as a view of hypothetical entities that is distinct from the extant views discussed in chapter 1. Rynasiewicz has developed suppositionalism more generally as a theory of supposition as employed across the arts and sciences, yielding a philosophy of fiction, a philosophy of mathematics, and a philosophy of science.<sup>13</sup> I will now turn to the task of sketching the details of suppositionalism, and showing how it accounts for the examples from fiction and mathematics. The natural deduction examples can be subsumed under mathematics. And the everyday life examples can, at least to some extent, be subsumed under the natural deduction examples, insofar as natural deduction can model our patterns of everyday reasoning. To be sure, a fuller defense of the suppositionalist views of fiction and of mathematics is still needed, and I won't offer one in what follows. I will take it that the notion of 'supposition' that suppositionalism is based on is the technical notion that I have just defended. Although Rynasiewicz has not advocated for this technical notion in print, I believe it to be within the spirit of his suppositionalism. And when it is otherwise clear, I will usually drop the quotations around 'supposition,' and take it as understood that this term covers both mere supposition and vivid imagination.

---

<sup>13</sup>See Rynasiewicz (2009) and Rynasiewicz et al. (2010). Much of what follows has its origin in a series of seminars taught by Rynasiewicz, and I am indebted to those who took part in those seminars.

2.4.1 PROPS, INTERPRETIVE ENGINES, AND SUPPOSITIONS

To begin with, Rynasiewicz et al. borrow from the work of Walton (1990) in order to develop a general picture of the practice of supposition, which they summarize as follows:

one begins with a *prop*—a picture, toy, statue, physical model, segment of fictional text—to which is applied an *interpretive engine* from which is generated a (perhaps fuzzy) set of sentences in some language, which may contain previously uninterpreted terms. The resulting set of sentences is a *representation* of the corresponding supposition. (Rynasiewicz et al. (2010), 6)

The set of sentences is not identical to the supposition, but merely a representation because, for Rynasiewicz et al., “[s]uppositions are historically located actions involving language use” (Rynasiewicz et al. (2010), 6). Some examples should make this general picture clear.

In mathematics, the prop is a set of axioms  $A$ , i.e., a decidable set of sentences in a given object language. Hence, the prop in this case is itself an abstract object. The interpretive engine closes  $A$  under logical consequence. Logical consequence will amount to different things in, say, classical and intuitionistic mathematics. But once we say what logical consequence amounts to, we can obtain  $Cn(A)$ . The representation of the supposition is given by a formal theory  $T$  in a given object language (i.e.,  $T = Cn(T)$ ).<sup>14</sup>

The fictional case is a bit more complicated than the mathematical one. When one engages with a work of fiction, one begins with a prop. In short, the prop is just the work of fiction. There are a number of examples of props here, including novels, short stories,

---

<sup>14</sup>This paragraph is a summary of a slide from Rynasiewicz (2009).

comic books, television programs, films, poems, drawings, and paintings. The prop may be considered as an abstract object (e.g., the set of sentences that constitutes *The Great Gatsby*), a type of empirical entity (e.g., a specific edition of *The Great Gatsby*), or a specific empirical entity (e.g., my copy of *The Great Gatsby*).

The interpretive engine is the means by which one generates the representation of the supposition. In other words, it is the means by which we engage with a work of fiction, and the sense of engagement with which I am concerned is the sense of determining what is true according to the supposition. The interpretive engine, according to Rynasiewicz et al., determines the elements of this set by “using a blend of deductive, inductive, and interpretive principles” (Rynasiewicz et al. (2010), 6). We make various inferences based on what we encounter in the work, and bring our background knowledge to bear in order to determine what is true according to the supposition.

What is true according to a supposition is not the same as what is explicitly stated, or otherwise depicted, in a work of fiction. The set of sentences that is the representation of the supposition generated by *The Great Gatsby* includes sentences that Fitzgerald never actually wrote. For example, it’s true that Jay Gatsby occasionally has to use the restroom, even if this is never explicitly stated in the work. The interpretive engine is meant to explain why this is the case, and how one can go from a work of fiction to a representation of a supposition.

While the basic idea is clear, it should be noted that, when it comes to the interpretive engine, the devil is in the details. Specifying the exact combination of deductive, inductive, and interpretive principles that are involved in our engagement with a work of fiction is

an outstanding issue, not just for suppositionalism, but for the philosophy of fiction more generally. To take one example, it would be problematic to close the representation of the supposition under logical consequence, as ordinarily understood, if the work in question contains a contradiction, since we don't want the representation of the supposition to contain every sentence. In that case, the interpretive engine needs to employ some non-explosive logic. To take another example, the importation of certain bits of background knowledge may be fair game when it comes to engaging with some works, but not for others. We may be able to infer that Warren G. Harding was president during the events of *The Great Gatsby*, since it takes place during the summer of 1922. But it's unclear whether Harry S. Truman was president during the events of the first Wile E. Coyote and Road Runner cartoon in 1949. In other words, one issue with the importation of background knowledge is that it seems that the less realistic a work is, the less you can import. Finally, the issue of which interpretive principles to employ is debated among critics, who often disagree sharply over whether, say, authorial intention matters at all to the interpretation of a work. In that case, it's far from obvious what interpretive principles should go into the interpretive engine.

There are at least two ways of understanding what the interpretive engine is. The first can be thought of as a normative view about the correct combination of principles that *should* go into the interpretive engine. It implies that there is one correct answer to the question as to what sentences belong in the (perhaps fuzzy) set that is the representation of the supposition. The second way of understanding the interpretive engine is in terms of what we *actually do* when we engage with a work of fiction, and in that case, the interpre-

tive engine must be psychologically realized in some sense. If so, it can, and often will, be the case that two individuals have distinct interpretive engines. And, as Rynasiewicz et al. note, “different interpretive engines can (and typically will) yield distinct suppositions” (Rynasiewicz et al. (2010), 6). In that case, two individuals may disagree about what, exactly, is true according to the work in question. These individuals could then attempt to argue in favor of their respective representations of the supposition, by appeal to the text and to the interpretive principles employed in their respective engines. This, in turn, may lead to an argument about the status of those principles. Those who work in literary and art criticism argue about these very issues. To take an example, there is some controversy over the role of authorial intention in interpreting a work, as William K. Wimsatt and Monroe Beardsley discuss in an important paper (Wimsatt Jr. and Beardsley (1946)). Suppositionalism makes it explicable why critics would argue about such matters. In short, the difference in interpretive engines at the psychological level leads to a discourse about what the interpretive engine *should* include, in which case we need to understand the interpretive engine in both of these two ways. To be sure, the exact nature of the interpretive engine is an issue for further research. But now that I have brought attention to this issue, for ease of exposition I will speak of *the* interpretive engine and *the* representation of a supposition in what follows.

#### 2.4.2 OBJECTS OF SUPPOSITION

In mathematics, we may suppose that  $p$  is an integer. In the course of engaging with a work of fiction, we may suppose that Superman has heat vision. Here we have two objects:

$p$  and Superman. These are examples of what Rynasiewicz et al. call “objects of supposition” (Rynasiewicz et al. (2010), 5). Objects of supposition, then, include mathematical objects and fictional objects, and importantly, not all objects of supposition are fictional objects (or, for that matter, mathematical objects).

An object of supposition is, at the most general level, an object, the existence of which a language user stipulates; and objects of supposition are to be distinguished from empirical entities as discussed in chapter 1. Rynasiewicz et al. “caution that *to suppose* [something] is not just the same as to entertain a proposition,” since “the act of supposing may involve the extension of our language by the introduction of new terminology” (Rynasiewicz et al. (2010), 5). If this new terminology lacks an interpretation, then sentences that contain such terminology may fail to express a proposition. The important point here is that, for Rynasiewicz et al., supposition is an attitude that we can take towards things (i.e., objects of supposition), as well as towards propositions.

One might then suspect that Rynasiewicz et al.’s suppositionalism immediately falls prey to something like the following objection. As emphasized above, most philosophers treat supposing as a propositional attitude. In that case, the only possible object of supposition would be a proposition, and the notion of an object of supposition that is not a proposition (like  $p$  or Superman) may seem contentious.

But we can use the technical notion of ‘supposition\*’ that I’ve introduced in order to answer this objection. Recall that, according to that notion, supposition\* includes both mere supposition and vivid imagination. Moreover, as mentioned above, philosophers like Walton and Yablo have admitted a notion of *objectual* imagining, which is to say that

one can imagine an object (say, a tiger) as opposed to a proposition (Walton (1990), 13; Yablo (1993), 27). It may be contentious to claim that there is a corresponding notion of objectual supposition (in the sense of mere supposition), and I will remain noncommittal about the prospects for such a notion.<sup>15</sup> But given that my technical notion of supposition\* covers vivid imagination, and given that there is an objectual notion of imagining, there is a corresponding objectual notion of supposition\* (in my technical sense). In that case, objects of supposition are, to a first approximation, the objects of our objectual supposing\*.

We can say a bit more about the nature of objects of supposition if we work with the representation of a supposition. A representation  $S$  is a set of sentences formulated in some language, and that language will include, among its constituents, various terms. The terms that occur in  $S$  refer to objects of supposition. The previously uninterpreted terms that  $S$  may contain are the terms used to introduce the objects of supposition. The terms that have been interpreted previously also refer to objects of supposition—it's just that those objects have been introduced before. Let  $T$  be the set of terms that occur in  $S$ . We can partition  $T$  into equivalence classes of co-referring terms, where co-reference is our equivalence relation. And we can model an object of supposition formally by identifying it with such an equivalence class.<sup>16</sup> Though, to be sure, when we don't have a need for formalism, we can simply speak of the object of supposition as what is referred to.

Some examples should make this clear. For an example from fiction, we can take the Superman comic books. It's true, according to the comic books, that the names 'Superman,'

---

<sup>15</sup>But when we start a proof with words like: 'Let  $n$  and  $m$  be integers such that ...,' there is a way to understand such an utterance as involving objectual, as opposed to propositional, supposition.

<sup>16</sup>Rynasiewicz put forward this proposal in a seminar taught in the spring of 2012.

‘Clark Kent,’ and ‘Kal-El’ all refer to the same person. In that case, they are co-referring terms, and we can model Superman (the object of supposition) as the equivalence class of these co-referring terms. For some examples from mathematics, we can start with Zermelo-Fraenkel set theory (ZF). It’s true, according to ZF, that ‘ $\omega$ ’ and ‘ $\mathbb{N}$ ’ refer to the same set, as do ‘2,’ ‘ $\{0, 1\}$ ,’ ‘ $\{\emptyset, \{\emptyset\}\}$ ,’ and ‘ $\{\{\}, \{\{\}\}\}$ .’ And just like in the fiction case, we can partition the set of terms into equivalence classes of co-referring terms, and identify the objects of supposition with these equivalence classes.

Rynasiewicz has claimed that suppositionalism thus depends on a seemingly counter-intuitive claim, namely, the claim that co-reference is more fundamental than reference.<sup>17</sup> Intuitively, we might say that co-reference depends on reference—two terms co-refer only if both refer to the same object. But recall that we’ve modeled objects of supposition as equivalence classes of co-referring terms. According to this method, we’ve used the notion of co-reference to model the referent. The trick is to determine which terms co-refer *according to the supposition*. Terms that may, in fact, lack a referent can still co-refer according to a supposition. Moreover, this isn’t simply a matter of empty names co-referring to the empty set. ‘Superman’ and ‘Batman’ may both lack an empirical referent, but they are not co-referring terms according to the supposition. Once we determine which terms co-refer according to the supposition, we can model objects of supposition as equivalence classes of such terms and refer to them as such.

One might object at this point that not all terms that occur in the representation of a supposition are terms that name objects of supposition. Ronald Reagan, for example, has

---

<sup>17</sup>Rynasiewicz has made this claim in a number of seminars.

appeared as a character in a number of Superman comic books, and the name ‘Ronald Reagan’ surely refers to a person, and not to an object of supposition. Or at least, so the objection goes.

Terence Parsons’ distinction between immigrants and natives offers a way to answer this objection (Parsons (1980), 51–52). A *native*, for Parsons, is an object introduced or created in a story, and if an object is not a native, then it is an *immigrant*. We can therefore divide objects of supposition into immigrants and natives. As stated earlier, the representation of a supposition may contain previously uninterpreted terms. These are terms for natives. All other terms are for immigrants. The Superman comic books have a number of immigrants, including Ronald Reagan (an empirical entity) and the Predator (a native from another work of fiction). The claim, then, is that both natives and immigrants are objects of suppositions. We can explain the intuition behind the objection by appeal to the fact that Ronald Reagan, *qua* immigrant, has much in common with Ronald Reagan, *qua* empirical entity. In short, the objection fails to distinguish between the empirical entity and the immigrant. The answer to the objection is that the term ‘Ronald Reagan’ sometimes refers to an empirical entity, and sometimes refers to an object of supposition, and it’s important to distinguish the two.

This last claim about referring to objects of supposition needs to be developed more fully. Rynasiewicz has suggested that it’s only once we talk *about* a supposition that we refer to objects of supposition. Peter van Inwagen makes a similar point regarding works of fiction, which we can illustrate by means of the following example. The narrator of the Max Fleischer animated shorts describes Superman as being “faster than a speeding bullet,

more powerful than a locomotive, and able to leap tall buildings in a single bound.” On van Inwagen’s view, what the narrator utters “does not represent an attempt at reference or description” (van Inwagen (1977), 301). This is because it would make no sense to tell the writers of the short that Superman is not, in fact, faster than a speeding bullet. But suppose, after watching one of the shorts, I remark: “Wow! Superman sure is strong!” I’d be making an assertion that’s false if Superman is not, in fact, strong. On van Inwagen’s view, I’m referring to a creature of fiction, and ascribing to it a property, while the writers of the short are doing no such thing (van Inwagen (1977), 305). Rynasiewicz’s account squares with this because we can only model objects of supposition once we apply our interpretive engines to a work of fiction, thereby yielding a representation of the supposition. Hence, the work of fiction itself does not make claims about such objects, or attribute properties to them—we can only talk about objects of supposition once we talk *about* a supposition.

This also yields an analogous view when it comes to mathematics. For example, the proponent of suppositionalism would deny that the axiom of choice is an assertion, or an attempt at reference or description. One may, of course, object to its use, on account of its counterintuitive results, e.g., that sets like  $\mathbb{R}$  can be well-ordered even if we can’t possibly conceive of how this could be. But for the suppositionalist, objecting to its use on the grounds that it’s false would be inappropriate. Its truth or falsity is determined in the same way as any other mathematical claim, that is, relative to the set of axioms that one uses as a prop. Determining the truth of an axiom is just like determining the truth of any mathematical claim—it involves the application of the interpretive engine, which then yields a representation of the supposition. In fact, it’s particularly easy to determine

the truth of an axiom, since it's trivial that a particular axiom which belongs to the set of axioms that one uses as a prop is itself a consequence of that set. Importantly, it's only once we have the representation in place that we can begin to talk about objects of supposition. To be sure, this view is not entirely uncontroversial, and it is opposed to any view that considers mathematical statements to be attempts at reference or description. A full defense of the suppositionalist view of mathematics would have to show why such a view is mistaken, but as I stated above, I won't offer such a defense here.

Rynasiewicz et al. claim that, for lack of a better term, an object of supposition's "ontological status is one of *existence-by-stipulation*" (Rynasiewicz et al. (2010), 5). In this case, one may object to suppositionalism on the grounds that it entails that fictional characters and mathematical entities, both of which are objects of supposition, exist. There is much debate over this issue. Regarding fictional characters in particular, those who follow Bertrand Russell (1905) or W. V. O. Quine (1948), for example, and deny that they exist, would take issue with this consequence. Moreover, if existence-by-stipulation implies some lesser grade of existence, followers of Quine would object on the grounds that existence doesn't come in grades. On the other hand, those who follow van Inwagen (1977) or David Lewis (1978), and grant that fictional characters exist, would be more sympathetic.

If one does find the consequence objectionable, there are a number of things to note. First of all, Rynasiewicz et al. are explicit about the fact that their "emphasis . . . is not on ontology, but on the phenomenology of language use" (Rynasiewicz et al. (2010), 5). Secondly, it's open to a proponent of suppositionalism to retort that objects of supposition exist, and terms refer to them, only in a Pickwickian sense. Thirdly, and relatedly, the on-

tological category of “existence-by-stipulation” may be worked out in a less objectionable way, though, to be sure, this would require a worked-out view, which is currently lacking.

However, perhaps the best response to those who find this consequence objectionable is to point to the work that suppositionalism can do. There are various existence puzzles that occur in mathematics and fiction, and suppositionalism gives us a straightforward way of making sense of these puzzles. In mathematics, I may suppose that there is a number greater than every natural number. If I make this supposition by using the axioms of Peano Arithmetic (PA) as a prop, I can show that there is no such number. But if I make the same supposition by using the axioms of ZF as a prop, I can show that  $\omega$  is such an ordinal number. So is there such a number? According to suppositionalism, such a number exists in the case of ZF, but not in the case of PA, and that’s all there is to say. To be sure, one can suppose that such a number exists in both cases. The important point here is that it’s true, according to the supposition, that there is a number greater than every natural number in ZF, but it’s not true, according to the supposition, in PA.

When it comes to fiction, I may ask whether Superman exists. Well, he certainly exists in the fictional world of Metropolis. But he doesn’t exist in our world. Moreover, when we claim that Superman is stronger than Tom Cruise, we seem to have some reason to grant Superman some kind of existence. Hence, the claims ‘Superman exists’ and ‘Superman does not exist’ both seem to be true in some sense, in which case we have a puzzle. But one who utters the former means to say that Superman is a character in a work of fiction, while one who utters the latter means to say that Superman is not a real person. In the terminology of suppositionalism, we say that Superman is an object of supposition, and

there is no empirical entity (i.e., real person) whose normal properties match the conformal properties of Superman, especially not Tom Cruise. I now turn to what Rynasiewicz et al. have to say about these two forms of predication.

### 2.4.3 NORMAL AND CONFORMAL PREDICATION

When it comes to objects of supposition and their properties, Rynasiewicz et al. distinguish between “two patterns of predication” (Rynasiewicz et al. (2010), 6). The first is *normal* predication. Let  $\varphi$  be a predicate, and  $\tau$  an object of supposition. If  $\varphi(\tau)$  is true, then (informally)<sup>18</sup> we can call  $\varphi$  a *normal property* of  $\tau$ . The second pattern of predication is *conformal* predication, which is to say that the predication is in conformity with the supposition. Let  $S$  be the representation of a supposition. Then if  $\varphi(\tau) \in S$ , we can (informally) call  $\varphi$  a *conformal property* of  $\tau$ . We can also introduce an ‘according to the supposition’ operator ( $\mathcal{S}$ ), so that  $\mathcal{S}(\varphi(\tau))$  is true iff  $\varphi(\tau) \in S$ . Importantly, this operator needn’t be explicated in terms of possible worlds, as is the case with Lewis’ truth-in-fiction operator (Lewis (1978)). And, indeed, this is a virtue of Rynasiewicz’s theory. Works of fiction may contain contradictions or other sentences that trigger imaginative resistance, and in light of such phenomena, they may only be instantiated in *impossible* worlds, in which case a Lewisian possible worlds analysis would be inappropriate.

The distinction between normal and conformal predication allows Rynasiewicz et al. to account for a variety of features that suppositions can and do possess. First of all, it allows them to distinguish between what we might call ‘truth inside of a supposition’ from

---

<sup>18</sup>I will return to the need for this parenthetical remark below.

‘truth outside of a supposition.’ The former has to do with conformal predication (e.g., Superman has heat vision), while the latter has to do with normal predication (e.g., Superman was created by Jerry Siegel and Joe Shuster). Other philosophers have recognized the need to distinguish between these two kinds of predication, and two kinds of solutions suggest themselves. First of all, some philosophers have distinguished between two different kinds of properties. For example, Parsons distinguishes between nuclear properties (e.g., is blue) and extra-nuclear properties (e.g., is thought about by me) (Parsons (1980), 23). Secondly, some philosophers have distinguished between two different kinds of relations that an object can bear to a property. For example, Bernard Linsky and Edward N. Zalta distinguish between exemplifying and encoding properties (Linsky and Zalta (1995), 536), while van Inwagen distinguishes between having a property and being ascribed a property (van Inwagen (1977), 305). Rynasiewicz et al., however, embrace neither of these purported solutions. According to them, there is only one kind of property, and one kind of relation that obtains between an object and a property, and this is tracked by the pattern of normal predication (Rynasiewicz et al. (2010), 6). To say that an object has a conformal property  $\varphi$  is just shorthand for saying that, according to the supposition, it has a normal property  $\varphi$ . Hence, there is not really a distinction between normal and conformal properties—Rynasiewicz et al. merely use these terms informally, as a means to explain their position, and I will follow their use.

Secondly, the distinction allows Rynasiewicz et al. to account for the fact that supposition and belief are not inconsistent with one another—in short, if a person  $S$  supposes a proposition  $p$ , that does not entail that  $S$  does not believe  $p$ . This is a point that some

philosophers have failed to recognize. White, for example, confines supposition to “conceptions that are *not known to be true*, and conceptions that are *known not to be true*” (White (1903), 502). And Denham, who views imagination in terms of supposition, claims that “*S* imagines that *p* if and only if he entertains the thought that *p* but does not judge and would not be disposed to judge that *p*, if put the question” (Denham (2000), 203). But Alan Leslie’s experimental work strongly suggests that we can pretend that *p* while at the same time believing that *p*. Leslie’s experiment consists of a pretend tea party with a child participant. After the child ‘fills’ two empty cups with imaginary tea, the experimenter turns over one of the cups and asks the child to point to the empty cup. The child believes that both cups are empty, but pretends that only one of the cups is empty (Leslie (1994), 223–225).<sup>19</sup> This kind of pretense falls within the domain of vivid imagination, and so is naturally subsumed under my notion of ‘supposition.’ In this case, Leslie’s results strongly suggest that supposition and belief are not inconsistent with one another.

Suppositionalism gives us a way to understand this. It may be the case that an object of supposition has a conformal property that matches with one of its normal properties. A bit more formally, for some  $\varphi$ , it may be the case that both  $\varphi(\tau)$  and  $\mathcal{S}(\varphi(\tau))$ . For example, it’s true, according to the Superman comic books, that Superman is famous. Hence, one of Superman’s conformal properties is that of being famous. But it’s also true, as a matter of normal predication, that Superman is famous. Hence, I may both believe and suppose that Superman is famous. And more generally, I may believe  $\varphi(\tau)$  and suppose  $\varphi(\tau)$ .

Thirdly, the distinction allows Rynasiewicz et al. to account for the appearance of truth-

---

<sup>19</sup>See Nichols (2006), 248 for discussion.

value gaps. The representation of a supposition will usually be incomplete, in the sense that there will be some  $\varphi$  such that  $\varphi(\tau) \notin S$  and  $\neg\varphi(\tau) \notin S$ . In general, for some conformal properties, it will be neither true nor false that an object of supposition has those properties. To take an example from fiction, in the Superman comic books, it is neither true nor false that Superman has an odd number of hairs. To take a mathematical example, the continuum hypothesis (CH) is neither true nor false in ZF. Hence, there are truth-value gaps when it comes to the conformal properties of an object of supposition. More accurately, there are gaps when it comes to what is true according to the supposition, and the appearance of truth-value gaps is explained by these truth-in-supposition gaps. However, objects of supposition, just like empirical entities, are complete when it comes to their normal properties, and there are no truth-value gaps when it comes to normal predication.

Fourthly, the distinction allows Rynasiewicz et al. to account for the comparison of objects of supposition with one another, and with empirical entities. For example, we might take two objects of supposition and compare their normal properties, as we do when we claim that Superman is more well-known than the set of all sets that do not contain themselves as members. We might take two objects from two distinct suppositions and compare the conformal properties of those objects, as we do when we claim that Superman is stronger than Sherlock Holmes. We might compare the conformal properties of an object of supposition with the normal properties of an empirical entity, as we do when we claim that Superman is stronger than Tom Cruise. And finally, we might compare the normal properties of an object of supposition with the normal properties of an empirical entity, as we do when we claim that Tom Cruise is more well-known than the set of all sets that do

not contain themselves as members.

Finally, the distinction between normal and conformal predication commits Rynasiewicz et al. to what they call “fine-grained identity” (Rynasiewicz et al. (2010), 6–7). This amounts to the claim that an object of supposition has the conformal properties that it does *necessarily*. Moreover, Rynasiewicz et al. claim that if  $\mathcal{S}(\varphi(\tau))$  is true, then it is analytically true. Their example is that ‘Superman has a red cape’ is analytically true (Rynasiewicz et al. (2010), 7). With regard to empirical entities, we may be able to conceive of the same entity as having normal properties different from those it does have. We can, for example, suppose that Aristotle was not born in Stagira. But we cannot do this with objects of supposition and their conformal properties. The Superman of the comic books was killed by Doomsday, while the Superman of the Max Fleischer animated shorts was not. Hence, the Supermen of the shorts and the comic books are distinct objects of supposition. If we suppose of the Superman of the Max Fleischer animated shorts, that he was killed by Doomsday, we thereby generate a distinct object of supposition. Likewise, in the mathematical case, we can take  $\mathbb{R}$  with CH and with  $\neg$ CH. In the former case, but not in the latter case,  $\mathbb{R}$  has the smallest possible cardinality that is greater than the cardinality of  $\mathbb{N}$ . Hence,  $\mathbb{R}$  with CH and  $\mathbb{R}$  with  $\neg$ CH differ in their conformal properties, and therefore count as distinct objects of supposition. To be sure, we can often get by without distinguishing objects in this way. But this is because oftentimes, we do not look for these discrepancies, or because oftentimes, the discrepancies don’t matter for our purposes.

Up until this point, I have sketched the details of suppositionalism in terms of examples from the domains of mathematics and fiction. There is a relatively straightforward way in

which these details map onto the scientific domain as well, and it is to the latter that I now turn.

## 2.5 HYPOTHETICAL ENTITIES IN THE SCIENCES

The central claim of this dissertation is that suppositionalism yields a view of hypothetical entities that is better than the extant views discussed in chapter 1. I now turn to introducing that view. This will involve sketching another important aspect of suppositionalism, namely, what Rynasiewicz et al. have called the “appropriation model,” which accounts for the ‘discovery’ of hypothetical entities (Rynasiewicz et al. (2010), 10).<sup>20</sup>

### 2.5.1 SUPPOSITIONALISM AND SCIENCE

As before, Rynasiewicz et al. develop the suppositionalist view in the scientific domain in terms of props, interpretive engines, and representations of suppositions. They claim that “[i]n scientific practice, the prop may be a set of phenomena to be explained, the interpretive engine rely on abductive inference, and the resulting supposition include terms for newly introduced, purported entities (hypothetical entities)” (Rynasiewicz et al. (2010), 6). They explicitly take their cue from Walton here. To use one of Walton’s examples, when children engage in pretense, they may use tree stumps as props, and declare that all stumps are bears (Walton (1990), 24). They may thereby generate a supposition. In the same way, a scientist may use a certain type of phenomenon as a prop, and declare that all phenomena of this type are effects of a hypothetical entity. This may also be a means of

---

<sup>20</sup>I will discuss the need for the scare quotes below.

generating a supposition.

In this case, there is an important difference between what Rynasiewicz et al. have to say about science, and what they have to say about fiction. Suppositionalism in the fictional domain is supposed to account for our engagement with works of fiction that someone has already created. Suppositionalism in the scientific domain, however, is supposed to account for how scientists create explanations that may posit hypothetical entities, or attempt to assert that there is some hypothetical entity responsible for some set of phenomena. But a closer analogy between the two domains would suggest that suppositionalism in the scientific domain would not attempt to account for the *creation* of works that result from scientific activity (i.e., articles, books, and lectures). Instead, a closer analogy would suggest that suppositionalism would attempt to account for our *engagement* with those works, in the sense of answering questions about how we do or should understand scientific discourse.

To be sure, the analogies that hold among the domains needn't be perfect. The analogy between the fictional and mathematical domains is perhaps as loose as that between the fictional and scientific domains. And this is not necessarily a problem. After all, these are distinct domains, in which case some differences are to be expected. And analogies aside, Rynasiewicz et al.'s treatment of the scientific domain may be used to elucidate an important part of scientific activity, namely, the creation of works that result from scientific activity.

Nonetheless, I will develop the details of the suppositionalist view of science in a different way, using the analogy with fiction as my starting point. In this case, the phenomenon to

account for is not the creation of scientific works like articles, books, and lectures, but our understanding of those works. These works, then, are the props. The interpretive engine is similar to that in the fictional case, and includes some combination of deductive, inductive, and interpretive principles. And finally, the representation of the supposition will, once again, be the (possibly fuzzy) set of sentences that are true according to the supposition.<sup>21</sup>

Importantly, not all scientific works are candidates for a suppositionalist treatment, and so only a subset of these works can function as props. There are many works in which scientists put forward hypotheses, attempt to explain phenomena in terms of theories that may not be all that well-supported, and, in general, speculate in a variety of ways. I'll call a work 'speculative' if it contains terms that were or are HE terms. Speculative work can be understood as triggering supposition\* (in the sense of mere supposition and vivid imagination), and this provides the motivation for a suppositionalist treatment.<sup>22</sup> But not all scientific work is speculative in this sense. For example, there are papers that report experimental results without speculating about the explanation of those results. Such papers are not candidates for a suppositionalist treatment, and therefore cannot function as props. In what follows, I'll limit my attention to speculative works, which *can* function as props.<sup>23</sup>

Moving on to the interpretive engine, we've already seen that there are at least a few issues that one must confront when it comes to specifying the exact combination of deductive, inductive, and interpretive principles in the fictional case. First of all, the interpretive engine cannot involve an explosive logic, since some works of fiction contain contradic-

---

<sup>21</sup>In what follows, I will drop the qualification about fuzzy sets.

<sup>22</sup>One may object that suppositionalism commits one to the claim that scientists are consciously and explicitly engaging in supposition, and that this is contentious. I'll consider such an objection below.

<sup>23</sup>This is not to say that speculative works are the only ones susceptible to a suppositionalist treatment. Works that deal with convenient fictions may also be susceptible, but I won't develop this view here.

tions. Secondly, it's sometimes unclear how much background knowledge we can import into the representation of a supposition. And thirdly, the set of interpretive principles involved is a matter of some debate among critics.

Unfortunately, these issues aren't all that more tractable in the scientific case than they are in the fictional case. First of all, just like works of fiction, works of science can contain contradictions.<sup>24</sup> Hence, an explosive logic may be inappropriate for determining what does and does not belong in the supposition. On the other hand, the fact that a work contains a contradiction may be reason enough to reject any hypotheses advanced in that work, and an explosive logic would give a way to state a *reductio* against such hypotheses. Secondly, the importation of background knowledge can sometimes be unclear, even in the scientific case. It may be the case that two articles that deal with similar or related phenomena are inconsistent with one another. In such a case, we seem to have reasons for and against importing material from the second article into the representation of the supposition generated by the first. Thirdly, it's a contentious issue as to what goes into the set of interpretive principles involved in our engagement with the works of science. I will consider one such case, which regards the role of scientists' intentions, below.

The fact that these issues are just as tractable in the scientific case might seem to be bad news for suppositionalism. But rather than being bad news for suppositionalism in particular, this is just the particular guise that some outstanding problems in philosophy take for the suppositionalist. Just as in the fictional case, there are two different ways of understanding the interpretive engine. First of all, there is the normative view about the correct

---

<sup>24</sup>Peter Vickers, for example, has emphasized the inconsistency of science. See, e.g., Vickers (2010).

combination of principles that *should* go into the interpretive engine. Secondly, there is the view of what we *actually do* when we engage with a work of science. Just as in the fictional case, specifying the exact combination of deductive, inductive, and interpretive principles that are/should be involved in our engagement with a work of science is an outstanding issue. The suppositionalist will frame these issues in terms of the interpretive engine, but they appear in different guises elsewhere; they are issues, not just for suppositionalism, but for philosophy more generally.

In what follows, I'll continue to make the simplifying assumption that there is one interpretive engine, and one representation of a supposition that is generated by that interpretive engine. Though, to be sure, it will be clear in what follows that suppositionalism depends on some non-trivial claims about what kinds of principles go into the interpretive engine.

### 2.5.2 HYPOTHETICAL ENTITIES AS OBJECTS OF SUPPOSITION

The representation of a supposition in the scientific domain is, once again, a set  $S$  of sentences that are true according to the supposition; and, just as in the other domains, we can understand objects of supposition in terms of the representation of a supposition. We take the set of terms  $T$  that occur in the sentences in  $S$ , and partition them into equivalence classes, where co-reference is our equivalence relation. The terms in  $T$  name objects of supposition, which we can model as equivalence classes of co-referring terms.

For Rynasiewicz et al., hypothetical entities are objects of supposition, but not all objects of supposition are hypothetical entities. They point out that, as in the mathematical and fictional cases, the representation of a supposition in the scientific case “may contain

previously uninterpreted terms,” and these are, at least in some cases, “terms for newly introduced, purported entities (hypothetical entities),” i.e., HE terms (Rynasiewicz et al. (2010), 6). The qualification ‘at least in some cases’ is necessary, since Rynasiewicz et al. are explicit that some of these previously uninterpreted terms may name members of another class of objects of supposition, namely, convenient fictions like “frictionless planes, incompressible fluids, image charges in electrostatics, epicycles and equants in Ptolemaic astronomy, valencies and quantum orbitals, Fermi spheres, Hamiltonians and Lagrangians” (Rynasiewicz et al. (2010), 10). When we partition the set  $T$  of terms that occur in the sentences in  $S$  into equivalence classes based on co-reference, and one HE term is an element of an equivalence class, then that equivalence class models a hypothetical entity.

Hypothetical entities, like other objects of supposition, have both conformal and normal properties. To take an example, caloric has the conformal property that it combines with oxygen to form oxygen gas. And it has the normal property of being employed by Antoine Lavoisier to explain changes in state.

Once again, we can get clearer on objects of supposition by applying Parsons’ distinction between immigrants and natives. Any term that occurs in a sentence in  $S$  refers either to an immigrant or a native, and immigrants and natives are both objects of supposition. The hypothetical entities that are introduced in a work, along with any convenient fictions that are introduced, are natives. The rest are immigrants, and immigrants can include hypothetical entities and convenient fictions introduced elsewhere, as well as empirical entities. We can treat the former cases as analogous to those cases in fiction in which a fictional character from another work shows up—e.g., when the Predator appears in a Superman

comic book. We can treat the latter cases as analogous to those cases in fiction in which an empirical entity shows up—e.g., when Ronald Reagan appears in a Superman comic book.

The same term can, on different occasions, refer to a hypothetical entity, to an immigrant that is not a hypothetical entity, or to an empirical entity. An example should make this clear. In the later years of the nineteenth century, Joseph Larmor's use of the term 'electron' can be understood as referring to a hypothetical entity, which is an object of supposition. In Larmor's work, 'electron' is an HE term, since he used the term before the discovery of the electron. In the early years of the twentieth century, after the discovery of the electron, Max Abraham and H. A. Lorentz put forward two competing theories of that particle.<sup>25</sup> According to Abraham's theory, the electron is a rigid sphere with either a uniform volume or a uniform surface charge. According to Lorentz's theory, the electron is deformable. For Abraham and Lorentz, 'electron' is not an HE term, since they used the term after the discovery of the electron. Their uses of the term 'electron' can be understood in two ways. First of all, if the works of these scientists can be used to generate suppositions, then the electron would be an immigrant in those suppositions. In that case, 'electron' may be understood as referring to an object of supposition. Importantly, this object of supposition is not a hypothetical entity, since the electron was no longer hypothetical. Secondly, since the electron was no longer hypothetical, their uses of the term 'electron' can be understood independently of the suppositions generated by their work, and can be understood as referring to an empirical entity, namely, the electron. In that case, a term like 'electron,' just like 'Ronald Reagan,' can have distinct referents.

---

<sup>25</sup>For more discussion of these two theories, see Miller (1981), 55, 70.

The notion of a term having distinct referents is something that needs to be made more precise, and there are a number of ways in which one might attempt to do so. Two proposals suggest themselves almost immediately. First of all, a term type can have tokens with distinct referents.<sup>26</sup> Secondly, there can be distinct, but homonymous, term types with distinct referents. So the sense in which they are the same terms is just the very minimal sense in which they are homonyms. Since I suspect that both proposals can be made to work, I will simply opt to develop the second without argument, and without presuming that it has any advantage over the first.

If we apply this second proposal to the example of the electron, then there are actually at least four distinct, but homonymous, term types. Larmor's 'electron<sub>1</sub>' refers to a hypothetical entity. Abraham's 'electron<sub>2</sub>' refers to an immigrant that has the conformal property of being a rigid sphere with either a uniform volume or a uniform surface charge. Lorentz's 'electron<sub>3</sub>' refers to a distinct immigrant that has the conformal property of being deformable. Finally, Abraham and Lorentz could use the term 'electron<sub>4</sub>' outside of the context of a supposition to refer to an empirical entity, namely, the electron.

### 2.5.3 THE APPROPRIATION MODEL

At this point, suppositionalism in the scientific domain may seem very implausible. After all, scientists are surely not in the business of talking about objects of supposition—they intend to explain various features of the natural world. There is something to this charge, but suppositionalism has an answer. In order to see this, I will now discuss what

---

<sup>26</sup>This proposal is similar to a theory of reference that Philip Kitcher develops, according to which one token of, say, 'dephlogisticated air' may refer to oxygen, while another may fail to refer (Kitcher (1993), 100–103).

Rynasiewicz et al. call the “appropriation model,” which has to do with the ‘discovery’ of hypothetical entities (Rynasiewicz et al. (2010), 11). In short, the appropriation model gives us a way of bridging the gap between a supposition and the natural world.

Hypothetical entities are objects of supposition, not empirical entities, and so it is impossible for scientists to literally discover them. Because of this, Rynasiewicz et al. put forward the appropriation model of discovery. According to Rynasiewicz et al., “[t]he hypothetical entity itself is not detected during the course of discovery, but rather a concrete empirical entity whose configuration of normal properties sufficiently matches certain of the conformal properties of the hypothetical entity in such a way as to give someone cause to make an association between the two” (Rynasiewicz et al. (2010), 10). In such a case, to a first approximation, the HE term that originally referred to a hypothetical entity (an object of supposition) is appropriated, so that it is taken to refer to an empirical entity.<sup>27</sup> I will continue to write about the discovery of hypothetical entities. But it should be emphasized that I intend a non-literal construal of such claims in terms of the appropriation model, since such entities cannot literally be discovered.

Some examples should shed light on the basic idea. Rynasiewicz et al. discuss the example of the neutron (Rynasiewicz et al. (2010), 11). In this case, the term ‘neutron’ originally referred to a hypothetical entity, but was later appropriated by James Chadwick to refer to the particle he is credited with discovering. Other cases are a bit more complicated. Around 1869, Dmitri Mendeleev employed his periodic law to predict the existence of a previously unknown element, which he called ‘ekasilicon.’ Ekasilicon, then, was a

---

<sup>27</sup>Here and elsewhere I will speak of first approximations, since I will develop these ideas in terms of homonymous term types below.

hypothetical entity, and in 1886, Clemens Winkler discovered a new element (an empirical entity) whose normal properties matched a number of the conformal properties of ekasilicon. But this element came to be known as germanium. This might seem to tell against the appropriation model, since ‘ekasilicon’ was not preserved as the name for the new element. But Winkler’s discovery was taken to bear out Mendeleev’s periodic law, and made it possible to take ‘ekasilicon’ as a term that refers to germanium, an empirical entity. This latter claim is all that the appropriation model is committed to.

Although Rynasiewicz et al. are not explicit about this point, we can also use the appropriation model to understand cases in which we ‘discover’ that a hypothetical entity does not exist. These are really cases in which we discover that there is no empirical entity whose normal properties sufficiently match the hypothetical entity’s conformal properties in such a way as to give someone cause to make an association between the two. For example, the celestial spheres of ancient astronomy were hypothetical entities at one point. And the discovery that ‘they’ don’t exist is really the discovery that there is no empirical entity whose normal properties match the conformal properties of the celestial spheres, *qua* objects of supposition. The role of the term ‘celestial sphere’ is now different, since it has been eliminated from scientific discourse. To a first approximation, we can understand this in terms of a failure of appropriation, according to which an HE term that originally named an object of supposition is not appropriated to name an empirical entity.

Rynasiewicz et al. develop their view by appeal to the notion of a sufficient match between the normal properties of an empirical entity and the conformal properties of a hypothetical entity, and I have followed them in this. One might then take them as claiming

that there is some number or combination of properties that is sufficient for there to be such a match. But a glance at the history of science suggests that any proposal for sufficiency here will face some serious challenges. To take one example, the atom was, at one point, a hypothetical entity, one of whose conformal properties was indivisibility, as the name indicates. But this did not stop scientists from appropriating the term ‘atom’ to refer to an empirical entity that is divisible. To take another example, scientists did not appropriate the term ‘ether’ to refer to the electromagnetic field, once it was accepted that there is no preferred frame of reference. Obviously, the cases of the atom and the ether are quite complicated. But even a cursory look at the cases suggests that a proposal for sufficiency faces serious difficulties. After all, in the case of the atom, indivisibility is a central conformal property of the hypothetical entity, while the empirical entity lacks this normal property. But that didn’t stop scientists from appropriating the term ‘atom.’ And in the case of the ether, it’s not clear that nineteenth-century wave theorists would classify the property of being-a-preferred-frame-of-reference as a central conformal property of the ether. But it seems that this was enough to stop scientists from appropriating the term ‘ether.’

How, then, should we understand the notion of a sufficient match between the normal properties of an empirical entity and the conformal properties of a hypothetical entity? Rynasiewicz et al. are clear that the match has to be sufficient “to give someone cause to make an association between the two” (Rynasiewicz et al. (2010), 10). So, instead of viewing this in terms of a sufficient number or combination of properties, we should take this claim about sufficiency as a claim about what scientists, in fact, judge to be sufficient. In this case, there is no general, across-the-board notion of sufficiency that will do the work required

of it. What is needed is a careful study of the cases in question, via an examination of the relevant primary source material. This will give us the reasons that individual scientists put forward, which sometimes reflect the influence of what we may take to be relatively contingent factors. It's possible that some may find this proposal philosophically lacking. But this notion of sufficiency allows us to use the appropriation model as a tool for examining the history of science to see what we can learn about cases in which hypothetical entities were and were not discovered. A more robust philosophical thesis about sufficiency would prejudice these very issues in advance of such an examination.

Since HE terms refer to objects of supposition, Rynasiewicz et al. claim that “theories trading in hypothetical entities are not even candidates to be true or false” (Rynasiewicz et al. (2010), 11). It's important to be clear about what Rynasiewicz et al. mean by this. Take a representation of a supposition  $S$ , and consider some sentence in  $S$  that predicates some property of a hypothetical entity. Some examples might include ‘Electrons are charged’ and ‘Caloric is conserved.’ To be sure, such sentences are candidates for being true or false, and can be understood as making claims to truth *simpliciter*, i.e., truth outside of a supposition. But such sentences will almost always be false. This is because the hypothetical entity *qua* object of supposition will most likely not have the property in question as a matter of normal predication. After all, the properties that a hypothetical entity has according to the supposition are conformal properties.<sup>28</sup> Rynasiewicz et al. are therefore most naturally understood as claiming that the sentences in  $S$  have an implicit

---

<sup>28</sup>However, there may be exceptions here, as when a hypothetical entity has a property both as a matter of normal and conformal predication. In other words, some analogue of the ‘Superman is famous’ example may exist in the scientific domain, though it's difficult to come up with such an example.

‘according to the supposition’ operator ( $\mathcal{S}$ ) prefixed to them, and therefore do not purport to make claims about what is true *simpliciter*. This is the sense in which such theories are not candidates for being true or false.

Instead, they claim that, in cases in which a hypothetical entity is discovered, we can speak of the theory being *vindicated*. Since Rynasiewicz et al. don’t explicitly address cases in which appropriation fails, we can introduce a bit more terminology and say that the theory has been *discredited* in such cases. A bit more precisely, a theory that has been vindicated has at least one HE term that has been appropriated, and a theory that has been discredited has at least one HE term that has failed to be appropriated. Moreover, the appropriation in question must catch on in the scientific community. In this case, we must distinguish between an attempted appropriation, considered as an act of an individual scientist, and an appropriation that has the backing of the community. A corresponding distinction applies to cases in which appropriation has failed—we must distinguish between an individual’s claim of failure, and a claim that has the backing of the community. Given this, the same theory may be both vindicated and discredited, if some HE terms have been appropriated, while others have failed to be appropriated. This may sound a bit odd, and the source of the oddness is the fact that, as I have defined them, vindication and discreditation come in degrees. A theory that has five HE terms that have been appropriated is more vindicated than one that has only three, and likewise for theories that are discredited.

The way in which we understand a theory before it is vindicated is different from the way in which we understand that theory after it is vindicated. The same goes for a theory before and after it is discredited. But what, exactly, does this difference amount to? For the

sake of giving a sufficiently worked-out picture of the appropriation model, I will attempt to answer this question on the basis of the idea, discussed above, that there can be distinct, but homonymous, term types with distinct referents. In short, the answer to the question is that, if a theory has been vindicated, then it is more truth-evaluable than it was before it was vindicated. Likewise, if a theory has been discredited, then it is more truth-evaluable than it was before it was discredited.

A full understanding of what it means to vindicate and discredit a theory requires some understanding of what a theory is. I propose that, just as the representation of a supposition in mathematics is a formal theory (in the sense that  $T = \text{Cn}(T)$ ), the representation of a supposition in science is a scientific theory. Though, to be sure, not all scientific theories are representations of some supposition. Since the representation of a supposition is a set of sentences, one may object to this proposal on the grounds that it is too close to the so-called syntactic or received view of theories, which we all presumably have very good reason to reject since the fall of logical positivism. But philosophers of science now seem to be realizing that the received view is not the straw man that one often sees presented in the literature, and that the semantic view is not obviously superior.<sup>29</sup> In that case, at the very least, the nature of scientific theories is still very much open to debate, and it is not clearly objectionable to advocate some version of the received view of theories. In what follows, I will assume that all theories are sets of sentences. The only argument that I will give in favor of this view of theories is that, unlike the semantic view, it fits well with the appropriation model, which gives us a good way of understanding scientific discourse

---

<sup>29</sup>See, for example, Halvorson (2012) and Lutz (2012).

concerning hypothetical entities.

At this point, it will be necessary to get clearer on appropriation. HE terms are terms that name hypothetical entities, which are objects of supposition. Therefore, they are not putatively referring expressions that have putative reference to empirical entities. But it is possible for them to be appropriated so as to refer to empirical entities. If we have a successful appropriation, then we have the introduction of a homonymous theoretical term (not an HE term) which is a putatively referring expression that has putative reference to an empirical entity, and, in fact, succeeds in referring to an empirical entity. If we have a failed appropriation, then we have the introduction of a homonymous theoretical term (not an HE term) which is a putatively referring expression that has putative reference to an empirical entity, and, in fact, altogether fails to refer to an empirical entity. Neither theoretical term is an HE term since neither term names a hypothetical entity. In this case, before and after appropriation (or failure of appropriation), we have two different term types.

Now it will be useful to introduce the notion of a *homonymous theory*. If we take a representation of a supposition, and make two sorts of replacements, the result is a homonymous theory. First of all, if appropriation has either succeeded or failed, we replace HE terms with the homonymous terms that putatively refer to empirical entities. Secondly, we replace terms that name immigrant empirical entities (a certain kind of object of supposition) with the homonymous terms that refer to the actual empirical entities. These two kinds of replacement give us a kind of mapping from theories *qua* representations of suppositions to theories *qua* homonymous theories. I call both of these ‘theories,’ but they may be more accurately characterized as different formulations of the same theory.

Rynasiewicz et al. claim that theories that contain HE terms are not candidates for being true or false, and so I introduce homonymous theories as a way of yielding theories that are candidates for being true or false, at least in part. In this way, the suppositionalist can secure some connection between a scientific theory and the world, and accommodate the fact that scientists do mean to be describing the world. A homonymous theory may not be true or false as a whole, since it may still be speculative, in the sense that it contains some HE terms. In that case, following Rynasiewicz et al., the theory as a whole may not be a candidate for being true or false. But it can be helpful to consider the sentences that are elements of the homonymous theory in isolation from one another. If a specific sentence in a homonymous theory contains no terms that refer to objects of supposition (whether hypothetical entities, convenient fictions, or immigrants), then that sentence is what I will call *truth-evaluable*. Truth-evaluable sentences should be understood as making claims to truth *simpliciter*, not truth according to a supposition. If a sentence is truth-evaluable, it must be capable of having a truth-value, though as a matter of fact, it may lack a truth-value.<sup>30</sup> A theory  $T_1$  is *more truth-evaluable* than a theory  $T_2$  just in case  $T_1$  contains more truth-evaluable sentences than  $T_2$ .

To be sure, this terminology leaves something to be desired, since, for any sentence  $\varphi$  in a representation of a supposition  $S$ , where  $\varphi$  has terms that refer to objects of supposition,  $\varphi$  can be true or false, and  $\mathcal{S}(\varphi)$  can be true or false. But  $\varphi$  will most likely be false as a matter of normal predication, and  $\mathcal{S}(\varphi)$  won't even purport to tell us about anything

---

<sup>30</sup>This would be the case if one were to adopt a semantics according to which sentences that contain non-referring terms lack a truth-value. Though, to be sure, suppositionalists need not commit themselves to such a semantics. There is therefore some similarity between my notion of truth-evaluable, and van Fraassen's notion of literal construal (van Fraassen (1980), 10). Indeed, truth-evaluable sentences are sentences that it would be appropriate to construe literally.

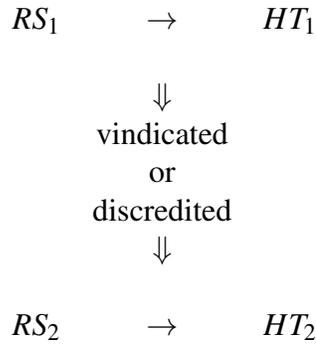


Figure 2.1: Vindicated and discredited theories, and their relation to the world.

outside of the supposition.  $\varphi$  and  $\mathcal{S}(\varphi)$  will not count as truth-evaluable in my sense, which aims to capture those sentences that it is best to construe as making claims to truth *simpliciter*, i.e., truth outside of a supposition.

Importantly, a vindicated/discredited theory will be more truth-evaluable than it was before it was vindicated/discredited. We can see this by examining Figure 2.1.  $RS_1$  is the representation of a supposition before it has been vindicated or discredited, and  $HT_1$  is the homonymous theory obtained from  $RS_1$ . Likewise,  $RS_2$  is the representation of a supposition after it has been vindicated or discredited, and  $HT_2$  is the homonymous theory obtained from  $RS_2$ . The single arrows indicate the mapping that takes one from a representation of a supposition to a homonymous theory by means of replacing various terms, as I described above. In the case in which the theory is vindicated,  $HT_2$  will have a term that refers to an empirical entity, whereas  $HT_1$  will have an HE term. In the case in which the theory is discredited,  $HT_2$  will have a term that fails to refer to an empirical entity, whereas  $HT_1$  will have an HE term. In both cases, where  $HT_1$  has an HE term that refers to an object of supposition,  $HT_2$  will have a putatively referring expression that has putative reference to an empirical entity. In short, there are HE terms in  $HT_1$  that have been replaced by pu-

tatively referring expressions in  $HT_2$ . Now let's consider the sentences in  $HT_1$  such that the only terms that refer to objects of supposition are a given type of HE term that will be replaced by a homonymous term in  $HT_2$ . These sentences in  $HT_1$  are not truth-evaluable, whereas the counterparts of those sentences in  $HT_2$  are truth-evaluable. Hence,  $HT_2$  is more truth-evaluable than  $HT_1$ .

We're now in a position to answer the objection stated above. The charge was that scientists are surely not in the business of talking about objects of supposition—they intend to explain various features of the natural world. To answer this charge, we can simply point to the notion of a homonymous theory. The truth-evaluable sentences in a homonymous theory are best construed as making claims to truth *simpliciter*, and therefore give us a connection to the features of the natural world that scientists are concerned to explain.

The last point of the appropriation model that I wish to discuss deals with the issue of dead hypotheticals, which I discussed briefly at the beginning of chapter 1. I claimed that hypothetical entities have a shelf-life that results from the fact that scientists, at some point, stop entertaining hypotheses because they are either confirmed or rejected. As a result, caloric is not a hypothetical entity today, and 'caloric' is no longer an HE term today. Instead, caloric is a *dead hypothetical*, and 'caloric' is what I will call a *dead HE term*. I will now argue that, after appropriation has failed, we need dead HE terms in addition to homonymous terms that have putative reference to empirical entities and, in fact, fail to refer. Rynasiewicz et al. claim that theories trading in hypothetical entities (objects of supposition) are not candidates for truth or falsity, and so I have introduced homonymous terms in order to make it the case that theories can be candidates for truth

and falsity. But it's also helpful to be able to talk about dead hypothetical entities as objects of supposition, so that there can be some sense in which claims like 'Caloric is conserved' are correct, and this is the reason for introducing dead hypotheticals and dead HE terms. Dead hypotheticals, like caloric, are objects of supposition, and they have the conformal properties that caloric had while it was still hypothetical. Dead HE terms, like 'caloric,' refer to those objects of supposition. They differ from hypothetical entities and HE terms because no scientists are entertaining hypotheses that involve them, and so appropriation is no longer possible. One way in which to conceive of the failure of appropriation, then, is in terms of the death of hypothetical entities and HE terms.

It should now be clear how the suppositionalist view of hypothetical entities differs from the extant views discussed in chapter 1. By way of review, and for the sake of clarity, the main claims regarding the semantics of scientific discourse that suppositionalism is committed to are as follows:

(S<sub>1</sub>) Theoretical terms are either putatively referring expressions that have putative reference to empirical entities, or else they refer to objects of supposition.

(S<sub>2</sub>) Representations of suppositions are theories capable of being vindicated or discredited, but they are not best understood as candidates for being true or false *simpliciter*.

(S<sub>3</sub>) Homonymous theories contain truth-evaluable sentences that are best understood as capable of being true or false *simpliciter*.

S<sub>HE</sub> HE terms refer to hypothetical entities, and hypothetical entities are objects of supposition.

2.5.4 THE ROLE OF INTENTIONS

I haven't yet said anything about the intentions of scientists, and one might be tempted to think that my neglect of intention leads to a fatal objection to the suppositionalist view of science. The basic idea is that scientists have intentions, that they intend to be talking about the natural world, and that the semantics of scientific discourse must take the intentions of scientists into account. However, the semantic claims that I've just outlined in (S<sub>1</sub>)–(S<sub>3</sub>) and S<sub>HE</sub> do not accommodate the intentions of scientists. Scientists surely don't intend to be talking about objects of supposition, and they don't intend to be putting forward representations of suppositions as theories. My final goal in this chapter is to respond to this objection.

At the very least, it is not obvious that every scientist, at every moment, intends to be talking about the natural world. Scientists sometimes entertain various theories and hypotheses without claiming that such theories and hypotheses truly describe the world.<sup>31</sup> In that case, the objection may presuppose a false claim about the intentions of scientists. But the objection still has force, insofar as some scientists, some of the time, do intend to be talking about the natural world when they put forward their hypotheses. Moreover, even those scientists who entertain theories without accepting them as true would presumably claim that the theories in question are at least candidates for being true or false, which is something that a suppositionalist would deny.

The objection may be based on another false presupposition. It presupposes that, according to the suppositionalist, scientists are talking about objects of supposition. But if

---

<sup>31</sup>For example, in chapter 3, I will discuss how Humphry Davy entertained the oxygen and phlogiston theories of acidity without accepting either.

we keep in mind the analogy with works of fiction, this is not true. Just as in the fictional case, in the scientific case, the author of a work does not make claims about objects of supposition. It's only once we use the work as a prop to generate a supposition that we can talk about objects of supposition. So, strictly speaking, suppositionalists are not committed to the claim that scientists are talking about objects of supposition. But now the suppositionalist seems to be committed to the claim that scientists and authors of works of fiction are engaged in similar activities, even though the intentions of the former may be very different from the intentions of the latter. So once again, the objection still has force.

In order to respond to the objection more fully, then, I will admit that scientists do intend to be talking about the natural world, but deny that this poses a problem for the suppositionalist. In general, the suppositionalist is not committed to any claim about scientists' intentions. More specifically, the suppositionalist is not committed to the claim that scientists intend to be engaging in supposition. Instead, the suppositionalist claims that the best way to understand speculative scientific works is by using them as props to generate suppositions. And, as a matter of fact, one can use any bit of text as a prop to generate a supposition, if one so desires, provided that that bit of text meets some minimal requirements of grammaticality. One can start with a sonnet, a passage from the Bible, or even a newspaper article, and use it as a prop to generate a supposition. And we can do this even if doing so does not accommodate the intentions of those who produced those texts, as it might not in the case of the writers of the Bible and the journalists who produce newspaper articles. It is debatable whether this method is the best way in which to understand some bit of text, and my goal in chapter 4 will be to show that this is the case when it comes to

speculative scientific works. The important point for now is that the focus, for the suppositionalist, is on the work produced, as opposed to the intentions of those who produced the work.

Suppositionalism, then, depends on the possibility of a semantics of scientific discourse that is independent of the intentions of scientists. At this point, one may wonder whether, and to what extent, such a semantics is possible. In order to make this view plausible, I now turn to a discussion of Irene Heim's file change semantics (FCS), and some analogies between Heim's account and the suppositionalist view. In short, FCS, like many semantic theories, makes it the case that the semantics of an utterance is at least partially independent of the intentions of the speaker of that utterance. I will argue that, given the analogies between FCS and suppositionalism, the latter is no more objectionable than the former when it comes to the charge of ignoring the intentions of speakers.

Heim begins by considering what she, following others, calls the "familiarity theory of definiteness," according to which:

A definite is used to refer to something that is already *familiar* at the current stage of the conversation. An indefinite is used to introduce a *new* referent.

(Heim (1983), 164)

She notes that, insofar as this theory is framed in terms of referents, it will be of no help when terms fail to refer. She goes on to discuss an attempt by Lauri Karttunen to defend something like the familiarity theory by replacing reference with his notion of "discourse reference" (Heim (1983), 165). The basic idea is that a noun phrase that lacks a referent

1	2
- is a woman	-is a dog
-was bitten	-bit 1
by 2	

Figure 2.2: File  $F_1$ : a two-card file.

can still have a discourse referent, and the difference between definite and indefinite noun phrases can be explained by appeal to the familiarity or novelty of discourse referents.

Heim proposes that Karttunen’s discourse referents be identified with what she calls “file cards”—hence the name “file change semantics” (Heim (1983), 167). She explains her file-keeping metaphor as follows:

A listener’s task of understanding what is being said in the course of a conversation bears relevant similarities to a file clerk’s task. Speaking metaphorically, let me say that to understand an utterance is to keep a file which, at every time in the course of the utterance, contains the information that has so far been conveyed by the utterance. (Heim (1983), 167)

She goes on to make this idea clear by means of an example (Heim (1983), 167–169).

Consider the following three sentences:

- (a) A woman was bitten by a dog. (b) She hit it. (c) It jumped over a fence.

The listener starts with an file  $F_0$  that contains no file cards. After (a) has been uttered, the listener introduces two file cards 1 and 2, which she uses to keep track of the information communicated in (a). She thus goes from  $F_0$  to  $F_1$ , a two-card file (see figure 2). After (b) has been uttered, the listener updates the two file cards in  $F_1$  to include the information

1	2
- is a woman	-is a dog
-was bitten	-bit 1
by 2	-was hit by 1
-hit 2	

Figure 2.3: File  $F_2$ : a two-card file.

1	2	3
- is a woman	-is a dog	-is a fence
-was bitten	-bit 1	-was jumped
by 2	-was hit by 1	over by 2
-hit 2	-jumped over 3	

Figure 2.4: File  $F_3$ : a three-card file.

communicated in (b). She thus goes from  $F_1$  to  $F_2$ , another two-card file (see figure 3). Finally, after (c) has been uttered, the listener updates the two file cards in  $F_2$ , and introduces a third card 3, in order to keep track of the information communicated in (c). She thus goes from  $F_2$  to  $F_3$ , a three-card file (see figure 4). More generally, Heim’s version of the familiarity theory of definiteness adopts something like the following rule:

For every indefinite, start a new card. For every definite, update an old card.

(Heim (1983), 168)<sup>32</sup>

Heim also explains how files can connect to the world. In short, a file  $F$  of  $n$  cards is true if there is an  $n$ -tuple that satisfies the cards in  $F$ ; otherwise,  $F$  is false (Heim (1983), 170–171).

While FCS and the suppositionalist view of science are distinct, there are many ways in which they are analogous. If we start with some speculative scientific work, FCS generates a file, while suppositionalism generates a supposition. The elements of the file are file cards,

<sup>32</sup>To be sure, this rule is a first approximation, and Heim goes on to develop a more sophisticated account later in her paper.

and these bear some similarity to the objects of supposition that the terms in the representation of a supposition name. The bits of information on a file card are analogous to the conformal properties of an object of supposition. According to FCS, the discourse referents of a speaker's utterance are file cards, and this is so regardless of the speaker's intentions. Likewise, according to suppositionalism, scientists who introduce hypothetical entities are putting forward objects of supposition, regardless of what their intentions are. According to FCS, it is a further question whether anything satisfies the file. Likewise, according to suppositionalism, it is a further question whether some empirical entity's normal properties sufficiently match the conformal properties of an object of supposition. According to FCS, the relationship between an utterance and its referents is mediated by a file—utterances generate a file, and the referents are the elements of the tuples that satisfy the file cards. Likewise, according to suppositionalism, the relationship between a speculative scientific work and the world is mediated by the representation of a supposition and its homonymous theory.

The upshot of FCS, for my purposes, is that it shows that, on some level, semantics does not accommodate speakers' intentions. An essential aspect of the semantics of a noun phrase, according to FCS, is its discourse referent, namely, a file card. And this aspect is independent of the intentions of the speaker, who presumably does not intend to be talking about file cards. FCS may have its problems. But I take it that the fact that, on some level, FCS is independent of the intentions of speakers, does not count as a decisive reason to reject it. Moreover, FCS does accommodate speakers' intentions on another level, insofar as it allows room for noun phrases to have referents in addition to discourse referents.

At this point, we can answer the imagined objection to suppositionalism. In short, if FCS is not obviously objectionable for its treatment of speakers' intentions, then neither is suppositionalism for its treatment of scientists' intentions. The central claim of the suppositionalist view of hypothetical entities, couched in Heim's metaphor, is that the semantics of HE terms happens exclusively at the level of the file, and a treatment in terms of possible referents that satisfy the file cards is inappropriate until appropriation either succeeds or fails. The file level does not take scientists' intentions into account. But if FCS is on the right track, it doesn't have to. And in this case, the objection to the suppositionalist view of science is not a fatal one.

## Chapter 3

---

### *A History of a Hypothetical Entity:*

#### *The Case of the Muriatic Radical*

### 3.1 INTRODUCTION

One of the hypothetical entities of late-eighteenth-century chemistry was the muriatic radical, a hitherto unknown element that was thought to combine with oxygen to form muriatic acid. As it turns out, muriatic acid is hydrochloric acid, and contains no oxygen.

The life of the muriatic radical, *qua* hypothetical entity, took place between two events: the discovery of a means to isolate chlorine, and the eventual acceptance, by the scientific community, of chlorine's elementary nature. Ira Remsen's brief gloss on the history of chlorine makes this clear, and if students of chemistry since the end of the nineteenth century know of the muriatic radical at all, they know of it from passages like the following, taken from Remsen's *A College Text-book of Chemistry*:

In 1774 Scheele first called attention to chlorine in his treatise on the black oxide of manganese or manganese dioxide. In accordance with the ideas then prevailing, he called it "dephlogisticated muriatic acid." Berthollet suggested in 1785 that it is oxidized hydrochloric acid, and it was then regarded as con-

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

sisting of the hypothetical element, *murium*, in combination with oxygen. In 1810 Davy pointed out that the idea, previously expressed by Gay-Lussac and Thénard, that the substance is an element in the highest degree probable, and he gave it the name chlorine (from  $\chi\lambda\omega\rho\acute{o}\varsigma$ , greenish-yellow). Since that time everything learned in regard to chlorine has gone to show that it is an element.

(Remsen (1901), 126)<sup>1</sup>

‘Murium’ is another term used to name the muriatic radical (or, at least, something quite similar), though not one that seems to have come into currency until sometime after 1785. In this case, putting the term in Berthollet’s mouth is somewhat anachronistic.<sup>2</sup> But, of course, the passage is anachronistic for a much more obvious reason, namely, that Berthollet would not have conceived of Scheele’s substance as oxidized hydrochloric acid. That, after all, would imply that he thought it was made up of hydrogen, oxygen, and chlorine. This passage from Remsen’s textbook illustrates the difficulty of simultaneously explaining the history of science to students in a way that they can understand, and in a way that is faithful to the facts.

There is a wealth of detail that Remsen’s brief gloss was not intended to capture, and in this chapter, I will give a fuller account of the history of the muriatic radical. In section 2, I will discuss the application of the oxygen and phlogiston theories of acidity to what Remsen calls hydrochloric acid and chlorine, respectively. In section 3, the heart of the chapter, I will discuss Davy’s subsequent determination of the composition of hydrochloric acid, and

---

<sup>1</sup>Emphasis is the author’s unless otherwise noted.

<sup>2</sup>The term was employed by the middle of the nineteenth century, in the work of Christian Friedrich Schönbein, who argued against the view that chlorine is an element, and held that it contains oxygen and murium (Schönbein (1857), 180; Schönbein (1858), 254).

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

of the elementary nature of chlorine. I will raise some puzzles concerning Davy's work, and will argue that these puzzles can be solved by tracing Davy's complicated relationship to the phlogiston and oxygen theories. In section 4, I will briefly examine some views of the muriatic radical put forward after Davy's work was completed.

The accounts of history that come down to us in textbooks certainly serve a purpose, but as Thomas Kuhn emphasized, they can also obscure important parts of scientific activity.<sup>3</sup> Because of this, my primary aim in this chapter is to give a picture of a hypothetical entity 'in the wild,' so to speak, before examining the prospects of the extant views of hypothetical entities and the suppositionalist view discussed in the earlier chapters of the dissertation. In the remainder of the dissertation, I will draw from this history of the muriatic radical in order to raise problems for the extant views, and in order to argue in favor of the suppositionalist view.

#### 3.2 THE BIRTH OF THE MURIATIC RADICAL

The muriatic radical was born in the later years of the eighteenth century, when chemists applied the oxygen theory of acidity to two as-yet undecomposed substances, which Remsen refers to as hydrochloric acid and chlorine, respectively. Hydrochloric acid was variously referred to as acid of sea-salt, marine acid, and muriatic acid, while chlorine was, at the time, not known to be an element, and was variously referred to as dephlogisticated marine acid, oxygenated muriatic acid, and oxymuriatic acid. I will begin this section with a discussion of the oxygen theory and its application to those two substances. I will then turn

---

<sup>3</sup>See Kuhn (1996), §XI.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

to a discussion of the phlogiston theory of acidity, and its application to the two substances.

3.2.1 THE OXYGEN THEORY OF ACIDITY

In the later years of the eighteenth century, Antoine Lavoisier developed his theory of oxygen, which was simultaneously a theory of combustion and a theory of acidity. My focus will be on the latter, which was meant to shed light on the nature of a kind (acids) which was already circumscribed by a number of properties. These properties are listed by William Nicholson in his *Dictionary of Chemistry*:

1. Their taste is sour and corrosive, unless diluted with water.
2. They change blue vegetable colors to a red.
3. Most of them unite with water in all proportions; and many have so strong an attraction to that fluid as not to be exhibited in the solid state.
4. At a moderate temperature, or in the humid way, they combine with alkalis so strongly as to take them from all other substances.
5. They combine with most bodies, and form combinations attended with many interesting phenomena; upon the due explanation of which a great part of the science of chemistry depends. (Nicholson (1795), 2)

Though Nicholson wrote his *Dictionary* several years after Lavoisier had published on his own theory, chemists at the time used these properties to distinguish acids from non-acids. These properties, then, were used to delineate a kind. But this list of properties does not tell us anything about the constitution of acids, and this is what Lavoisier's theory was meant to illuminate.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

Lavoisier develops his theory in a number of papers,<sup>4</sup> but it reached maturity at least by the time he wrote his *Traité élémentaire de Chimie*, originally published in 1789, and so I will focus on his formulation of it in that work. Lavoisier discusses a number of experiments, which show that three combustible bodies combine with oxygen to form acids. A combustible body for Lavoisier is just “a body which possesses the power of decomposing oxygen gas, by attracting the oxygen from the caloric with which it was combined” (Lavoisier (1802), 111).<sup>5</sup> The combustible bodies that Lavoisier employs in these experiments are phosphorus, sulphur, and carbon, which, he claims, when combined with oxygen, form phosphoric acid, sulphuric acid, and carbonic acid, respectively. He labels the process by which these bodies are converted into acids “oxygenation,” and writes of “oxygenating” a combustible body in order to convert it into an incombustible acid (Lavoisier (1802), 110–111).<sup>6</sup>

Although Lavoisier could employ more examples, he generalizes to a theory of acidity from the three mentioned above:

In the mean time, however, the three examples above cited, may suffice for giving a clear and accurate conception of the manner in which acids are formed. By these, it may be clearly seen, that oxygen is an element common to them all, and which constitutes or produces their acidity; and that they differ from each other, according to the several natures of the oxygenated or acidified substances. We must, therefore, in every acid, carefully distinguish between the

---

<sup>4</sup>See, for example, Lavoisier (1776) and Lavoisier (1778). See Lavoisier (1783) for English translations of both.

<sup>5</sup>Translation of Lavoisier (1801), 66.

<sup>6</sup>See Lavoisier (1801), 66 for the original French.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

acidifiable base, which Mr de Morveau calls the radical, and the acidifying principle, or oxygen. (Lavoisier (1802), 114)<sup>7</sup>

In short then, on Lavoisier's theory, an acid is just an acidifiable base or radical plus oxygen. It should be emphasized that Lavoisier, along with all working chemists around this time, did not use 'base' to mean what we mean today by that term. Moreover, 'base' is not a term that is specific to the constitution of acids. Gases, for Lavoisier, are compounds of caloric and a base, in which case oxygen gas is caloric plus a base (oxygen) (Lavoisier (1802), 63).<sup>8</sup> This shows the need for Lavoisier's use of the qualifier 'acidifiable' in the above quotation. And when it comes to acidifiable bases, Lavoisier emphasizes that they needn't be simple—in fact, "there exist acids . . . having double and triple bases" (Lavoisier (1802), 170).<sup>9</sup> Likewise, 'radical' does not have the meaning that it does in today's chemistry either. Lavoisier borrows the term from Louis Bernard Guyton de Morveau, and uses it as a synonym for 'base.'

The above passage gives the basic idea behind the theory, but it is still necessary to make a couple of clarifications. First of all, it isn't the case that every substance that contains oxygen as a constituent is an acid. Lavoisier notes that "when metallic substances are oxygenated in atmospheric air, or in oxygen gas, they are not converted into acids" (Lavoisier (1802), 130).<sup>10</sup> Secondly, substances containing oxygen, and acids in particular, can be distinguished from each other in terms of the amount of oxygen they contain. That is to say, substances composed from the same elements may differ in their degree of

---

<sup>7</sup>Translation of Lavoisier (1801), 69.

<sup>8</sup>See Lavoisier (1801), 17 for the original French.

<sup>9</sup>Translation of Lavoisier (1801), 123.

<sup>10</sup>Translation of Lavoisier (1801), 83.

oxygenation. Lavoisier explains this in the following way:

The first or lowest degree of oxygenation in bodies, converts them into *oxyds*; a second degree of additional oxygenation constitutes that class of acids, of which the specific names, drawn from their particular bases, terminate in *ous*, as the *nitrous* and *sulphurous* acids; the third degree of oxygenation changes these into that division of acids, which are distinguished by the termination in *ic*, as the *nitric* and *sulphuric* acids; and, lastly, we can express a fourth, or highest, degree of oxygenation, by adding the word *oxygenated* to the name of the acid, as has been already done with the *oxygenated muriatic acid*. (Lavoisier (1802), 131–132)<sup>11</sup>

In this case, even if the base in question is capable of combining with oxygen to form an acid, it needs to combine with the right amount of oxygen in order to do so. For example, arsenic, molybdenum, and tungsten were known to be capable of forming both oxyds and acids.<sup>12</sup>

### 3.2.2 TWO UNDECOMPOSED ACIDS

There were a number of substances that were considered acids in Lavoisier's day that were yet to be decomposed, and so one could not yet prove that they contain oxygen. One such substance was variously referred to as acid of sea-salt, marine acid, and muriatic acid. It is now called hydrochloric acid. Another such substance was variously referred to as

---

<sup>11</sup>Translation of Lavoisier (1801), 84.

<sup>12</sup>See the table in Guyton de Morveau, Lavoisier, Berthollet, and Fourcroy (1788), 73.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

dephlogisticated marine acid, oxygenated muriatic acid, and oxymuriatic acid. It is now called chlorine.

Towards the end of the eighteenth century, the first substance was well-known by chemists. But, as the quotation from Remsen makes clear, the second substance was not isolated until 1774. It was in this year that Carl Wilhelm Scheele combined muriatic acid with what he called ‘manganese,’ which we now call ‘manganese dioxide,’ and isolated a new substance. He had this to say about the nature of the reaction:

The following is the theory of the solution: The manganese is first attacked by the acid, and thus we have a brown solution. The manganese, when dissolved, acquires, by means of the acid, a strong attraction for phlogiston, . . . and really attracts it from the particles with which it is combined. These particles having thus lost one of their constituent parts, and being but very loosely combined with the phlogisticated manganese, are expelled from it by the remaining muriatic acid, which has not yet suffered any decomposition, and now appear with an effervescence, as an highly elastic air; the brown colour has now disappeared, and the solution is become limpid. (Scheele (1786), 92)

As a proponent of the phlogiston theory, Scheele naturally explains the reaction in the terms of that theory. Scheele goes on to describe the properties of the highly elastic air that he isolated:

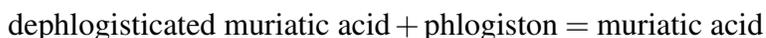
The marine acid separated from phlogiston, one of its constituent parts, unites with water in a very small quantity only, and gives it a slight acid taste: But

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

whenever it is enabled to combine with phlogiston, it assumes its former nature, and again becomes a true muriatic acid. (Scheele (1786), 92–93)

Hence, the substance has some of the properties that Nicholson lists in his dictionary, though, to be sure, it has them to a much lesser degree than most acids. Indeed, Nicholson himself notes as much in his *Dictionary* (Nicholson (1795), 27). This, however, sufficed for Scheele to label the substance an acid, and in accordance with the phlogiston theory, he called it “dephlogisticated muriatic acid” (Scheele (1786), 93). A bit more schematically, Scheele’s view of the two acids is as follows



Importantly, for Scheele, the highly elastic air that he isolated (dephlogisticated muriatic acid) is a constituent of muriatic acid, and hence, a simpler substance. One can obtain the latter by adding phlogiston to the former.

As Remsen notes, in 1785, Claude Louis Berthollet put forward another view of the nature of this reaction, one at odds with Scheele’s view. Berthollet, like Scheele, notes that given its properties, “dephlogisticated marine acid [is] almost entirely deprived of acidity” (Berthollet (1905), 15).<sup>13</sup> He goes on to explicitly apply Lavoisier’s theory of acidity to the experimental results that he obtained:

It is therefore to the vital air of the manganese, which combines with the marine acid, that the formation of the dephlogisticated marine acid is due. I ought

---

<sup>13</sup>Translation of Berthollet (1788), 279.

to state that this theory was presented and announced some time ago by M. LAVOISIER ... (Berthollet (1905), 16)<sup>14</sup>

These experiments ought to dispel any doubts which might remain as to the nature of dephlogisticated marine acid; the latter is manifestly formed by the combination of vital air with marine acid ... (Berthollet (1905), 20)<sup>15</sup>

Vital air is oxygen, Lavoisier's principle of acidity. In that case, we have an acid (dephlogisticated muriatic acid) which contains oxygen, in accordance with Lavoisier's theory. Moreover, we have a reversal of Scheele's view, since, on this view, muriatic acid is a constituent of dephlogisticated muriatic acid, and hence, is a simpler substance than the latter.

Lavoisier had a more specific hypothesis concerning the nature of what Scheele and Berthollet call 'dephlogisticated muriatic acid.' But in order to get clear on this hypothesis, it will be necessary to turn to the other acid, namely, muriatic acid. To begin with, in 1787's *Méthode de nomenclature chimique*, Guyton de Morveau applies Lavoisier's oxygen theory to the as-yet undecomposed muriatic acid:

Analogy induces us to think that the muriatic acid has an acidifiable base, as well as the carbonic, sulphuric, and phosphoric acids, which like the bases of these latter, serves to give a distinct and particular property to the produce of a combination of oxygen. We could not express this substance otherwise than by the name muriatic *radical* or muriatic radical principle, in fine, that a name

---

<sup>14</sup>Translation of Berthollet (1788), 281.

<sup>15</sup>Translation of Berthollet (1788), 284.

should not be given to an unknown substance, and that the expression should be limited to the simple property with which we are acquainted, and which is to produce this acid. We have the same caution in respect to all the other acids with which we are not yet well acquainted, some of whose bases will probably be discovered in the substances which we have already named. (Guyton de Morveau et al. (1788), 33–34)<sup>16</sup>

Guyton de Morveau mentions the three examples that Lavoisier employs in the *Traité* to generalize to his theory of acidity. And, indeed, we find a similar passage in the *Traité*, in which Lavoisier displays his confidence that his theory of acidity also covers the as-yet undecomposed muriatic acid:

Although we have not yet been able, either to compose or to decompose this acid of sea-salt, we cannot have the smallest doubt that it, like all other acids, is composed by the union of oxygen with an acidifiable base. We have therefore called this unknown substance the *muriatic base*, or *muriatic radical*, deriving this name, after the example of Mr Bergman and Mr de Morveau, from the Latin word *muria*, which was anciently used to signify sea-salt. Thus, without being able exactly to determine the constituent parts of *muriatic acid*, we design by that term a volatile acid, which retains the form of a gas in the common temperature and pressure of our atmosphere; which combines with great facility, and in great quantity, with water, and whose acidifiable base adheres so very intimately with oxygen, that no method has hitherto been devised for

---

<sup>16</sup>Translation of Guyton de Morveau, Lavoisier, Berthollet, and Fourcroy (1787), 46–47.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

separating them. If ever this acidifiable base of the muriatic acid is discovered to be a known substance, though now unknown in that capacity, it will be requisite to change its present denomination for one analogous with that of its base. (Lavoisier (1802), 121–122)<sup>17</sup>

We can locate the birth of the muriatic radical in these two passages, in which we can see the hypothesis that muriatic acid is made up of oxygen, the acidifying principle, combined with some unknown acidifiable base or radical. This is the birth of a new hypothetical entity, in the sense of the rough guide put forward in chapter 1. The muriatic radical is a new kind of purported entity that a scientist put forward as a kind of purported empirical entity in advance of decisive empirical reasons to do so. Theoretical considerations lead Lavoisier and Guyton de Morveau to put forward this entity as an empirical entity, but those considerations were less than decisive. Moreover, we have a number of different HE terms, including: ‘muriatic radical,’ ‘muriatic base,’ ‘acidifiable base of muriatic acid,’ and ‘muriatic radical principle.’

We can now understand Lavoisier’s more specific hypothesis concerning the nature of dephlogisticated muriatic acid, which he calls ‘oxygenated muriatic acid,’ and others call ‘oxymuriatic acid.’<sup>18</sup> Lavoisier held that two different acids can have the same constituent elements, and that what makes them different is the different proportions of those elements that the acids contain. For example, he believed that both sulphurous acid and sulphuric acid contain nothing but sulphur and oxygen. What makes them different acids

---

<sup>17</sup>Translation of Lavoisier (1801), 75–76.

<sup>18</sup>For Lavoisier’s terminology, see Lavoisier (1802), 123–124 (Lavoisier (1801), 77 for the original French). For the alternative terminology, see, for example, Davy (1810b).

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

is that the former is “under-saturated with oxygen,” while the latter is “completely saturated” (Lavoisier (1802), 117–118).<sup>19</sup> In the same way, Lavoisier held that oxymuriatic acid results from a combination of muriatic acid and oxygen, in which case oxymuriatic acid contains the muriatic radical as well (Lavoisier (1802), 123–124).<sup>20</sup> To put the point another way, muriatic acid is “under-saturated with oxygen,” while oxymuriatic acid is “completely saturated,” though, to be sure, both contain the muriatic radical.

In sum, the oxygen theory, as applied to muriatic and oxymuriatic acid, yields the following:

muriatic radical + oxygen = muriatic acid

muriatic acid + oxygen = oxymuriatic acid

It is also important to note that the muriatic and oxymuriatic acids were not the only acids at this time that chemists were unable to decompose. Boracic acid and fluoric acid, like muriatic acid, had not yet been decomposed, and indeed, Lavoisier lists the boracic and fluoric radicals, along with the muriatic radical, in his table of simple substances (Lavoisier (1802), 291).<sup>21</sup>

### 3.2.3 THE PHLOGISTON THEORY OF ACIDITY

In 1789, the year in which Lavoisier’s *Traité* first appeared, the second edition of Richard Kirwan’s *An Essay on Phlogiston and the Constitution of Acids*, complete with responses from Lavoisier and other French chemists, was published. Kirwan was a fer-

---

<sup>19</sup>See Lavoisier (1801), 71–72 for the original French.

<sup>20</sup>See Lavoisier (1801), 77 for the original French.

<sup>21</sup>See Lavoisier (1801), 192 for the original French.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

vent defender of the phlogiston theory, and phlogiston figures prominently in his theory of acidity.

Kirwan, like other latter-day phlogiston theorists, identified phlogiston with what he and other chemists called “inflammable air.”<sup>22</sup> According to Kirwan,

inflammable air, before its extrication from bodies in which it exists in a concrete state, was the very substance to which all the characters and properties of the phlogiston of the ancient chymists actually belonged . . . (Kirwan (1789), 4–5)

Lavoisier and his colleagues used the term ‘hydrogene’ to refer to the substance that Kirwan referred to as ‘inflammable air,’<sup>23</sup> though the French chemists obviously didn’t identify this, or any substance of which they were familiar, with “the phlogiston of the ancient chymists.”

Kirwan’s theory of acidity can be seen as a modification of Lavoisier’s. On Lavoisier’s theory, acids are made up of an acidifiable base or radical plus oxygen. Kirwan agrees with this much, but argues that the radical always contains phlogiston. An acid forms when the phlogiston leaves the radical to combine with “pure air,” i.e., oxygen, to form “fixed air” (Kirwan (1789), 39). Acids in general, then, are composed of an (at least partially) dephlogisticated radical, fixed air (which Kirwan thought to be a compound of phlogiston and oxygen), and perhaps some additional phlogiston that did not unite with oxygen.<sup>24</sup>

---

<sup>22</sup>Joseph Priestley identified phlogiston and inflammable air as early as 1782 in his correspondences (Priestley (1892), 33). See Kitcher (1978), 534 for more citations, along with some discussion. Henry Cavendish admits the possibility that Priestley and Kirwan are correct, but holds that there is more evidence in favor of the view that inflammable air is phlogiston united with water (Cavendish (1784), 137).

<sup>23</sup>See, for example, Kirwan (1789), 21–22, an excerpt from Lavoisier’s note on Kirwan’s introduction.

<sup>24</sup>See Kirwan (1789), 62 for an example, namely, Kirwan’s treatment of sulphurous acid and sulphuric acid, which he calls volatile vitriolic acid and fixed vitriolic acid, respectively.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

Moreover, for Kirwan, fixed air, and not oxygen, is the acidifying principle (Kirwan (1789), 78, 80).

Kirwan's claim that fixed air contains only oxygen and phlogiston is bound to be confusing if one knows that the substance that Kirwan called 'fixed air' is what other chemists at the time called 'carbonic acid,' and what we today call 'carbon dioxide.' But for Kirwan, fixed air contains no carbon—only phlogiston (i.e., inflammable air, or hydrogen) and pure air (i.e., oxygen). Lavoisier laments this consequence of Kirwan's view in the following passage:

Mr. Kirwan agrees with us in this article, that water is a compound of dephlogisticated air and inflammable gas, or, according to our manner of expression, of oxigene and hydrogene; we have therefore no reply to make on this head. But he thinks that, from this combination of oxigene and hydrogene, there does not always result water; that in some circumstances the result is fixed air, or carbonic acid; that in others it is phlogisticated air, or azotic gas; and, lastly, that the result is sometimes nitrous gas and nitrous acid. All these substances, according to Mr. Kirwan, are composed of the same principles ... (Kirwan (1789), 56)

Such a view can be maintained only if one holds that, in these different cases, hydrogen and oxygen combine in different ways. For example, one might hold that they combine in different proportions. The important point for now is just that, for Kirwan, fixed air consists of the combination of hydrogen and oxygen.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

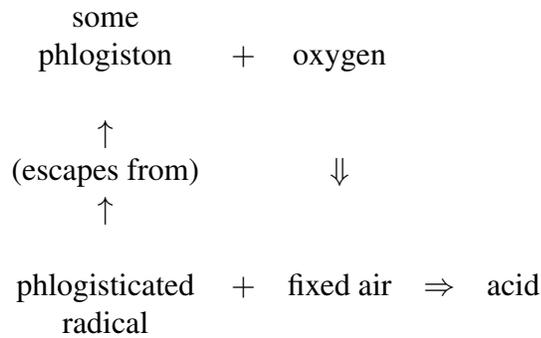


Figure 3.1: Phlogisticated acids.

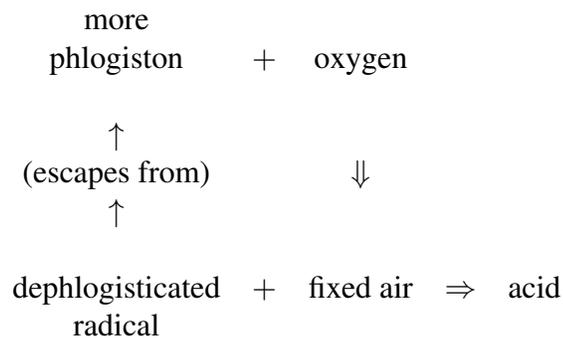


Figure 3.2: Dephlogisticated acids.

Coming back to Kirwan’s theory of acidity, there are really two cases to consider—that of phlogisticated acids, and that of dephlogisticated acids, as figures 3.1 and 3.2 make clear. In both cases, phlogiston leaves the radical to unite with hydrogen in order to form fixed air. The latter combines with the radical to form the acid. Both figures are best read by starting at the bottom left corner, and proceeding clockwise by following the ‘escapes from’ arrows. The plus sign indicates that two or more substances are combining, and the double arrows indicates the result of their combination.

After sketching his theory, Kirwan goes on to deal with some concrete examples, including the as-yet-undecomposed muriatic and oxygenated muriatic acids, which he calls marine acid and dephlogisticated marine acid, respectively, and which we would call hy-

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

drochloric acid and chlorine, respectively. He believed marine acid “to consist of a peculiar basis united to phlogiston, and a certain proportion of fixed air” (Kirwan (1789), 126). Dephlogisticated marine acid, then, results “from the union of the dephlogisticated basis, with an excess of fixed air” (Kirwan (1789), 126). The former, then, is a phlogisticated acid, while the latter is a dephlogisticated acid. More schematically, we get the following view of the acids:

phlogisticated radical + fixed air = muriatic acid

dephlogisticated radical + fixed air = dephlogisticated muriatic acid

Importantly, then, like Scheele, Kirwan holds that dephlogisticated muriatic acid is simpler than muriatic acid. Moreover, one can obtain the latter by adding phlogiston (i.e., hydrogen) to the former.

For Kirwan, the heart of the controversy between the proponents of the oxygen theory and of the phlogiston theory is the following:

The controversy is therefore at present confined to a few points, namely, whether the *inflammable principle* be found in what are called phlogisticated acids, vegetable acids, fixed air, sulphur, phosphorus, sugar, charcoal, and metals.  
(Kirwan (1789), 6–7)

This controversy was decided, in Kirwan’s mind, by 1791, two years after the publication of the second edition of his *Essay*, when he gave up the phlogiston theory:

I know of no single clear decisive experiment by which one can establish that

fixed air is composed of oxygen and phlogiston, and without this proof it seems to me impossible to prove the presence of phlogiston in metals, sulphur or nitrogen (Saltpeterluft). (quoted in Partington (1961), 664)

But although it lost one of its most fervent defenders, the phlogiston theory did not disappear from the scene.

### 3.3 FROM MURIATIC ACID TO HYDROCHLORIC ACID

I now turn to the task of exploring the subsequent history of the muriatic radical, primarily by means of an examination of Humphry Davy's work leading up to his determination of the composition of muriatic acid. As the quotation from Remsen makes clear, we now credit Davy with establishing the elementary nature of chlorine. Moreover, we credit him with the discovery that muriatic acid is composed of hydrogen and chlorine, which makes it clear why we now refer to that acid as hydrochloric acid (HCl). Because this acid contains no oxygen, Davy's work on muriatic acid is seen as delivering the killing blow to the oxygen theory of acidity, the theory which gave birth to the muriatic radical.<sup>25</sup>

To be sure, there are good reasons to credit Davy with these accomplishments. Indeed, Davy begins his 1810 Bakerian Lecture by claiming that oxymuriatic acid gas is "elementary as far as our knowledge extends" (Davy (1811a), 1). He recognized that, "[t]o call a body which is not known to contain oxygene, and which cannot contain muriatic acid, oxymuriatic acid, is contrary to the principles of that nomenclature in which it is adopted," and therefore proposed that "chlorine" and "chloric gas" be used to name that body (Davy

---

<sup>25</sup>For claims similar to these, see, for example, Le Grand (1974), 224 and Chang (2012b), 33.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

(1811a), 32). However, Davy's concerns with nomenclature didn't extend to muriatic acid. In the same lecture, he states that "[t]he name muriatic acid, as applied to the compound of hydrogen and oxymuriatic gas, there seems to be no reason for altering" (Davy (1811a), 33). Moreover, earlier in 1810, he claimed that muriatic acid "may be considered as having hydrogen for its basis, and oxymuriatic acid for its acidifying principle" (Davy (1810b), 243). The addition of another acidifying principle is a strong point against the oxygen theory.

In this episode from the history of science, we see a transition from discourse involving hypothetical entities to discourse involving empirical entities. The state of chemistry when Davy began his work involved various hypotheses about the composition of muriatic acid and oxymuriatic acid. And these hypotheses involved a hypothetical entity, namely, the muriatic radical. Davy then went on to determine that muriatic acid is composed of hydrogen and oxymuriatic acid, two empirical entities.

But as is often the case in the history of science, the full story is much more complicated, and much more interesting. In particular, three issues stand out as puzzling. First of all, there is the question of why Davy succeeded, when so many other chemists failed to determine the natures of muriatic acid and oxymuriatic acid. What was special about Davy? After all, when it came to another as-yet-undecomposed acid, namely, boracic acid, Davy had to share his success in decomposing this acid with Joseph Louis Gay-Lussac and Louis Jacques Thénard. What makes this issue even more puzzling is that, as Jan Golinski points out, in making his argument concerning the natures of muriatic acid and oxymuriatic acid, "Davy brought forward no dramatically new experimental evidence" (Golinski

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

(1992), 223). If this is the case, then why was Davy capable of succeeding where other chemists had failed?

Secondly, Davy's views about the elementary nature of chlorine are a bit confusing. After introducing the name "chlorine," Davy has this to say:

Should it [chlorine] hereafter be discovered to be compound, and even to contain oxygene, this name can imply no error, and cannot necessarily require a change. (Davy (1811a), 32)

It is possible that oxymuriatic gas may be compound, and that this body and oxygene may contain some common principle; but at present we have no more right to say that oxymuriatic gas contains oxygene than to say that tin contains hydrogene . . . (Davy (1811a), 34)

There *may* be oxygene in oxymuriatic gas; but I can find none. (Davy (1811a), 35)

Such statements sound somewhat odd coming from the chemist credited with establishing the elementary nature of chlorine.

Thirdly, Davy's remarks about the composition of the muriatic and oxymuriatic acids display his complicated relationship to what he calls the "phlogistic" and "antiphlogistic" theories.<sup>26</sup> Earlier in 1810, in the same paper in which he successfully reports the composition of muriatic acid, Davy has the following to say about the ideas of chlorine's discoverer regarding the nature of muriatic acid and oxymuriatic acid:

---

<sup>26</sup>See, for example, Davy (1808a), 32–33 for this terminology.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

It is evident from this series of observations, that SCHEELE'S view, (though obscured by terms derived from a vague and unfounded general theory,) of the nature of the oxymuriatic and muriatic acids, may be considered as an expression of facts; whilst the view adopted by the French school of chemistry, and which, till it is minutely examined, appears so beautiful and satisfactory, rests in the present state of our knowledge, upon hypothetical grounds. (Davy (1810b), 237)

If one rejects the antiphlogistic oxygen theory of acidity, it surely doesn't follow that one must accept some sort of phlogistic theory. How, then, can Davy, working in the early years of the nineteenth century, claim that Scheele, a proponent of the presumably discredited phlogiston theory, is essentially correct?

As it turns out, Davy's relationship to the phlogistic and antiphlogistic theories is a complicated one. In the spring of 1799, Davy wrote a "Prospectus of Experiments," which includes his hope "[t]o decompose the muriatic, boracic, and fluoric acids; to try triple affinities, and the contact with heated combustible bodies at a high temperature" (quoted in Davy (1839), 56). Davy goes on to focus on muriatic acid in particular, and asks:

May we not be able to decompose muriatic acid by heating some of the muriates of the metals red, and sending sulphur in vapour through them? The muriate of lead might be tried in this way, or the muriate of copper. The attraction of copper for sulphur, and the attraction of sulphur for oxygen, would most probably effect the decomposition. (quoted in Davy (1839), 57)

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

Davy's brother John claims, in his commentary on this passage, that the fact that Davy speculates about methods of drawing the oxygen out of muriatic acid shows that he views acids as compounds of oxygen and a base or radical, in conformity with the oxygen theory (Davy (1839), 57). However, although Davy's views were in conformity with the oxygen theory, this does not entail his acceptance of that theory, or his outright rejection of the phlogiston theory. Two years later, in a letter to James Watt, Davy raises a rhetorical question concerning hydrogen: "May not this substance come from the metal & be the old phlogiston?" (Davy (1801)). Perhaps, then, John's claim about his brother's allegiance to the oxygen theory at this time is not quite right.

In the papers leading up to his determination of the composition of the muriatic and oxymuriatic acids, Davy occasionally points to the limitations of the oxygen theory, and devotes a fair amount of discussion to the prospects for a modified phlogiston theory. Davy's concern with the latter may seem puzzling to us today. As David Knight puts it, "[b]ecause we have been brought up with the idea that phlogiston was proved to be illusory by Lavoisier, it is a surprise to find the old theory appearing about twenty years later in such a context" (Knight (1992), 68). Moreover, this issue has been ignored by many of those who have written about Davy. As Robert Siegfried has pointed out, "Davy's biographers have generally ignored his extensive flirtation with phlogiston, perhaps from a false sense of embarrassment for a man who should have known better" (Siegfried (1964), 118). Knight hints that Davy's concern with the phlogiston theory stems from a desire to triumph over the French (Knight (1992), 68–69). And Siegfried argues that Davy's desire to reduce the number of chemical elements drove him to consider defensible modifications

of the phlogiston theory (Siegfried (1964), 119). Siegfried and Knight thus provide us with some insight as to why Davy put so much effort into developing his modified phlogiston theory.

There still might seem to be a puzzle here regarding Davy's attitude towards the phlogiston and oxygen theories, at least if one thinks that he had to accept at least one of these theories. My goal at this point is to examine Davy's work leading up to his determination of the composition of the muriatic and oxymuriatic acids, while also tracing his complicated relationship to the phlogiston and oxygen theories. This will make it clear that Davy entertained both theories while accepting neither. Later, I will argue that a proper understanding of this complicated relationship will help us to solve the three puzzles discussed above. One upshot of the account that I'll defend here is that the fact that Davy entertained both the phlogiston theory and the oxygen theory allowed for scientific progress. To put my point another way, this episode gives us an example of a case from the history of science in which two wrongs (the phlogistic and antiphlogistic theories) make a right (a proper understanding of the muriatic and oxymuriatic acids).<sup>27</sup>

### 3.3.1 DAVY ON OXYGEN, PHLOGISTON, AND THE ACIDS

#### **The 1807 Bakerian Lecture**

In his 1807 Bakerian Lecture, delivered in November of that year, Davy displays some ambivalence towards both the phlogiston theory and the oxygen theory. To be sure, at this point, he still thinks that muriatic acid contains oxygen, and he believes that this view is

---

<sup>27</sup>I owe this way of putting the point to Steven Gross.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

supported by his experiments on the decomposition of various substances by means of an electric circuit, which show that:

as far as our knowledge of the composition of bodies extends, all substances attracted by positive electricity, are oxygene, or such as contain oxygene in excess; and all that are attracted by negative electricity, are pure combustibles, or such as consist chiefly of combustible matter. (Davy (1808a), 43)

Davy goes on to claim that, since acids are attracted by positive electricity, “[t]he idea of muriatic acid, fluoric acid, and boracic acid containing oxygene, is highly strengthened by these facts” (Davy (1808a), 43). Davy’s views therefore seem to be in conformity with the oxygen theory.

However, in the same lecture, Davy expresses a tentative attitude toward the oxygen theory. He claims that, although “the antiphlogistic solution of the phenomena has been uniformly adopted, yet the motive for employing it has been rather a sense of its beauty and precision, than a conviction of its permanency and truth”; though, to be sure, he also admits that, “in the present state of our knowledge, it appears the best approximation that has been made to a perfect logic of chemistry” (Davy (1808a), 32–33). So, while admitting that the oxygen theory may turn out to be false, passages like these suggest that Davy prefers the oxygen theory to the phlogiston theory.

But in a footnote that occurs shortly after this passage, Davy goes on to sketch a way to defend a phlogiston theory. On such a theory,

the metals are compounds of certain unknown bases with the same matter as

that existing in hydrogen; and the metallic oxides, alkalies and acids compounds of the same bases with water . . . (Davy (1808a), 33)

Once one recalls the phlogiston theorists' identification of phlogiston with hydrogen, it's clear that this is a version of the phlogiston theory, and the similarities with Kirwan's theory are striking. For Kirwan, an acid is composed of a base plus fixed air, the latter being his principle of acidity, which, in turn, is composed of phlogiston (i.e., hydrogen) and oxygen. As Davy uses the term 'carbonic acid' instead of 'fixed air,' he follows Lavoisier, as opposed to Kirwan, and holds that this substance is a compound of carbon and oxygen.<sup>28</sup> In that case, Davy doesn't commit himself to the view that carbonic acid is a constituent of all acids, and to this extent, his phlogiston theory is in conflict with Kirwan's. But Kirwan also emphasizes that the real controversy concerns whether phlogiston is a constituent of acids (among other substances). And, on Davy's phlogiston theory, acids do contain phlogiston. Davy, of course, knew that water is a compound of hydrogen and oxygen.<sup>29</sup> In that case, if an acid is composed of a base plus water, then Davy's phlogiston theory squares with Kirwan's, since, on both theories, an acid is composed of a base plus phlogiston (i.e., hydrogen) and oxygen.

Although Davy puts this theory forward as a defensible phlogiston theory, he immediately notes three reservations about such a theory. First of all, he claims that "in this theory more unknown principles would be assumed than in the generally received theory," i.e., the oxygen theory (Davy (1808a), 33). Secondly, "[i]t would be less elegant and less dis-

---

<sup>28</sup>See, Lavoisier (1802), 116 for an example of Lavoisier's usage of the term (Lavoisier (1801), 70 for the original French), and Davy (1808a), 44 for an example of Davy's usage.

<sup>29</sup>See, for example, Davy (1807), 12.

tinct” (Davy (1808a), 33). And thirdly, Davy notes that, although he initially thought that he had some experimental evidence for the existence of hydrogen in potassium, his newly discovered metal, a closer examination revealed that the hydrogen was evolved from some water present in the experimental apparatus. All things considered, then, the 1807 Bakerian Lecture shows Davy’s ambivalence towards both the oxygen theory and the phlogiston theory.

#### **“Electro-Chemical Researches ...” and the 1808 Bakerian Lecture**

In June of 1808, Davy read a paper to the Royal Society in which he communicated a number of experimental results. In a footnote, he discusses his attempts to decompose the as-yet-undecomposed acids, namely, the fluoric, boracic, and muriatic acids, by means of potassium. Knowing that potassium has a strong affinity for oxygen, he attempted to use that metal in order to decompose these acids. Regarding muriatic acid, his results were as follows:

When potassium was heated in muriatic acid gas, as dry as it could be obtained by common chemical means, there was a violent chemical action with ignition; and when the potassium was in sufficient quantity, the muriatic acid gas wholly disappeared, and from one-third to one-fourth of its volume of hydrogen was evolved, and muriate of potash was formed. (Davy (1808b), 343)

A few paragraphs later, Davy claims that

[i]n all these instances there is great reason to believe that hydrogen was pro-

duced from the water adhering to the acids; and the different proportions of it in the different cases, are a strong proof of this opinion. Admitting this idea, it seems that muriatic acid gas must contain at least one-eighth or one-tenth of its weight of water . . . (Davy (1808b), 344)

Davy then proposes that “[t]hose persons who have supposed hydrogen to be the basis of muriatic acid may, perhaps, give another solution of the phenomena, and consider the experiment I have detailed as a proof of this opinion” (Davy (1808b), 344). As the quotation makes clear, this idea is not one that originates in Davy. To take one example, Lavoisier’s English translator, Robert Kerr, notes Christoph Girtanner’s claim that hydrogen is the muriatic radical, for which Kerr suggests the name ‘muriogen’ (Lavoisier (1802), 122).

What Davy has to say in this footnote is in conformity with both the oxygen theory and the phlogiston theory from the 1807 Bakerian Lecture. His attempt to draw the oxygen out of muriatic acid by means of potassium is in conformity with the oxygen theory of acidity. But it is also in conformity with Davy’s phlogiston theory, according to which an acid is a base plus water. And the claim that hydrogen is the basis of muriatic acid, is likewise consistent with both theories. On the oxygen theory, it makes sense to label hydrogen as the basis, as opposed to the acidifying principle, since, unlike oxygen, and like the bases of other acids, it is attracted by negative electricity (Davy (1807), 28–29). But it is also consistent with Davy’s phlogiston theory. To label something as a base, according to that theory, does not commit one to the claim that oxygen is the principle of acidity. Just as Lavoisier can speak of bases as constituents of substances other than acids, Davy can as well.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

Later in the paper, Davy goes on to discuss the prospects for the phlogiston theory from the 1807 Bakerian Lecture, in light of some recent experimental results (Davy (1808b), 362–363). These experimental results have to do with the nature of ammonia, also known as the volatile alkali at that time, and they lead Davy to believe that ammonia is a compound of oxygen and a “binary basis” composed of hydrogen and nitrogen (Davy (1808b), 353–354). Davy calls this basis “ammonium” and believes that it is metallic in nature (Davy (1808b), 361–362).

While admitting that the phlogiston theory is still inferior to the oxygen theory, Davy claims that “the only good arguments in favour of a common principle of inflammability, flow from some of the novel analogies in electrochemical science” (Davy (1808b), 363). He goes on to explain what he has in mind:

Assuming the existence of hydrogene in the amalgam of ammonium, its presence in one metallic compound evidently leads to the suspicion of its combination in others. And in the electrical powers of the different species of matter, there are circumstances which extend the idea to combustible substances in general. Oxygene is the only body which can be supposed to be elementary, attracted by the positive surface in the electrical circuit, and all compound bodies, the nature of which is known, that are attracted by this surface, contain a considerable proportion of oxygene. Hydrogene is the only matter attracted by the negative surface, which can be considered as acting the opposite part to oxygene; may not then the different inflammable bodies, supposed to be simple, contain this as a common element? (Davy (1808b), 363)

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

The basic idea is that any matter attracted to the negative surface must contain hydrogen, and that ammonium, the base of ammonia, gives us one example of this. Davy explicitly labels this as a “hypothesis” (Davy (1808b), 363–364), and continues to spell out the idea of oxygen and hydrogen acting in opposite ways, in terms of the example of ammonia:

by the absorption of oxygene, does the amalgam of ammonium produce the volatile alkali; and if we suppose that ammonia is metallized, by being combined with hydrogene, and freed from water, the same reasoning will likewise apply to the other metals, with this difference, that the adherence of their phlogiston or hydrogene, would be exactly in the inverse ratio of their attraction for oxygene. (Davy (1808b), 364)

Davy develops his hypothesis by generalizing from his beliefs about the nature of ammonia, and his explicit identification of phlogiston and hydrogen is noteworthy.

Davy ends his 1808 paper on a speculative note. He considers a hypothesis concerning the connection between electrical states and chemical states, according to which “still fewer elements than those allowed in the antiphlogistic or phlogistic theory might be maintained” (Davy (1808b), 368). Davy’s idea here is that there is a correlation between electrical states and chemical states. Substances attracted to the positive surface of an electric circuit, like the acids, are negative, while substances attracted to the negative surface, like the alkalies, are positive. Moreover, when the acids are positively electrified, they lose their acidic properties, and when the alkalies are negatively electrified, they lose their alkaline properties. In a footnote, Davy goes on to claim that, on this hypothesis, “[w]ater positively electrified

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

would be hydrogene, water negatively electrified, oxygene” (Davy (1808b), 369). He goes on to note some ways in which such a theory would perform better than the oxygen and phlogiston theories, especially when it comes to the nature of ammonia.

To be sure, though, Davy’s all-things-considered view at this time is that “the age of chemistry is not yet sufficiently mature for such discussions; the more subtile powers of matter are but just beginning to be considered; and all general views concerning them, must as yet rest upon feeble and imperfect foundations” (Davy (1808b), 369). Hence, he remains noncommittal regarding the oxygen theory, the phlogiston theory, and this new electro-chemical theory.

In his 1808 Bakerian Lecture, delivered in December of that year, Davy devotes a fair amount of discussion to the as-yet-undecomposed acids. By this point, he had successfully decomposed boracic acid, proposing the name “boracium” for what he calls the “boracic basis” (Davy (1809a), 85). This may have been news in England, but in France, Gay-Lussac and Thénard had reported the decomposition of boracic acid little more than a month before, and proposed the name “bore” (i.e., “boron”) for the “radical boracique” (Gay-Lussac and Thénard (1808), 173). With regard to muriatic acid, Davy was still unsuccessful (Davy (1809a), 91–100).

Davy concludes the 1808 Bakerian Lecture by, once again, considering the prospects for the phlogiston theory advanced in the 1807 lecture:

The facts advanced in this Lecture, afford no new arguments in favour of an idea to which I referred in my last communication to the Society, that of hydrogene being a common principle in all inflammable bodies; and except in

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

instances which are still under investigation, and concerning which no precise conclusions can as yet be drawn, the generalization of LAVOSIER happily applies to the explanation of all the new phenomena. (Davy (1809a), 103)

This passage, once again, shows that Davy did not accept the phlogiston theory. But one gets the impression that he would have liked to have uncovered some evidence in favor of that theory.

In the 1809 appendix to the lecture, which appeared in print in March of that year, Davy is clear about the moral that we should draw from the 1808 Bakerian Lecture. He claims that the experiments reported in that lecture

shew that the ideas which had been formerly entertained respecting the difference between the muriatic acid and the oxymuriatic acid are not correct. They prove that muriatic acid gas is a compound of a substance, which as yet has never been procured in an uncombined state, and from one-third to one-fourth of water, and that oxymuriatic acid is composed of the same substance, (free from water) united to oxygene. They likewise prove, that when bodies are oxydated in muriatic acid gas, it is by a decomposition of the water contained in that substance, and when they are oxydated in oxymuriatic acid, it is by combination with the oxygene in that body, and in both cases there is always a union of the peculiar unknown substance, the dry muriatic acid with the oxydated body. (Davy (1809c), 468)

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

A bit more schematically, Davy claims the following:

dry muriatic acid + water = muriatic acid

dry muriatic acid + oxygen = oxymuriatic acid

oxymuriatic acid + hydrogen = muriatic acid

Interestingly, this is the first time that Davy puts forward a view of the acids according to which oxymuriatic acid is a simpler substance than muriatic acid, and is actually contained in the latter. Davy's view resembles the phlogiston theories of Scheele and Kirwan in this respect, and therefore differs from Lavoisier's oxygen theory, according to which the exact opposite is true. Moreover, if we follow Kirwan and identify hydrogen and phlogiston, then oxymuriatic acid would, indeed, be a dephlogisticated acid. However, it would be a bit hasty to conclude from this that Davy rejects the oxygen theory and accepts the phlogiston theory at this point. Since both acids contain the acidifying principle, namely, oxygen, one could still claim that Davy's view squares with the essential details of the oxygen theory. Davy's confidence in this view is revealed by his use of the language of proof in the passage quoted above. But his view did not yet receive the ultimate proof, since Davy goes on to report that all of his attempts to procure this unknown dry muriatic acid have failed.

In order to more clearly situate Davy's view on the muriatic and oxymuriatic acids, we can contrast it with the view that Gay-Lussac and Thénard put forward in February of 1809, roughly the same time that Davy put forward the view under consideration. Gay-Lussac

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

and Thénard claim that water, or perhaps hydrogen and oxygen separately, are contained within muriatic acid (Gay-Lussac and Thénard (1809), 321–322). To this extent, their view shares some similarities with Davy’s. Moreover, both Gay-Lussac and Thénard, and Davy, admit oxygen as a constituent of oxymuriatic acid (Gay-Lussac and Thénard (1809), 341). But unlike Davy, Gay-Lussac and Thénard still view muriatic acid as a constituent of oxymuriatic acid. Indeed, they still set themselves the goal of “extract[ing] [muriatic acid] from oxygenated muriatic gas, by removing the oxygen by means of combustible bodies” (Gay-Lussac and Thénard (1905), 39).<sup>30</sup> To put their view schematically,<sup>31</sup>

muriatic radical + water = muriatic acid

muriatic acid + oxygen = oxymuriatic acid

Interestingly, Gay-Lussac and Thénard consider the hypothesis that oxymuriatic acid is a simple substance. But they ultimately reject this hypothesis in favor of the view that it is a compound body (Gay-Lussac and Thénard (1809), 357–358).<sup>32</sup> So, Gay-Lussac and Thénard’s view accommodates the claim of the oxygen theory that muriatic acid is a constituent of oxymuriatic acid, and can be seen as an extension of Lavoisier’s oxygen theory. Davy’s view, on the other hand, contradicts this claim.

---

<sup>30</sup>Translation of Gay-Lussac and Thénard (1809), 342.

<sup>31</sup>It should be noted that Gay-Lussac and Thénard do not mention the muriatic radical in their paper. But they do discuss the boracic radical (Gay-Lussac and Thénard (1809), 297) and the fluoric radical (Gay-Lussac and Thénard (1809), 298).

<sup>32</sup>For Davy’s recognition of this point, see Davy (1810b), 237. Crosland and Knight claim that it was Berthollet who convinced Gay-Lussac and Thénard not to abandon the oxygen theory. See Crosland (1980), 102 and Knight (1978), 135.

### The 1809 Bakerian Lecture

In November of 1809, Davy delivered another Bakerian Lecture, which touches on the composition of the muriatic and oxymuriatic acids, along with the prospects of the oxygen theory, the phlogiston theory, and the electro-chemical theory. Regarding the former, his views remain essentially unchanged from the previous year's Bakerian Lecture, as he still thinks that muriatic acid "is a compound of a body unknown in a separate state, and water" (Davy (1810a), 67). He does, however, speculate that the muriatic radical is metallic in nature (Davy (1810a), 68–69). Moreover, Davy puts even more distance between his view and the oxygen theory, when he claims that "in oxymuriatic acid, the acid energy seems to be blunted by oxygene, and is restored by the addition of hydrogene," which, on Davy's current way of seeing things, would yield muriatic acid (Davy (1810a), 70). Far from being the principle of acidity, oxygen actually plays a role in suppressing the acidic properties of a substance.

Since Davy distances himself from the oxygen theory, one might think that he comes closer to accepting the phlogiston theory. He does discuss this theory in the lecture, as when he writes:

If hydrogene, according to an hypothesis to which I have often referred, be considered as the principle which gives inflammability, and as the cause of metallization, then our list of simple substances will include oxygene, hydrogene, and unknown bases only; metals and inflammable solids will be compounds of these bases, with hydrogene; the earths, the fixed alkalies, metallic oxides, and

the common acids, will be compounds of the same bases, with water. (Davy (1810a), 69)

The hypothesis to which he has often referred is the phlogistic hypothesis from the 1807 Bakerian Lecture. But Davy makes a point that he doesn't make in 1807, namely, that the simple substances on this hypothesis are oxygen, hydrogen, and unknown bases.

The fact that he labels this as a hypothesis, though, indicates that Davy isn't quick to embrace the phlogiston theory either. At one point, he stresses that his "conclusions must be considered as entirely independent of hypothetical opinions, concerning the existence of hydrogen in combustible bodies, as a common principle of inflammability" (Davy (1810a), 37). Davy goes on to devote a fair amount of discussion to contrasting the phlogistic and antiphlogistic explanations of various phenomena associated with the nature of ammonia (Davy (1810a), 56–58). He explicitly labels the latter as a "hypothesis," but points to some phenomena that seem to him to tell in favor of that hypothesis (Davy (1810a), 57). However, his all-things-considered view appears at the end of the lecture, where he concludes that neither theory deals adequately with the phenomena (Davy (1810a), 73).<sup>33</sup>

The electro-chemical theory of Davy's 1808 paper also makes an appearance in the 1809 Bakerian Lecture. Davy brings up "an hypothesis which [he has] before brought

---

<sup>33</sup>Interestingly, later in November of 1809, the same month in which he delivered the 1809 Bakerian Lecture, Davy displays some enthusiasm for the phlogiston theory in a letter to John Dalton. After discussing the work of Gay-Lussac and Thénard, along with some work relating to Davy's 1808 Bakerian Lecture, Davy asks rhetorically: "Had these facts appeared 20 years ago, who would have given up phlogiston" (Davy (1809b)). Such enthusiasm might go some length towards explaining why Davy continued to entertain the phlogiston theory. Moreover, this enthusiasm is consistent with the claim that he did not accept the phlogiston theory, and is also consistent with his claim, at the end of the 1809 Bakerian Lecture, that the phlogiston theory does not provide an adequate explanation of the phenomena.

before the Society,” according to which:

the power of chemical attraction and electrical action, may be different exhibitions of the same property of matter, and . . . oxygene and inflammable bodies, are in relations of attraction which correspond to the function of being negative and positive respectively, . . . (Davy (1810a), 62)

At this point, the theory looks like nothing more than the electro-chemical theory from 1808, and in that earlier paper, Davy is clear that the electro-chemical theory is distinct from both the phlogistic and antiphlogistic theories (Davy (1808b), 368).

#### **“Researches on the oxymuriatic Acid . . .” and the 1810 Bakerian Lecture**

In a paper read in July of 1810, Davy reports his view that “oxymuriatic acid gas is converted into muriatic acid gas, by combining with hydrogene” (Davy (1810b), 236). While this is a consequence of the view that he put forward in the appendix to the 1808 Bakerian Lecture, Davy also makes a novel claim, namely, that the assumption that oxymuriatic acid contains oxygen “has not yet been proved” (Davy (1810b), 236). This view is quite different from Davy’s earlier view. Davy no longer entertains dry muriatic acid as a hypothetical substance. And while he formerly claimed in the appendix that he had proof of the view that oxymuriatic acid contains dry muriatic acid and oxygen, he now claims that oxymuriatic acid has not been proven to contain oxygen. On his new view, then, neither oxymuriatic acid nor muriatic acid has been proven to contain oxygen, which, of course, spells trouble for the oxygen theory. And, indeed, as Davy himself remarks, his own view is more akin

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

to the phlogiston theory of Scheele, according to which muriatic acid is oxymuriatic acid plus phlogiston (i.e., hydrogen) (Davy (1810b), 231, 237).

Davy's argumentative strategy in his 1810 paper is to show that results which seem puzzling on the oxygen theory, can be made sense of by adopting his own view. Davy discusses a number of such results, but two are particularly important for the conclusion that he draws.

First of all, Davy notes that water is always present when muriatic acid and oxygen are procured from oxymuriatic acid. Davy uses the work of Gay-Lussac and Thénard to provide evidence for this claim. They had shown that in all such cases, oxygen cannot be procured from oxymuriatic acid without the presence of water, and that in such cases in which water is present, muriatic acid is also formed (Davy (1810b), 236). On a similar note, Davy claims that it is a consequence of the account given in his 1808 Bakerian Lecture "that muriatic acid can in no instance be procured from oxymuriatic acid, or from dry muriates, unless water or its elements be present" (Davy (1810b), 231–232). The oxygen theory, then, must be committed to the claim that oxymuriatic acid cannot be decomposed into its constituents (oxygen and muriatic acid) without the presence of water.

Secondly, Davy provides evidence for his claim that oxymuriatic acid and hydrogen combine to form muriatic acid. Once again, he uses the work of Gay-Lussac and Thénard in order to support this claim, which Davy had also verified for himself (Davy (1810b), 235–236). The oxygen theory would then have to contend with the fact that, by combining with hydrogen, oxymuriatic acid can be converted into one of its purported constituents, namely muriatic acid.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

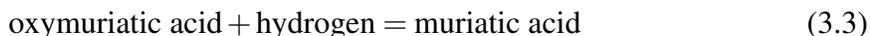
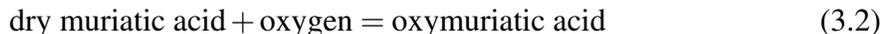
---

Davy draws the following conclusion from these two results:

it is scarcely possible to avoid the conclusion, that the oxygene is derived from the decomposition of water, and, consequently, that the idea of the existence of water in muriatic acid gas, is hypothetical, depending upon an assumption which has not yet been proved—the existence of oxygene in oxymuriatic acid gas. (Davy (1810b), 236)

The oxygen theorists would claim that water is necessary for the decomposition of oxymuriatic acid into muriatic acid and oxygen. Davy, however, claims that what really happens here is that the oxymuriatic acid decomposes the water, combining with the hydrogen to form muriatic acid and leaving oxygen as a byproduct. Hence, we have good reason to think that the oxygen comes, not from the oxymuriatic acid, but from the water.

In the passage quoted in the previous paragraph, Davy is arguing against his former view from the appendix of the 1808 Bakerian Lecture. On Davy's previous view,



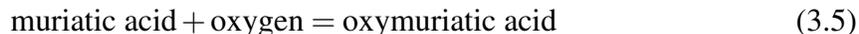
Both Davy's previous view, and his 1810 view, are committed to (3.3), which Davy takes to be established by experiment. Assuming, then, that (3.3) is true, the only way (3.1) could be true is if (3.2) were true. (3.2), however, is what Davy thinks hasn't been established, since

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

we have no reason to think that oxymuriatic acid contains oxygen. In that case, hydrogen cannot combine with that oxygen to form the water that muriatic acid allegedly contains, in which case we have no reason to think (3.1) is true. This is the sense in which the idea of the existence of water in muriatic acid, depends on the assumption that oxymuriatic acid contains oxygen.

In the above-quoted passage, Davy is also concerned to argue against Gay-Lussac and Thénard's application of the oxygen theory to the two acids, which they put forward in February of 1809 (Gay-Lussac and Thénard (1809)). On Gay-Lussac and Thénard's view, muriatic acid is a constituent of oxymuriatic acid:



As Davy puts it, "their general conclusion is, that muriatic acid gas contains about one quarter of its weight of water; and that oxymuriatic acid is not decomposable by any substances but hydrogene, or such as can form triple combinations with it" (Davy (1810b), 232). Their view therefore differs from Lavoisier's insofar as they hold that both acids contain hydrogen, in addition to oxygen. But insofar as both acids contain oxygen, their theory is still recognizably a version of the oxygen theory.

One might think that, on Gay-Lussac and Thénard's view, the idea of the existence of water in muriatic acid does not depend on the assumption that oxymuriatic acid contains oxygen. The dependence for Davy comes from his commitment to (3.3), which Gay-Lussac

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

and Thénard reject. But if (3.3) is, indeed, an experimentally established fact, then Gay-Lussac and Thénard would have to accept it as well. In that case, the only reason to think that muriatic acid contains water is the assumption that oxymuriatic acid contains oxygen. But Davy's reasons for rejecting (3.2) apply equally well to (3.5). And (3.3), combined with a rejection of (3.5), show that we have no reason to accept (3.4) either. Hence, Davy's argument applies to Gay-Lussac and Thénard's view as well.

Davy goes on to conclude that, when it comes to the nature of the muriatic and oxymuriatic acids, the oxygen theory is found to be wanting, while Scheele's original view, framed in terms of the phlogiston theory, gives us a more accurate characterization of the phenomena (Davy (1810b), 237). At this point, Davy's attitude toward the oxygen theory, at least when applied to the two acids, is best characterized as one of outright rejection, whereas before, he entertained the theory, but without accepting it. This is not to say, however, that Davy accepts the phlogiston theory.

To be sure, he does consider a way in which one may argue from an analogy in order to support the oxygen theory; but, without naming it as such, he thinks that this analogy actually supports some version of the phlogiston theory. One might attempt to support the claim "that oxymuriatic acid consists of an acid basis united to oxygene" by appeal to "the general analogy of the compounds of oxymuriatic acid and metals, to the common neutral salts" (Davy (1810b), 239). Davy, however, claims that, in addition to the analogy being "very indistinct,"

it may be applied with as much force to support an opposite doctrine, namely,  
that the neutral salts are compounds of bases with water; and the metals of

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

bases with hydrogen; and that in the case of the action of oxymuriatic acid and metals, the metal furnishes hydrogen to form muriatic acid, and a basis to produce the neutral combination. (Davy (1810b), 239)

Davy doesn't label this as a phlogistic doctrine, but the claim that the metals are combinations of bases plus hydrogen is enough to conclude that it is. But given that Davy does think that the analogy is indistinct, it would be unfair to conclude from this passage that Davy accepts the phlogiston theory.

Moreover, later in the paper, Davy notes that, while he

once thought that the phænomena of metallization might be explained according to a modified phlogistic theory, . . . [t]he phænomena of the action of potassium and sodium upon muriatic acid . . . seem however to overturn these speculations so far as they concern the metals from the fixed alkalies. (Davy (1810b), 246)

Hence, it's clear that Davy's rejection of the oxygen theory does not entail his acceptance of the phlogiston theory.

Davy's rejection of the oxygen theory goes hand-in-hand with his view that the presence of oxygen in oxymuriatic acid has not been proven. Hence, in his paper, he is also concerned to discuss some views regarding the nature of oxymuriatic acid. Davy recognizes, as Scheele, Nicholson, and Berthollet did before him, that oxymuriatic acid either lacks the properties that most acids have, or either has them to a lesser degree. And given that we have no reason to think that it has been decomposed, Davy suggests that "it may

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

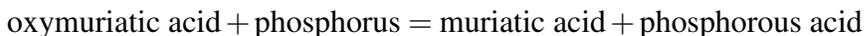
---

possibly belong to the same class of bodies as oxygene,” and may be “a *peculiar* acidifying and dissolving principle” (Davy (1810b), 243). In support of this claim, he notes that both oxygen and oxymuriatic acid are attracted to the positive surface of an electric circuit, and therefore, like oxygen, combines with bodies that are attracted to the negative surface (Davy (1810b), 250). Davy claims that, “[o]n this idea muriatic acid may be considered as having hydrogen for its basis, and oxymuriatic acid for its acidifying principle” (Davy (1810b), 243). It should be emphasized that Davy’s language here indicates that he’s speculating, as opposed to reporting facts, and it wouldn’t be fair to saddle him with the claim that there is, in fact, another principle of acidity, namely, oxymuriatic acid.

Davy continues to discuss oxymuriatic acid in his 1810 Bakerian Lecture, which he read in November of that year. Davy begins the lecture by claiming that, according to the results reported in his previous paper, oxymuriatic acid “has not as yet been decomposed,” and is “elementary as far as our knowledge extends” (Davy (1811a), 1).

Later in the lecture, he gives some more reasons for thinking that oxymuriatic acid does not contain oxygen or muriatic acid. He notes that, in cases in which oxymuriatic acid reacts with various oxides, the amount of oxygen released depends on the amount of the oxide used, and is independent of the amount of oxymuriatic acid used (Davy (1811a), 28). This is puzzling if the oxygen comes from the decomposition of oxymuriatic acid, as the oxygen theorists hold. But it is quite intelligible on Davy’s view, according to which the oxygen is procured from the oxides, and not from oxymuriatic acid. Davy goes on to consider the nature of the combination of oxymuriatic acid and phosphorus. According to Lavoisier’s version of the oxygen theory, the results of such a combination should be the

following:



This is because the oxymuriatic acid would decompose into muriatic acid and oxygen. The latter would then combine with the phosphorus to form phosphorous acid. However, when Davy actually performs this experiment, he finds that the resulting substance lacks the properties that one would associate with the muriatic and phosphorous acids (Davy (1811a), 28). Hence, once again, Davy concludes from all of this that there is no proof of the existence of oxygen or acid matter in oxymuriatic acid.

Given that there is no such proof, Davy presses for a change in terminology:

To call a body which is not known to contain oxygene, and which cannot contain muriatic acid, oxymuriatic acid, is contrary to the principles of that nomenclature in which it is adopted; and an alteration of it seems necessary to assist the progress of discussion, and to diffuse just ideas on the subject. . . . After consulting some of the most eminent chemical philosophers in this country, it has been judged most proper to suggest a name founded upon one of its obvious and characteristic properties—its colour, and to call it *Chlorine*, or *Chloric* gas. (Davy (1811a), 32)

Davy emphasizes that this change in terminology can be happily adopted, even by a proponent of the oxygen theory. His claim has always been that the existence of oxygen in oxymuriatic acid has not been proved. As Davy puts it later in the lecture: “There *may* be

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

oxygene in oxymuriatic gas; but I can find none” (Davy (1811a), 35). Davy’s attitude, then, is perfectly consistent with the possibility that oxygen is a constituent of this acid. Indeed, Davy claims that, “[s]hould [oxymuriatic acid] hereafter be discovered to be compound, and even to contain oxygene, this name can imply no error, and cannot necessarily require a change” (Davy (1811a), 32). So, while he rejects the oxygen theory, he is careful to make sure that his terminology is not rejected by the oxygen theorists.<sup>34</sup> Davy’s revisions in terminology do not, however, extend to “[t]he name muriatic acid, as applied to the compound of hydrogen and oxymuriatic gas,” which he sees “no reason for altering” (Davy (1811a), 33).

Once again, Davy’s rejection of the oxygen theory does not entail acceptance of the phlogiston theory. In putting forward this new terminology, Davy explicitly rejects Scheele’s terminology, since “dephlogisticated marine acid is a term which can hardly be adopted in the present advanced æra of the science” (Davy (1811a), 32). Moreover, Davy explicitly distinguishes his own view from both Scheele’s phlogiston theory and Berthollet’s oxygen theory, when he considers possible explanations for the bleaching powers of oxymuriatic acid. Scheele attempts to explain these powers by appeal to oxymuriatic acid’s ability to combine with phlogiston, while Berthollet appeals to its power to supply oxygen. Davy, however, notes that “the pure gas is incapable of altering vegetable colours,” and explains these bleaching powers by appeal to the fact that oxymuriatic acid can decompose water, combining with its hydrogen, leaving its oxygen in a free state (Davy (1811a), 29).

Although Davy continued to work on both muriatic acid and oxymuriatic acid, we can

---

<sup>34</sup>See Knight (1978), 38 for recognition of this point.

see the 1810 Bakerian Lecture as the culmination of his work on these acids. At this point, he has given his reasons for claiming that, as far as we know, oxymuriatic acid is an elementary body, and it combines with hydrogen to form muriatic acid. Though, as we'll see later, it took some time for the rest of the scientific community to agree with these claims.

### **Davy's Unpublished Notes and Post-1810 Writings**

While Davy's work on the muriatic and oxymuriatic acids culminated in 1810, his speculations concerning modified phlogiston theories and his electro-chemical theory continued for some time. Before discussing those post-1810 speculations, however, I wish to discuss some excerpts from Davy's unpublished notes, which his brother John reproduces in volume I of his *Memoirs of the Life of Sir Humphry Davy*.<sup>35</sup> In these excerpts, we see a blending of theories resembling the phlogistic theory and the electro-chemical theory from his 1808 paper "Electro-Chemical Researches . . ." John doesn't tell us the dates of the unpublished notes from which these excerpts are drawn, though he does tell us that they were composed during Davy's convalescence (Davy (1836), 403). And John tells us elsewhere that Davy "took to his bed about the 23d of November [of 1807], and nine weeks after he was only just convalescent" (Davy (1839), 110). In that case, the earliest that Davy could have composed these notes would have been near the end of January of 1808. And given that these notes involve a blending of the phlogistic and electro-chemical theories from the 1808 paper, read in June of that year, it makes sense that Davy composed them sometime

---

<sup>35</sup>John reproduces the same passages in volume I of Davy's *Collected Works*. See Davy (1839), 155–157.

in 1808.

In a section of his unpublished notes titled “*My Opinions concerning the Elements of Bodies*,” Davy sketches a way in which “[a] new phlogistic theory might be established, which would explain all the phenomena, as well as the antiphlogistic” (Davy (1836), 403–404). The theory is recognizably a phlogiston theory, since it has the consequence that combustible bodies contain hydrogen. But Davy also devotes a fair amount of discussion to the electrical states of bodies, focusing on the idea, already familiar from the above discussions of the electro-chemical theory, that “oxygen is a principle possessing negative electricity, and hydrogen positive; and as all bodies assembling at the positive contain oxygen, so may all bodies assembling at the negative contain hydrogen” (Davy (1836), 404).

After sketching the details of this theory, Davy goes on to a section titled “THIRD THEORY. *Extension and Improvement of the last*” (Davy (1836), 405). The implication here is that these are Davy’s modifications of the phlogiston theory discussed above. But the basic idea behind this theory also bears a striking resemblance to the electro-chemical theory, especially as Davy discusses it in his 1808 paper. On this theory, water is a simple substance. Oxygen is negatively electrified water, while hydrogen is positively electrified water. The metals, charcoal, sulphur, phosphorus, and nitrogen are made up of unknown bases and positively electrified water, i.e., hydrogen. In that case, the theory is still recognizably a phlogiston theory. The acids, oxides, alkalies, and earths, on the other hand, are made up of unknown bases and negatively electrified water. i.e., oxygen (Davy (1836), 405–406). The elements on this theory, then, are water and the unknown bases.

If Davy did, in fact, compose these notes during his convalescence, then it would seem

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

that they predate his 1808 paper. In this case, Davy presumably conceived of the phlogistic and electro-chemical theories in a blended form before distinguishing them in his 1808 paper. But, as John isn't completely clear about the dates of the unpublished notes, we also need to leave open the possibility that they were composed after the 1808 paper. This possibility is less likely, given what John does say about Davy's convalescence. But if this is the case, then Davy first distinguished the two theories in the 1808 paper, and only later blended them together.

In any case, these excerpts from Davy's unpublished notes are revealing, not just for the details that they provide about the phlogistic and electro-chemical theories, but for hinting at Davy's purpose for entertaining either theory in the first place. Both theories have to do with his "Opinions concerning the Elements of Bodies."

We can see something resembling the blended electro-chemical phlogiston theory in a lecture from 1811, in which a digression on alchemy leads Davy to speculate about the following theory:

The analogy of the properties of the metals . . . would all lead us to the idea of their not being entirely different kinds of matter; but would rather incline one to suppose that they contain some common element or elements. . . . A series of proportions may be formed in which the metals may be supposed composed of hydrogen, and another substance in definite quantities; and, in this hypothesis, the lightest would contain the largest quantity of hydrogen, and possess as they are found to possess, the strongest attraction for oxygen and chlorine. (Davy (1840), 330)

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

To be sure, Davy doesn't try and correlate the electrical and chemical states of bodies. And there is no explicit mention of phlogiston here. But the idea of metals as composed of hydrogen plus a base, is enough to show that this is some kind of modified phlogiston theory. Moreover, Davy uses the example of ammonia to support these speculations, just as he does with the phlogistic theory in his 1808 paper. And this theory bears a strong resemblance to the theory that Davy discusses in his unpublished notes. The main difference is that, on Davy's 1811 theory, there is only one unknown base, and the metals differ from one another insofar as they differ in their proportions of hydrogen and this unknown base.

Siegfried has shed some light on the question as to why Davy would consider such theories in the first place. Based on a close examination of Davy's work, he argues that Davy was led to consider defensible modifications of the phlogiston theory because of his dissatisfaction with the number of chemical elements on the oxygen theory, and his desire to reduce the number of elements (Siegfried (1964), 118–119).<sup>36</sup> Siegfried supports this claim with the above-quoted passage from the 1811 lecture, among others (Siegfried (1964), 121–123). They include passages from another lecture from 1811, in which Davy discusses the issue of definite proportions, and claims that “the true solution of this most important part of philosophy will be found in a great simplification of elementary materials” (Davy (1840), 344). Siegfried also discusses a passage from Davy's book *Elements of Chemical Philosophy*, in which Davy claims that “by supposing [the metals] and the inflammable bodies different combinations of hydrogen with another principle as yet unknown in the separate form; all phenomena may be easily accounted for, and will be found in harmony

---

<sup>36</sup>See also Chang (2009), 249–250 for similar points.

with the theory of definite proportions” (Davy (1812a), 275). Moreover, the excerpts from Davy’s unpublished notes, which Siegfried does not discuss, add further support to his claim. In particular, the section title in the unpublished notes (“*My Opinions concerning the Elements of Bodies*”) does bear out his claim that Davy’s interest in the phlogiston theory has something to do with the chemical elements.

Siegfried’s conclusion, however, needs to be modified. He isn’t careful to distinguish the phlogistic and electro-chemical theories, and his analysis is therefore not sensitive to Davy’s claim that the electro-chemical theory is a hypothesis according to which “still fewer elements than those allowed in the antiphlogistic or phlogistic theory might be maintained” (Davy (1808b), 368). But Siegfried’s point can plausibly be extended to the electro-chemical theory as well. It is perhaps still within the spirit of Siegfried’s analysis to claim that Davy’s concern with both the phlogistic and electro-chemical theories, in both their separate and blended forms, has to do with reducing the number of chemical elements.

Siegfried claims that there came a point when Davy no longer entertained the phlogistic theory, namely, after the publication of *Elements of Chemical Philosophy* in 1812 (Siegfried (1964), 124). And in February of 1814, he comes closest to an outright rejection of the phlogiston theory:

The chemists in the middle of the last century had an idea, that all inflammable bodies contained phlogiston or hydrogen. It was the glory of LAVOISIER to lay the foundations for a sound logic in chemistry, by shewing that the existence of this principle, or of other principles, should not be assumed where they could not be detected. (Davy (1814), 71–72)

## CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

If Siegfried is right, then it's natural to suppose that Davy also stopped entertaining the electro-chemical theory around the same time. We cannot be sure without a detailed study of Davy's later writings. But this will take us too far afield, as we now have enough of the historical details to offer solutions to the three puzzles discussed above.

### 3.3.2 THREE PUZZLING ISSUES

At this point, it remains to solve the three puzzles discussed above. The first puzzle I'll discuss has to do with Davy's complicated relationship to the phlogistic and antiphlogistic theories. The solution to this puzzle will, in turn, help to solve the two remaining puzzles. These deal with the reasons for Davy's success with the muriatic and oxymuriatic acids, and his puzzling views on the elementary nature of chlorine, respectively. After discussing all three puzzles and their solutions, I'll argue that there's an important moral that we should draw from all of this. Specifically, because Davy simultaneously entertained two theories that turned out to be wrong in many important respects, he was able to make progress on the issue of the nature of the muriatic and oxymuriatic acids. In this case, we have an example from the history of science in which two wrongs make a right.

#### **Puzzle 1: Davy's Relationship to the Phlogistic and Antiphlogistic Theories**

As we've seen, Davy's relationship to the phlogistic and antiphlogistic theories is somewhat complicated. Davy claims that, while Scheele's phlogiston theory is "vague and unfounded," it also "may be considered as an expression of facts"; the oxygen theory, while "beautiful and satisfactory" in appearance, "rests in the present state of our knowledge,

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

upon hypothetical grounds” (Davy (1810b), 237). More generally, as we’ve seen, Davy often contrasts explanations that follow from the oxygen theory, with those that follow from the phlogiston theory. And he has both positive and negative things to say about each of these theories—both have some points in their favor, and both have trouble accounting for at least some phenomena. The puzzle, then, is to account for Davy’s attitudes towards these theories.

One way in which we might be tempted to solve the puzzle is by claiming that Davy accepted certain parts of each theory, and rejected certain other parts of each theory. But Davy’s general distrust of theories tells against such a solution. To take one example, Davy claimed in an 1808 paper that “the age of chemistry is not yet sufficiently mature for such discussions; the more subtle powers of matter are but just beginning to be considered; and all general views concerning them, must as yet rest upon feeble and imperfect foundations” (Davy (1808b), 369). This claim applies equally to all theories. To take another example, in a lecture given in 1811, Davy laments the fact that, for years, he “experienced great difficulties in conquering the prejudices adopted from the French school of chemistry,” and described his mind as being “fettered” by the French theories (Davy (1840), 312). This general distrust of theories has not gone unnoticed, and indeed Siegfried has argued that it has its source in the critical response to “hasty generalization” that Davy made early in his career (Siegfried (1959), 198–199). Davy went on to regret these hasty generalizations, and Siegfried further argues that Davy’s rejection of the atomic theory is a further consequence of his distrust of theories. In sum, then, Davy’s general distrust of theories tells against a piecemeal solution to the puzzle that would have him accept certain parts, and reject other

parts.

Instead, the solution to the puzzle involves recognizing that there are at least three attitudes that we can take towards a theory. One can accept or reject a theory.<sup>37</sup> But one can also entertain a theory, without accepting it or rejecting it.<sup>38</sup> The sense of entertaining that I have in mind here is the sense according to which one's primary concern is not the truth or falsity of the theory, but what one can do with the theory—one entertains the theory for the sake of argument. This is the attitude that Davy took towards both the phlogiston and oxygen theories for a time—he entertained both theories without accepting or rejecting either. Davy's earlier writings, like the "Prospectus of Experiments" from 1799, may indicate that Davy did accept the oxygen theory at one point, though this is not completely clear. In any case, at least by the time Davy delivered his 1807 Bakerian Lecture, he entertained both theories without accepting or rejecting either. Beginning with 1808's "Electro-Chemical Researches . . .," Davy also began to entertain another theory, namely, the electro-chemical theory, without accepting or rejecting it. By the time Davy put forward his view of the muriatic and oxymuriatic acids in 1810's "Researches on the oxymuriatic Acid . . .," he rejects the oxygen theory, at least when it comes to acidity. And by 1814, Davy also rejects the phlogiston theory. But, for at least two years, between 1807 and 1809, Davy entertained both theories without accepting or rejecting either.<sup>39</sup>

---

<sup>37</sup>One might claim that there are many more attitudes that one can take towards a theory, since one can accept a theory as true, or as empirically adequate, etc. This does not affect my point, but it's worth pointing out that these may be understood as different ways of understanding what the attitude of acceptance does or should involve, as opposed to distinct attitudes of acceptance.

<sup>38</sup>It may be the case that there is room for some overlap here, in the sense that one might simultaneously entertain a theory while also accepting it or rejecting it. I remain non-committal about this possibility.

<sup>39</sup>Siegfried has claimed that, between 1807 and 1812, although Davy considered a number of phlogistic explanations of various phenomena, he never rejected the oxygen theory (Siegfried (1964), 118). Siegfried's claim may be technically correct, insofar as Davy still entertains various explanations that the oxygen theory

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

Of course, one may still ask: ‘Why entertain the phlogiston theory at all?’ Here, Siegfried’s work supplies us with the answer. It is very plausible that Davy’s concern with the phlogistic and electro-chemical theories stems from a desire to reduce the number of chemical elements that we would be committed to on the oxygen theory. And, as Siegfried makes clear, “[Davy’s] discussions of [the phlogistic theory] were always clearly indicated as speculative and carefully distinguished from observed behavior” (Siegfried (1964), 124). Davy’s aim, namely, the reduction of the number of chemical elements, would seem to be a scientifically respectable one, especially given his method of carefully marking out speculations whenever he employed them.

However, as Knight remarks, Davy’s concern with the phlogiston theory may still seem puzzling if one takes the view that the phlogiston theory ceased to be scientifically respectable after Lavoisier’s arguments against it (Knight (1992), 68). But, as a growing number of scholars have recognized, this view is simply false. Hasok Chang is representative of this growing consensus, when he claims that “a reasonably close look at the primary literature should make it evident that there were numerous chemists who decided not to jump on the Lavoisier bandwagon even by 1790 and beyond” (Chang (2010), 63). Chang goes on to examine the primary literature, and brings in help from recent scholarship in the history of chemistry in order to support this claim.<sup>40</sup> Indeed, in the course of his survey, Chang takes special notice of Davy as “[p]erhaps the most interesting case of the new generation of anti-Lavoisier chemists,” and notes that he “entertained various systems of

---

provides. Moreover, we don’t see Davy explicitly rejecting the oxygen theory as a whole. But if I’m right, Davy certainly didn’t accept the oxygen theory at this time either—he merely entertained it. Moreover, by 1810, he certainly rejected the oxygen theory of acidity.

<sup>40</sup>See See Chang (2010), 63–68 and Chang (2012b), 30–34 for Chang’s discussion, along with citations to the primary and secondary literature.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

chemistry involving the revival of phlogiston” (Chang (2010), 67). Davy, then, numbers among many respectable chemists who entertained the phlogiston theory after 1790. I’ll go on to argue below that Davy’s use of the phlogiston theory wasn’t just scientifically respectable—it was scientifically productive, in the sense that it led to progress on the issue of the nature of the muriatic and oxymuriatic acids.

At this point, one potentially puzzling aspect still remains. It’s clear why Davy, in the paper in which he successfully reports the composition of muriatic acid, claims that the oxygen theory “rests in the present state of our knowledge, upon hypothetical grounds” (Davy (1810b), 237). But why, in the very same passage, does Davy side with Scheele’s phlogiston theory?

To this question, two answers can be given. First of all, while Davy’s praise for Scheele doesn’t entail the former’s acceptance of the phlogiston theory, Davy does think that Scheele’s view gets something right. Scheele’s view, like Davy’s view, has it that oxymuriatic acid is simpler than muriatic acid, and, indeed, the former is a constituent of the latter. This is one sense in which Scheele’s view “may be considered as an expression of facts” (Davy (1810b), 237). Davy’s characterization of Scheele’s view in the opening of the paper is illustrative of the same point:

THE illustrious discoverer of the oxymuriatic acid considered it as muriatic acid freed from hydrogen; and the common muriatic acid as a compound of hydrogen and oxymuriatic acid; and on this theory he denominated oxymuriatic acid dephlogisticated muriatic acid. (Davy (1810b), 231).

Moreover, once one identifies hydrogen with phlogiston, Scheele's view can be seen as correct in another sense, given that muriatic acid contains hydrogen, while oxymuriatic acid does not.

Secondly, as Knight has pointed out, "triumphing over the French was something never very far from Davy's mind" (Knight (1992), 68–69). Some of the puzzling aspects of this passage may then dissolve if one thinks that, for Davy, any opportunity to disparage the French is a good one.

### **Puzzle 2: Why Davy? Why Not Gay-Lussac and Thénard?**

When it comes to the nature of the muriatic and oxymuriatic acids, why is it that Davy succeeded, whereas Gay-Lussac and Thénard failed? As we've seen, both Davy, and Gay-Lussac and Thénard, succeeded in determining the composition of another as-yet-undecomposed acid, namely, boracic acid. Both succeeded in isolating the so-called boracic radical, which we now know as boron. To be sure, Gay-Lussac and Thénard do consider the idea that oxymuriatic acid "is a simple body," and hold that "[t]he phenomena . . . can be explained well enough on this hypothesis"; but they go on to claim that the phenomena "are still better explained by regarding oxygenated muriatic acid as a compound body" (Gay-Lussac and Thénard (1905), 48).<sup>41</sup> Davy then goes on to argue for the very hypothesis that Gay-Lussac and Thénard here reject. Moreover, as Golinski points out, in making this argument, "Davy brought forward no dramatically new experimental evidence" (Golinski (1992), 223). If Davy and the French chemists were in possession of the same

---

<sup>41</sup>Translation of Gay-Lussac and Thénard (1809), 358.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

experimental evidence, then we cannot appeal to such evidence to explain why the former succeeded where the latter failed. What difference can we point to, then, in order to explain why Davy succeeded and Gay-Lussac and Thénard failed?

The heart of the solution I will offer appeals to the fact that Davy, on the one hand, and Gay-Lussac and Thénard, on the other, had very different theoretical backgrounds. The French chemists accepted the oxygen theory, which turned out to be inconsistent with the true nature of the muriatic and oxymuriatic acids. Although Gay-Lussac and Thénard saw the possibility of the elementary nature of oxymuriatic acid, there's a sense in which their commitment to the oxygen theory prevented them from making progress on the nature of the acids. Davy, however, entertained both the oxygen and phlogiston theories, while accepting neither. And as we've seen, there are some ways in which the phlogiston theory is a better guide to the true nature of these acids. In entertaining both theories, Davy possessed a kind of theoretical openness, in the sense that he lacked any commitment to those theories that may have blocked his success.

This is also consistent with the fact that all three chemists succeeded when it came to determining the composition of boracic acid. As it turns out, boracic acid contains oxygen, and so accords with the oxygen theory. Gay-Lussac and Thénard were able to succeed because they accepted the oxygen theory. Similarly, we can attribute Davy's success to the fact that he entertained the oxygen theory. Hence, in this case, the three chemists shared the relevant theoretical background. In the case of boracic acid, this shared theoretical background suffices to explain why all three succeeded. In the case of the muriatic and oxymuriatic acids, the difference in theoretical background accounts for Davy's success

and Gay-Lussac and Thénard's failure.

Based on this difference in theoretical background, one might offer a kind of Kuhnian solution to the puzzle. In *The Structure of Scientific Revolutions*, Kuhn asks why Lavoisier was able to see oxygen gas in Joseph Priestley's experiments, when Priestley himself was unable to see it. In order to answer this question, Kuhn appeals to an important difference between Lavoisier and Priestley: "Long before he played any part in the discovery of the new gas, Lavoisier was convinced both that something was wrong with the phlogiston theory and that burning bodies absorbed some part of the atmosphere" (Kuhn (1996), 56). Priestley, obviously, had no such conviction, and Kuhn claims that this is why Lavoisier could see oxygen as an agent in Priestley's experiments, while Priestley himself saw dephlogisticated air. Given the opportunity, Kuhn might have explained the difference between Davy and the French chemists in similar terms. Gay-Lussac and Thénard accepted the oxygen theory, and so could not see oxymuriatic acid as a simple body; instead, they saw it as a compound body containing oxygen. Davy, on the other hand, did not accept the oxygen theory, or, for that matter, the phlogiston theory, though he entertained both. Because of his attitude toward these theories, he, unlike Gay-Lussac and Thénard, could see oxymuriatic acid as a simple body.

However, based on what Gay-Lussac and Thénard write, it doesn't seem plausible to attribute their failure to an inability to see something that Davy could see. After all, they explicitly consider the hypothesis that oxymuriatic acid is a simple body, only to reject it. Hence, Gay-Lussac and Thénard saw the same possibility that Davy did. In that case, we must look to some other difference between the two, and the difference in theoretical

background is a good candidate. The claim that Gay-Lussac and Thénard failed to see the same possibility as Davy may be literally false. But it does seem impossible to deny that theoretical background can, to a large extent, color a scientist's perceptions of various phenomena, and influence her decisions to put forward claims in print, and to this extent, a weakened version of the Kuhnian solution should suffice.

Other scholars have noted differences between Davy and the French chemists that might explain why the former, but not the latter, succeeded. Here I'll note two such differences. Bernadette Bensaude-Vincent points to one possible difference, namely, national allegiance. She notes that French chemists at this time downplayed any inconsistencies in Lavoisier's system (Bensaude-Vincent (1983), 69). Of Gay-Lussac and Thénard in particular, she claims that they "strongly suspected that they were dealing with an element, but had not dared deal such a blow to the doctrine of the master," namely, Lavoisier (Bensaude-Vincent (1983), 70). Davy, on the other hand, had no such allegiance, as we've already seen.

Maurice Crosland has pointed to another difference that may go some length towards explaining Davy's success and Gay-Lussac and Thénard's failure, namely, a difference in temperament. While Gay-Lussac was quite hesitant to put forward any claims that went beyond the data that he collected, Davy was more confident (Crosland (1980), 107). Gay-Lussac's temperament interacted with his education in such a way as to foster his commitment to the oxygen theory. Crosland notes that Gay-Lussac's education

had impressed on him the orthodoxy of the French scientific establishment  
...What Gay-Lussac really feared was the responsibility of innovation. He

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

could work within the framework of established ideas without assuming any great burden. (Crosland (1980), 109)

Gay-Lussac, in marked contrast to Davy, “did not want to destroy Lavoisier’s theory but only to amend it where necessary” (Crosland (1980), 114). Crosland claims that, while Gay-Lussac’s education led him to be a more disciplined chemist than Davy, the former “lost the freedom to explore any area of the natural world from any point of view” (Crosland (1980), 111). Insofar as Davy was both confident and free in a sense that Gay-Lussac was not, we can explain why only the former succeeded with the muriatic and oxymuriatic acids.

Bensaude-Vincent and Crosland both present essential aspects of a solution to this puzzle; but national allegiance and temperament by themselves do not suffice for a complete explanation of Davy’s success. I suggest that national allegiance and temperament be seen as reasons for the difference in theoretical background between Davy and the French chemists, as opposed to reasons for their differential success, and that this difference in theoretical background is the primary reason for Davy’s success and Gay-Lussac and Thénard’s failure. Gay-Lussac and Thénard’s acceptance of the oxygen theory, then, most likely involves a complicated mix of empirical evidence, national allegiance, and temperament. The same goes for Davy, who, as we’ve seen, entertained both the oxygen and phlogiston theories while accepting neither. The important point for my purposes is that the French chemists did, in fact, accept the oxygen theory, while Davy did not. The kind of freedom that Crosland claims that Gay-Lussac lacks is one that Davy possesses—the fact that he entertained both theories without accepting either allowed him to “explore any area

of the natural world from any point of view.”

It’s interesting to note that, given his general distrust of theories, which I have discussed above, Davy may have given an account of his own success similar to the one that I defend here. In a lecture from 1811, in which he discusses the muriatic and oxymuriatic acids, he goes on to rail against the oxygen theory, and also against a specific use of theories more generally. Davy notes that he himself “experienced great difficulties in conquering the prejudices adopted from the French school of chemistry, and strengthened by its nomenclature” (Davy (1840), 312). He accuses “the French inquirers [of] clos[ing] for nearly a third of a century this noble path of investigation,” namely, the path that Davy was advocating (Davy (1840), 318). Even more generally, Davy claims:

Nothing is so fatal to the progress of the human mind as to suppose that our views of science are ultimate; that there are no mysteries in nature; that our triumphs are complete, and that there are no new worlds to conquer. (Davy (1840), 318)

It is passages like these that John Hedley Brooke has in mind when he argues that Davy was opposed to being bound by theories in general (Brooke (1980), 129).<sup>42</sup> This general opposition to being bound by theories fits within the view that I’ve argued for here, namely, that Davy entertained both the oxygen and phlogiston theories, while neither accepting nor rejecting either, and that this is the reason for his success regarding the muriatic and oxymuriatic acids.

---

<sup>42</sup>In a similar vein, Knight claims that, “perhaps because of his caution about large scale theorizing, [Davy] was often at his best demolishing the theories of others” (Knight (1992), 69). See also Knight (1978), 136.

**Puzzle 3: Is Chlorine an Element for Davy?**

As the quotation from Remsen makes clear, Scheele is credited with the discovery of chlorine, while Davy is credited with establishing that it is an element. And, as Chang puts it, “[Davy] made his name . . . by putting a nail in the coffin of Lavoisier’s theory of acids with his argument that chlorine was an element and that muriatic acid (hydrochloric acid) did not contain oxygen, only hydrogen and chlorine” (Chang (2012b), 33).

But in a number of passages, Davy is fairly noncommittal regarding the elementary nature of chlorine. He claims that it “has not as yet been decomposed,” and is “elementary as far as our knowledge extends” (Davy (1811a), 1). And later in the same paper, he admits the possibility that “[t]here *may* be oxygene in oxymuriatic gas; but [he] can find none” (Davy (1811a), 35). If Davy admits this possibility, it’s difficult to see him as establishing the elementary nature of chlorine, and thereby killing off the oxygen theory of acidity. The puzzle, then, is to explain why the chemist who is credited with these achievements says the things that he says about chlorine.

In order to solve this last puzzle, it will first be necessary to get clear on Davy’s views concerning the nature of elements. Once again, Siegfried’s work on Davy provides us with the necessary information. Siegfried argues that Davy’s views on the nature of elements owe much to Lavoisier’s (Siegfried (1964), 119). According to Lavoisier,

if, by the term *elements*, we mean to express those simple and indivisible atoms of which matter is composed, it is extremely probable we know nothing at all about them; but, if we apply the term *elements* or *principles of bodies*, to

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

express our idea of the last point which analysis is capable of reaching, we must admit, as elements, all the substances into which we are able to reduce bodies by decomposition. (Lavoisier (1802), xxvii–xxviii)<sup>43</sup>

As Lavoisier recognizes, on this use of the term ‘element,’ it may be the case that a substance is classified as an element today, and loses this status tomorrow as the result of some advance in chemical analysis. Davy is largely in agreement with Lavoisier, as we can see from the following passage from 1802. At the beginning of a section titled: “OF UNDECOMPOUNDED SUBSTANCES OR SIMPLE PRINCIPLES,” Davy writes:

Though the corpuscular theory supposes the existence of bodies composed of similar particles; yet we are not certain that any such bodies have been yet examined. The simple principles of the chemists are substances which have not been hitherto composed or decomposed by art; and they are elements, only in relation to other known substances. (Davy (1802), 4–5)

Davy, like Lavoisier, refuses to identify elements with atoms, and instead identifies them with undecomposed substances.

As Siegfried has pointed out, once one understands Davy’s views on the nature of elements, his initially puzzling remarks about chlorine actually make sense (Siegfried (1959), 195–196). Given these views, Davy would have to admit the possibility that chlorine may eventually be decomposed, and may even contain oxygen. But according to Davy, we should consider chlorine an element because it hadn’t been decomposed yet.

---

<sup>43</sup>Translation of Lavoisier (1801), xvii.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

If this is the case, then, even in light of Davy's work, the oxygen theory has a glimmer of hope, which depends on the existence of oxygen in chlorine. Interestingly, Davy's response to this possibility involves the claim that "at present we have no more right to say that oxymuriatic gas contains oxygene than to say that tin contains hydrogene," thereby putting the oxygen theory in the same boat as the phlogiston theory (Davy (1811a), 34) Davy's work, then, still amounts to strong evidence against the oxygen theory, since, as he points out, there is no reason to think that chlorine contains oxygen.

Davy's desire to reduce the number of chemical elements, which led to his concern with the phlogistic and electro-chemical theories, also fits with what he says about chlorine. Given this desire, it makes sense that Davy would hold out some hope that chlorine would be decomposed. Davy is, of course, famous for the discovery of a number of elements, and it's a bit ironic that he would have held out some hope for their decomposition. Even so, given his views, such hope is understandable.

This way of solving the puzzle explains why Davy says the things that he says. But one might object to it, on the grounds that it doesn't quite make sense of why Davy is credited with the achievements which we typically credit him. Given his views on the nature of elements, it seems that Davy presents a rather weak case for the conclusion that chlorine is an element in the sense of our current understanding of the nature of elements. And if Davy allows a glimmer of hope to the oxygen theory, he can hardly be seen as putting the nail in its coffin.

But rather than amounting to an objection to my solution, this actually illustrates an important point. There are many cases from the history of science in which there is a gap

between the achievements with which a scientist is credited, and the way in which that scientist conceives of her own achievements. Kenneth Caneva has discussed such cases in which “what a scientist is typically credited with having discovered often differs significantly from the way in which the scientist himself characterized his work” (Caneva (2001)).<sup>44</sup> Caneva discusses a number of examples, one of which is that of Hans Christian Ørsted, who is credited with discovering electromagnetism. We now take the term ‘electromagnetism’ to name a certain class of phenomena that Ørsted was presumably among the first to investigate. However, Ørsted himself applied the term, not to a class of phenomena, but to a specific form of action that is operative in what we now call electromagnetic phenomena. For the reasons discussed in the previous paragraph, Davy’s work on chlorine amounts to another example of this general trend. The way we tend to conceive of his achievements is quite different from the way he conceived them.

### **Two Wrongs *Can* Make a Right**

Based on my solution to the three puzzles, we can see that there’s a strong sense in which the fact that Davy entertained two false theories (the phlogiston theory and the oxygen theory) actually contributed to his success in determining the nature of the muriatic and oxymuriatic acids. In this case, we have an example from the history of science in which two wrongs make a right.

I’ve argued that, in the time leading up to Davy’s success, there were at least a couple of years in which he entertained both theories, while neither accepting nor rejecting ei-

---

<sup>44</sup>See also Caneva (2005) for a much more in-depth discussion.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

ther theory. And I went on to argue that, in taking this attitude toward both theories, Davy's theoretical background did not prohibit him from considering and defending possibilities at odds with the oxygen theory. The same could not be said for Gay-Lussac and Thénard, who did not share in Davy's success when it came to the muriatic and oxymuriatic acids. Moreover, his reasons for entertaining both theories were quite reasonable. Davy is clear that both theories had their merits and flaws—each dealt with some phenomena adequately, and each had trouble dealing with some phenomena. And Davy's concern with the phlogiston and electro-chemical theories was driven by a desire that was quite reasonable at the time, namely, the desire to reduce the growing number of chemical elements. In Davy's work, then, we have an example from the early nineteenth century in which a chemist employed the phlogiston theory, in some capacity, in order to bring about scientific progress.

This conclusion is largely in agreement with Chang's view of the phlogiston and oxygen theories in the late eighteenth and early nineteenth centuries. Chang argues for the claim that phlogiston actually suffered a premature death—there was no good reason to give it up, and keeping it around would have led to progress in science (Chang (2011), 412–423). More recently, he has engaged in some actual and counterfactual history of science, in order to argue for a number of conclusions (Chang (2012b), 42–50). First of all, the phlogiston theory had a number of points in its favor that the oxygen theory lacked. Secondly, the death of the phlogiston theory actually retarded scientific progress. Thirdly, the interaction between the phlogiston and oxygen theories actually had a number of beneficial effects. Fourthly, and lastly, this interaction would have had even more beneficial effects had the phlogiston theory not been rejected. Chang suggests that the correct conclusion to draw

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

from all of this is what he calls “conservationist pluralism,” which is a rejection of the idea that the sciences have room for only one dominant theory (Chang (2011), 428). Such a pluralistic view would have allowed for the side-by-side development of the oxygen and phlogiston theories.

The example of Davy’s work fits fairly comfortably within Chang’s view. Although he mentions Davy’s concerns with the phlogiston theory (Chang (2012b), 33), Chang doesn’t devote any kind of detailed discussion to Davy’s use of that theory in the course of his work. But, insofar as I am correct that Davy’s success with the muriatic and oxymuriatic acids is partially due to the fact that he entertained both the phlogiston and oxygen theories, this example lends some support to Chang’s conclusions.

#### 3.4 THE AFTERMATH OF DAVY’S WORK AND THE DEATH OF THE MURIATIC RADICAL

By the time Remsen composed the passage that I quoted in the introduction to this chapter, something like Davy’s view of chlorine and muriatic acid came to prevail. But, as we’ve already seen, his view was changed in the process—Davy would have characterized his achievements in a way different from the way in which we now tend to characterize them. We now tend to see Davy as giving us conclusive evidence for the elementary nature of chlorine, and for killing off the oxygen theory of acidity. This is what I will refer to as the orthodox view. And according to that view, the muriatic radical is a dead hypothetical entity—it was born in Lavoisier’s oxygen theory of acidity, which Davy destroyed, and

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

could not survive outside of it for long.<sup>45</sup>

However, the muriatic radical did not die immediately after Davy's work on the acids culminated in 1810. In fact, it lived on in various forms for quite some time. While the subsequent history of the muriatic radical deserves close attention, I'll confine myself to a brief discussion of the aftermath of Davy's work.

A number of chemists still maintained some form of the oxygen theory, and defended it against Davy's arguments. Jöns Jacob Berzelius and John Murray are notable in this regard.<sup>46</sup> Both accept the claim that oxymuriatic acid combines with hydrogen to form muriatic acid. But they hold on to the idea that oxymuriatic acid contains oxygen, and that muriatic acid therefore contains water. They attempt to show that this view accommodates the facts at least as well as Davy's view, if not better.

Berzelius's view is particularly noteworthy insofar as he explicitly employs the "muriatic radicle," which "is capable of combining with different doses of oxygen" (Berzelius (1813), 254). He admits that muriatic acid "cannot be obtained in a separate state," and "does not seem capable of existing except in combination with some oxide or other" (Berzelius (1813), 254). He claims that, "[w]hen combined with water, it constitutes common muriatic acid gas" (Berzelius (1813), 254). But he goes on to write of a simpler substance, which he also calls 'muriatic acid,' that does not contain water. This simpler substance therefore bears a striking resemblance to the dry muriatic acid that Davy discusses in the 1809 appendix of his 1808 Bakerian Lecture (Davy (1809c), 468), and in

---

<sup>45</sup>Chang holds something like the orthodox view as I have described it. See Chang (2012b), 33 for his claims about Davy's work on the elementary nature of chlorine, and his subsequent destruction of the oxygen theory of acidity. See Chang (2011), 417 for Chang's claim that the muriatic radical doesn't exist.

<sup>46</sup>See, for example, Berzelius (1813), Berzelius (1816), and Murray (1811a).

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

order to be clear about Berzelius' view, I will call muriatic acid without water 'dry muriatic acid.' Berzelius employs Dalton's terminology, and claims that dry muriatic acid "is composed of one atom of radicle and two atoms of oxygen," while oxymuriatic acid "consist[s] of one atom of radicle combined with three . . . atoms of oxygen" (Berzelius (1813), 254). On Berzelius' view, if one adds hydrogen to oxymuriatic acid, the result is muriatic acid. This is because the hydrogen combines with some of the oxygen of oxymuriatic acid to form water, which is necessary for muriatic acid. Oxymuriatic acid therefore contains more oxygen than dry muriatic acid, though muriatic acid and oxymuriatic acids would seem to contain the same amount of oxygen.

When it comes to Murray's attacks on Davy, John Davy came to his brother's aid, and went on to defend the latter's view, entering into a prolonged dispute with Murray.<sup>47</sup>

One might think that the issue was decided in the Davys' favor once the dispute between Murray and John Davy concluded. But this is not the case. As late as 1858, we see something like the muriatic radical in the work of Christian Friedrich Schönbein. In a letter to Humphry Davy's former assistant, Michael Faraday, Schönbein reports his view that chlorine contains oxygen and what he calls "murium," which is the term that Remsen employs in his gloss on the history of chlorine. Hydrochloric acid, for Schönbein, contains murium and water (Schönbein (1858), 180).<sup>48</sup> Schönbein's theory contains some additional aspects that suggest that murium and the muriatic radical are not quite identical. But it's clear that the two are at least closely related, insofar as they share some striking similarities.

But by 1871, we can see something like the orthodox view in the work of William

---

<sup>47</sup>See, for example, Davy (1811b), Murray (1811b), and Davy (1812b).

<sup>48</sup>See also Schönbein (1857), 254.

CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY:  
THE CASE OF THE MURIATIC RADICAL

---

Odling. Odling discusses the views of Davy, Berzelius, and Schönbein, among others. He goes on to conclude that “whatever may hereafter prove to be its real nature, chlorine has hitherto proved undecomposable, and is accordingly regarded by the generality of chemists as an element” (Odling (1871), 2). If chlorine is an element, then there is, of course, no room for the muriatic radical.

While Odling frames his discussion in terms of Schönbein’s murium, Josiah Parsons Cooke frames his in terms of the muriatic radical. In an 1889 article in *Popular Science*, Cooke, like Odling, presents something like the orthodox view. After a brief discussion of Lavoisier’s boracic, fluoric, and muriatic radicals, in which Cooke claims that only the first of these conformed to Lavoisier’s theory, Cooke writes the following:

With the exception of caloric and two of the radicals above referred to [the muriatic and fluoric radicals], Lavoisier’s list of elements includes no substance not regarded as elementary at the present day ... (Cooke (1889), 743)

The fact that Cooke groups the two radicals with caloric suggests that he thinks that the radicals are not elementary because of the fact that they simply don’t exist, in accordance with the orthodox view mentioned above. But just two pages before this passage, Cooke claims that, in Lavoisier’s day, “the radical of muriatic acid, chlorine, was then a well-known substance, having been discovered by Scheele in 1774” (Cooke (1889), 741). This identification of chlorine with the muriatic radical would, of course, allow the latter to survive as an element, in which case it’s curious that Cooke dismisses this possibility just two pages later.

### CHAPTER 3. A HISTORY OF A HYPOTHETICAL ENTITY: THE CASE OF THE MURIATIC RADICAL

---

This completes my history of the muriatic radical. In the next and final chapter of the dissertation, I will draw on the details of this history in order to argue against the extant views of hypothetical entities, and for the suppositionalist view. In order to do this, I will consider some questions one might reasonably raise, given the puzzling things that Cooke says about the muriatic radical in his article. Does the muriatic radical exist? And if so, what could it be?

## Chapter 4

---

### *An Objection to the Extant Views and a Defense of Suppositionalism*

#### 4.1 INTRODUCTION

In chapter 3, I examined a case from the history of chemistry involving a hypothetical entity, namely, the muriatic radical. Toward the end of the eighteenth century, Antoine Lavoisier hypothesized that all acids are composed of the principle of acidity (oxygen), combined with an acidifiable base or radical. The composition of muriatic acid was not well understood at this time, and so Lavoisier predicted that it contains oxygen combined with the muriatic radical, a hitherto undiscovered element. Chemists continued to work with muriatic acid, but their efforts to isolate the radical did not meet with success. Then, in the early years of the nineteenth century, Humphry Davy succeeded in determining the composition of muriatic acid—it contains hydrogen and chlorine, and is what we now call hydrochloric acid (HCl). This spelled bad news for Lavoisier's theory of acidity, and although there were subsequent attempts to defend the hypothesis that there is such a muriatic radical, these efforts did not succeed, and chemists eventually stopped entertaining this

hypothesis.<sup>1</sup>

My goal in this chapter is to use the case of the muriatic radical as a test case for views of hypothetical entities. In chapter 1, I extracted a number of extant views of hypothetical entities from scientific realism, constructive empiricism, and the various versions of structural realism. According to the extant views, the HE term<sup>2</sup> ‘muriatic radical’ is a putatively referring expression that has putative reference to an empirical entity. The suppositionalist view, which I discussed in chapter 2, has it that ‘muriatic radical’ refers to a hypothetical entity, understood as an object of supposition. My aim is to argue that, based on the historical details of the muriatic radical case, the suppositionalist view is preferable to the extant views.

To that end, in section 2, I will begin by arguing that the extant views are committed to some substantive claims regarding the HE term ‘muriatic radical,’ and I will outline these claims. My argument in this chapter makes use of two claims, which I defend in section 3. First of all, as a matter of fact, the term ‘muriatic radical’ today fails to refer to an empirical entity. Secondly, there are counterfactual histories according to which the term ‘muriatic radical’ would refer to an empirical entity today. The question that I pose to proponents of the extant views and the suppositionalist view is: When Lavoisier published his *Elements* in 1789, to what did the term ‘muriatic radical’ refer? In section 4, I consider the answers that proponents of the extant views could give. I argue that their answers commit them to give either a whig history or an incomplete history of the muriatic radical. I conclude that

---

<sup>1</sup>The historical details will be important to my argument in what follows. In this chapter, I will attempt to give a clear explanation of these details when possible. But the reader is advised to refer to chapter 3 for more detailed explanations and citations to the primary source material.

<sup>2</sup>‘HE term’ is short for ‘hypothetical entity term.’

this gives us good reason to reject the extant views. Finally, in section 5, I consider the answers that a proponent of the suppositionalist view could give. Based on these answers, I conclude that the suppositionalist view is preferable to the extant views.

## 4.2 THE EXTANT VIEWS AND THE MURIATIC RADICAL

The historical details of the previous chapter give us some examples of scientific discourse concerning a hypothetical entity, namely, the muriatic radical. It is my claim that the extant views of hypothetical entities should have something to say about the muriatic radical, and my first aim in this section is to answer an objection to this claim. I will then lay out what the extant views are committed to.

### 4.2.1 THE SCOPE OF THE EXTANT VIEWS

Scientific realism, constructive empiricism, and the various forms of structural realism are the three main positions in the scientific realism debate, and, as such, they are supposed to tell us something about how to understand science. But there is some reason to think that these positions are not unrestricted in scope, in the sense that they are meant to apply, not to the whole of what we call science, but to a proper part of it. To take one recent example, Anjan Chakravartty characterizes scientific realism as “a positive epistemic attitude towards the content of our best theories,” as opposed to all theories (Chakravartty (2011)). I have drawn the example of the muriatic radical from Lavoisier’s oxygen theory of acidity, and I claim that the extant views should have something to say about this example. But it may be objected that the extant views are restricted in scope, and that Lavoisier’s oxygen theory of

acidity falls outside of the scope of their application. Chakravartty, for example, may claim that, since the oxygen theory of acidity is not one of our best theories, scientific realism is not applicable to it. And the proponents of constructive empiricism and the various forms of structural realism may say the same thing. If this objection is successful, then I would be unable to use the example of the muriatic radical in order to argue against the extant views.<sup>3</sup>

Before attempting to answer this objection, it will be necessary to explain the need for such a restriction in scope. It traces back to a certain kind of response to Larry Laudan's attack on scientific realism (Laudan (1981), 22–29, 32–36).<sup>4</sup> The realist, according to Laudan, claims that the explanatory and predictive success of our theories is good evidence that those theories are approximately true, and that their theoretical terms are genuinely referential. Laudan's attack on this claim involves a laundry list of counterexamples, namely, successful theories with central theoretical terms that do not refer, even by the realist's own admission. The theories in question therefore can't be true or even approximately true. The list includes the caloric theory of heat, the humoral theory of medicine, the wave theory of light, and, importantly for my concerns here, the phlogiston theory of chemistry (Laudan (1981), 33).

Realists have responded to Laudan's attack in a number of ways, many of which involve attempts to reduce the size of Laudan's list of counterexamples. One way, which I will call the 'restriction strategy,' involves the restriction of realism to the so-called 'ma-

---

<sup>3</sup>I thank Steven Gross and Ronald Duchovic for raising this objection, and for helping me to recognize the need to devote more thought to answering it.

<sup>4</sup>In what follows, 'realism' should be understood as 'scientific realism.' When I mean to be talking about another form of realism, I will indicate this in the text.

ture' sciences. The basic idea is that some theories on Laudan's list, like the humoral theory of medicine and the phlogiston theory of chemistry, are drawn from sciences which had not yet reached maturity. And since realism is not meant to apply to immature sciences, these theories cannot serve as counterexamples to the realist's claims. Clyde L. Hardin and Alexander Rosenberg were the first to use the restriction strategy against Laudan (Hardin and Rosenberg (1982), 609–610), and more recently, realists like Stathis Psillos have followed their lead in this regard (Psillos (1999), 107–108).

To be sure, the realists who Laudan is attacking are explicit about the fact that realism applies only to the mature sciences,<sup>5</sup> and so it's no surprise that Laudan anticipates the restriction strategy. He argues that it is inadequate for two reasons (Laudan (1981), 34–35). First of all, he points out that the realist wants to explain the success of science. But if we restrict the application of realism to the mature sciences, then realism leaves the success of the *immature* sciences a mystery. Secondly, Laudan argues that, without an account of what maturity amounts to, there is a danger of rendering realism tautologous. As Laudan notes, realists sometimes claim that realism is an empirical hypothesis about the sciences. But if 'mature sciences' are understood as 'sciences of which the tenets of realism are true,' then realism is transformed from an empirical hypothesis into a tautology. Hence, in order for the restriction strategy to work, the realist needs an independent handle on what counts as a mature science.

Philosophers of science have not been altogether silent when it comes to giving an account of maturity. Hardin and Rosenberg suggest that the distinction between mature and

---

<sup>5</sup>See, for example, Putnam (1978), 20–21.

immature sciences be understood as analogous to Thomas Kuhn's distinction between pre-paradigmatic and paradigmatic sciences (Hardin and Rosenberg (1982), 609–610).<sup>6</sup> They also rely on Richard Boyd's notion of a "take-off point" in order to give an account of maturity. According to Boyd, a "take-off point" is "a point in the development of the relevant scientific discipline at which the accepted background theories are sufficiently approximately true and comprehensive" (Boyd (1981), 627). While Hardin and Rosenberg admit that the notion of a take-off point is "far from sharp," they suggest that a mature science be understood as one that has already reached its take-off point (Hardin and Rosenberg (1982), 609), and, once again, Psillos has followed their lead (Psillos (1999), 107–108). John Worrall has also been critical of realists for leaving the notion of maturity undefined, and has gone on to develop his own account. For Worrall, "a science counts as mature once it has theories within it which are ... predictive of general *types* of phenomena, without these phenomena having been 'written into' the theory" (Worrall (1989), 114).<sup>7</sup> The kind of predictive success that Worrall has in mind is what philosophers of science sometimes call 'novel predictive success.'<sup>8</sup>

It should now be clear why some philosophers of science have adopted the restriction strategy, and how that strategy is supposed to work. At this point, I will respond to the objection that the muriatic radical falls outside of the scope of the extant views.

**Response 1:** *Not all of the extant views are committed to the restriction strategy.* In particular, constructive empiricism and ontic structural realism (OSR) are not committed

---

<sup>6</sup>See Kuhn (1996).

<sup>7</sup>Emphasis is the author's unless otherwise noted.

<sup>8</sup>See Musgrave (1988), 232 for an example of the use of this terminology.

to the restriction strategy. This is because proponents of these positions do not see Laudan's argument as a threat. Bas van Fraassen holds that, according to constructive empiricism, empirical content is the only part of a theory's content that is always preserved over theory change (van Fraassen (2001), 163; van Fraassen (2008), 110–111). And James Ladyman and Don Ross claim that, according to OSR, structure, in addition to empirical content, is always preserved over theory change, and that there are no 'Kuhn-losses' in such cases (Ladyman and Ross (2007), 157). Proponents of both constructive empiricism and OSR, then, hold that they can account for theory change in a way that does not run afoul of Laudan's argument. Moreover, both van Fraassen and Ladyman explicitly apply their positions to the realist's paradigm case of a theory from an immature science, namely, the phlogiston theory (van Fraassen (1992), 14–20; van Fraassen (2008), 84–85, 261; Ladyman (2011)).<sup>9</sup> In this case, the muriatic radical does fall within the scope of these positions.

We've already seen that realists are committed to the restriction strategy, and as it turns out, epistemic structural realists are as well. Worrall commits himself to this strategy, and claims that the size of Laudan's list can indeed be reduced if one uses Worrall's account of maturity (Worrall (1989), 114–115). Epistemic structural realism (ESR) and realism come apart when it comes to their explanations of novel predictive success, and since Worrall believes that the realist explanation is flawed, he goes on to develop ESR. But, for my purposes, we can treat realism and ESR together, insofar as both are committed to the restriction strategy. In what follows, then, I'll focus on realism, with the understanding that

---

<sup>9</sup>See also van Fraassen (1980), 32, where van Fraassen applies his constructive empiricism to a hypothetical example from pre-Lavoisierian chemistry, which is presumably just as much a paradigm case of an immature science, at least according to the realist.

what I have to say also applies to ESR.

**Response 2:** *Some realists who adopt the restriction strategy explicitly claim that realism is meant to apply to Lavoisier's oxygen theory.* In order to rule out the phlogiston theory as a potential counterexample to realism, some realists have claimed that chemistry did not reach maturity until Lavoisier. To take one example, in their discussion of the distinction between a mature and an immature science, Hardin and Rosenberg claim that “[i]f one is to draw such a line in chemistry, for example, it would most plausibly come with the publication of Lavoisier's *Elements of Chemistry* and thus would exclude phlogiston theory as a counterexample” (Hardin and Rosenberg (1982), 610). In this case, Lavoisier's oxygen theory of acidity would be drawn from a mature science, and so the muriatic radical would fall within the scope of realism.

Incidentally, there is some reason to think that this way of adopting the restriction strategy won't accomplish the realist's aim. As we've seen in the previous chapter, phlogiston lived on for quite some time after Lavoisier's death. In this case, the line that Hardin and Rosenberg draw is not even effective in ruling out the phlogiston theory as a counterexample. In response, realists may want to claim that chemistry was still immature at the beginning of the nineteenth century. But such realists will have to be careful that they don't face the problem that I will discuss next.

**Response 3:** *The restriction strategy cannot be too restrictive—otherwise, the extant views will have nothing to say about theory change.* The realist may be tempted to identify our best theories with those we currently accept, and to claim that the sciences reached maturity when those theories became accepted. But realists have taken great pains to explain

various features of theory change, and have developed fairly worked-out views concerning what does and does not get preserved over theory change. If realists who adopt the restriction strategy do give in to this temptation, then they will have nothing to say about theory change. Realists must therefore claim that some of our best theories are theories that we no longer accept, and some sciences reached maturity before our current theories were accepted. In that case, it is possible that the muriatic radical will fall within the scope of the extant views.

**Response 4:** *There is good reason to think that chemistry was a mature science by the time Lavoisier developed his oxygen theory.* Of course, in order to make this claim, it is necessary to say what maturity amounts to, and so I will consider the proposals discussed above.

Worrall argues that a science counts as mature once it has theories that are capable of novel predictive success. Lavoisier's oxygen theory of acidity is such a theory. As we've seen in the previous chapter, according to the oxygen theory, boracic acid should contain oxygen and the boracic radical, a hitherto unknown element. And both Joseph Louis Gay-Lussac and Louis Jacques Thénard, on the one hand, and Humphry Davy, on the other hand, went on to decompose boracic acid, proving that it contains oxygen and boron. Things are a bit complicated by the fact that boracic acid has the chemical formula  $H_3BO_3$ , and Lavoisier's hypothesis makes no mention of hydrogen. Nonetheless, his prediction that the acid would contain oxygen and a hitherto unknown element turned out to be correct. And so we have a case of novel predictive success, since this result was unknown to Lavoisier at the time he formulated his theory. And in that case, chemistry was mature by the time

Lavoisier formulated his theory.

Although this seems to me to be a clear case of a theory that enjoyed novel predictive success, I admit that there is room for disagreement here. One might deny that the oxygen theory enjoyed such success, and attempt to defend the claim that chemistry during Lavoisier's time was immature. But, assuming that we accept Worrall's account of maturity, in order to defend this latter claim, one would have to show that there were no examples of theories in chemistry that enjoyed novel predictive success before Lavoisier's death.<sup>10</sup> Although this kind of issue cannot be settled *a priori*, it is highly unlikely that chemistry during Lavoisier's time was immature in this sense.

Hardin and Rosenberg tell us that the distinction between mature and immature sciences is supposed to be analogous to Kuhn's distinction between pre-paradigmatic and paradigmatic science. Kuhn himself claims that Lavoisier's *Elements* ushered in a new paradigm (Kuhn (1996), 10), though he also writes of the "phlogiston paradigm" (Kuhn (1996), 57) which this new paradigm replaced. If we use the notion of a paradigm as Kuhn understood it, then we would have to conclude that chemistry was a paradigmatic science well before Lavoisier. And if the realist wants to understand the notion of a paradigm in such a way that chemistry did not reach maturity even during Lavoisier's lifetime, then she owes us an account of what this notion amounts to. Otherwise, the notion of a paradigm can provide no guidance regarding the maturity of a science.

Hardin and Rosenberg also claim that a science has reached maturity once it has reached what Boyd calls a "take-off point," which requires the relevant background theories to

---

<sup>10</sup>See, for example, Hasok Chang's discussion of some examples of novel predictive success that were shared by Lavoisier's oxygen theory and the phlogiston theory (Chang (2012b), 54).

be approximately true and comprehensive. I must admit that I'm not at all sure whether chemistry had reached a take-off point by the time Lavoisier published his *Elements*. It's certainly true that chemists at this time had gotten very good at performing various kinds of experiments, and at analyzing and synthesizing various kinds of substances. But it's not clear whether a set of background theories is at work here, as opposed to general know-how. Moreover, if the set of background theories makes reference to entities like caloric, then it's unlikely that they are approximately true. With regard to this understanding of maturity, it may be the case that chemistry in Lavoisier's day had not reached a take-off point, and was therefore immature. But this would conflict with what Hardin and Rosenberg believe concerning Lavoisier's *Elements*. And, rather than being a reason in favor of thinking that chemistry at this time was immature, this may indicate a problem with Boyd's notion of a take-off point. Realists will have to ask themselves whether this notion is worked out well enough in order to do the work that realists require of it. This leads to my next response.

**Response 5:** *Laudan's objections to the restriction strategy still apply.* One of the ways in which Laudan objects to the restriction strategy is to point out that, if by 'mature sciences,' realists mean 'sciences of which the tenets of realism are true,' then realism is transformed from an empirical hypothesis into a tautology. As we've seen, realists have attempted to answer Laudan's objection by appealing to Boyd's notion of a take-off point in order to provide an independent account of what maturity amounts to. Hardin and Rosenberg admit that the notion is "far from sharp," but it doesn't seem to me to be of *any* help in answering Laudan's objection. This is because take-off points require the approximate truth of background theories, and it is presumably one of the tenets of realism that back-

ground theories are approximately true. So, the notion of a take-off point does not give us an independent handle on maturity.

The other way in which Laudan objects to the restriction strategy is to point out that if realists restrict the application of realism to the mature sciences, then realism leaves the success of the *immature* sciences a mystery. An analogous point can be made against a realist who restricts the application of realism so as to exclude Lavoisier's oxygen theory of acidity. In this case, the realist will be unable to explain success in general. And, more importantly for my purposes here, the realist will be unable to offer an account of hypothetical entities in general.

**Response 6:** *Even if the muriatic radical falls outside the scope of the extant views, there are most likely similar examples of hypothetical entities that do fall within their scope.*

Although I am relying on the example of the muriatic radical in order to argue against the extant views, I do not think that this is the only hypothetical entity that can raise problems for those views. I suspect that one could make the same points in terms of another hypothetical entity, from another branch of science, or from another time period. To take some examples from nineteenth- and twentieth-century physics, which was surely a mature science by this point, one might use particles like the electron, positron, and neutron in order to raise problems for the extant views.<sup>11</sup> I won't be able to defend this claim here. But it is important to emphasize that I'm not claiming that everything rides on the case of the muriatic radical.

**To Sum Up:** In general, the realist who wants to object to my claim, and deny that

---

<sup>11</sup>For some brief remarks on the neutron in this regard, see Rynasiewicz et al. (2010), 10–11.

## CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

realism applies to Lavoisier's theory of acidity and the muriatic radical, will have to do a number of things. She will have to draw a distinction between the theories within the scope of realism, and the theories outside of its scope. This distinction should be worked out well enough so as to be able to tell us of specific parts of science whether or not realism applies to them. It may be a distinction between mature and immature sciences, between pre-paradigmatic and paradigmatic sciences, or between our best theories and our 'other' theories. What she says about such distinction will have to entail that Lavoisier's theory falls outside of the scope of realism. It will also have to entail that realism applies to more than just our current theories, at least if realism is still supposed to tell us something about theory change. And such a distinction will also have to be principled—the way in which it sorts the history of science should not be *ad hoc* or arbitrary. The responses I have made do not amount to a proof that this is impossible. But they do show that it is difficult, and the burden of proof is on the realist to show that such a position can be made to work.

### 4.2.2 WHAT THE EXTANT VIEWS ARE COMMITTED TO

Now that I have responded to the objection that the muriatic radical falls outside of the scope of the extant views, we can see what these views have to say about that hypothetical entity. These views of the muriatic radical are easily recoverable from the characterizations of the extant views from chapter 1. Here, then, by way of review, and for the sake of clarity, are what those views are committed to regarding the muriatic radical.

Scientific realism is committed to  $R_{HE}$ , according to which the term 'muriatic radical' is a putatively referring expression that has putative reference to an empirical entity.

According to the realist trichotomy, ‘muriatic radical’ either

(R<sub>T1</sub>) refers to an empirical entity; or

(R<sub>T2</sub>) it altogether fails to refer to an empirical entity; or

(R<sub>T3</sub>) it ‘kind of’ refers to that entity.

The epistemic structural realist view is a variation on the scientific realist view. This view is also committed to R<sub>HE</sub>, but failure of reference is not possible, as is reflected by ESR’s commitment to (R<sub>T1</sub>), but not (R<sub>T2</sub>) or (R<sub>T3</sub>).

Constructive empiricism is committed to CE<sub>HE</sub>, according to which the term ‘muriatic radical’ is a putatively referring expression that has putative reference to either an empirical entity or a mathematical entity. According to the constructive empiricist trichotomy, ‘muriatic radical’ either

(CE<sub>T1</sub>) refers to either an empirical or a mathematical entity; or

(CE<sub>T2</sub>) it altogether fails to refer to an empirical or mathematical entity; or

(CE<sub>T3</sub>) it ‘kind of’ refers to that entity.

Ontic structural realism is committed to OSR<sub>HP</sub>, according to which the HP locator ‘muriatic radical(L)’<sup>12</sup> is a putatively referring expression that has putative reference to a real pattern. According to the trichotomy that OSR is committed to, ‘muriatic radical(L)’ either

(OSR<sub>T1</sub>) refers to a real pattern; or

---

<sup>12</sup>See Ladyman and Ross (2007), 122 for this terminology.

(OSR<sub>T2</sub>) it altogether fails to refer to a real pattern; or

(OSR<sub>T3</sub>) it ‘kind of’ refers to that real pattern.

Before moving on, one point should be emphasized. I’ve framed the commitments of the extant views in terms of the HE term ‘muriatic radical.’ But, as I pointed out in the previous chapter, there are a number of other HE terms that chemists employed to name the same entity, including ‘muriatic base,’ ‘muriatic basis,’ ‘acidifiable base of muriatic acid,’ and ‘muriatic radical principle.’ In short, the extant views are committed to the same claims regarding these terms, as they are for ‘muriatic radical.’

### 4.3 ACTUAL AND COUNTERFACTUAL HISTORIES OF THE MURIATIC RADICAL

My aim is to argue against the extant views of hypothetical entities from chapter 1, and in favor of the suppositionalist view from chapter 2. My argument will make use of the actual history of the muriatic radical, which I discussed in chapter 3, along with some counterfactual histories. In particular, my argument will make use of two claims about the reference of ‘muriatic radical’ today. Before stating these two claims, it’s worth emphasizing that, today, ‘muriatic radical’ is not an HE term. This is because scientists no longer entertain any hypotheses involving the muriatic radical, in the sense that they are no longer concerned with attempting to confirm or reject such hypotheses. Indeed, such hypotheses have been rejected. The first claim, then, is that today, the term ‘muriatic radical,’ so understood, actually fails to refer to an empirical entity. The second claim is

that there are counterfactual histories according to which the term ‘muriatic radical’ would refer to an empirical entity today.

#### 4.3.1 ACTUAL HISTORY: FAILURE OF REFERENCE

If one looks at the actual historical details of the case of the muriatic radical, then the term ‘muriatic radical’ fails to refer to an empirical entity today.

If there were such an empirical entity as the muriatic radical, then based on what Lavoisier tells us, it would, at the very least, have to be a component of muriatic acid. Since muriatic acid has exactly two components, there are exactly two candidates that fit the bill so far: hydrogen and chlorine. At this point, we can examine each possibility.

First of all, it may seem that a good case can be made for the claim that ‘muriatic radical’ refers to hydrogen. Before Davy conducted his work on muriatic acid, Christoph Girtanner claimed that the base of muriatic acid is hydrogen; and Lavoisier’s English translator, Robert Kerr, suggests that the name ‘hydrogen’ might therefore be rejected in favor of ‘muriogen,’ with an obvious nod to the muriatic radical (Lavoisier (1802), 122). However, one might object to the claim that hydrogen is the radical, on the grounds that the only way in which to isolate the radical of an acid is to remove the principle of acidity, i.e., to subtract all of the oxygen, from that acid. And as muriatic acid contains no oxygen, there is no way of discerning which of its components is the radical. In reply to this objection, one could cite Davy’s speculations regarding the nature of chlorine (Davy (1810b), 243, 250). After noting some analogies between oxygen and chlorine, Davy considers the possibility that chlorine is another principle of acidity. On this idea, we could isolate the radical by

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

subtracting the oxygen-like substance from the acid, namely, chlorine. This might seem to provide some support for the claim that ‘muriatic radical’ refers to hydrogen.

But in order for this claim to stand, it cannot be the case that there are equally good reasons in favor of the claim that ‘muriatic radical’ refers to chlorine. In order for this second possibility to appear at all plausible, one would need some reason to think that in muriatic acid, hydrogen is playing the role of a principle of acidity. Indeed, some chemists in the early nineteenth century claimed that hydrogen does sometimes play such a role. For example, in John Webster’s 1826 textbook, one finds the following passage:

oxygen is not essential to the acidity of a compound, for some bodies are rendered acid by union with chlorine, others by hydrogen; and the theory of Lavoisier which considered oxygen as the essential principle of acidity, and in conformity to which its present name was assigned to it can no longer be received as correct. (Webster (1826), 88)

Since Webster also admits that chlorine can play the same role, this passage provides no guidance when it comes to determining which of the constituents of muriatic acid is the radical. However, based on two of the three theories of acidity that we accept today, it is hydrogen, and neither oxygen nor chlorine, which plays an essential role in acids. A substance is an Arrhenius acid if it increases the concentration of hydrogen ions in water, and a substance is a Brønsted-Lowry acid if it is a proton donor.<sup>13</sup> If hydrogen is playing the role of the principle of acidity, then that leaves chlorine as the radical. Our current theories

---

<sup>13</sup>According to the third theory, a substance is a Lewis acid if it is an electron pair acceptor. See Chang (2012a) for a discussion of these theories of acidity and the ways in which they do not line up.

of acidity, then, might seem to lend some support to the claim that ‘muriatic radical’ refers to chlorine.<sup>14</sup>

Based on the discussion in the previous two paragraphs, one might suspect that we can determine which substance ‘muriatic radical’ refers to by comparing the strengths of the reasons in favor of each possibility. But this strategy doesn’t hold much promise. The reasons on each side are more-or-less equally strong, and this is because all of the reasons are pretty bad.

First of all, we can consider the reasons in favor of claiming that ‘muriatic radical’ refers to hydrogen. Davy may have been correct to note some analogies between chlorine and oxygen. But as he himself recognizes, these analogies are not sufficient to conclude that chlorine is another principle of acidity. This latter claim is mere speculation. Moreover, it’s a claim that cannot be sustained in light of any of the theories of acidity that we currently accept. As we’ve seen, if any element is playing an essential role in acids, it is hydrogen. In that case, it’s difficult to support the claim that ‘muriatic radical’ refers to hydrogen.

Secondly, we can consider the reasons in favor of claiming that ‘muriatic radical’ refers to chlorine. Given what our current theories of acidity tell us about hydrogen, one might hold that an application of the principle of charity would yield the conclusion that, when Lavoisier spoke of the principle of acidity, we can understand him as referring to hydrogen. But, as Hilary Putnam has emphasized, even charity can be overdone (Putnam (1978), 22, 25). It’s one thing to say that, based on two of our current theories of acidity, hydrogen plays an essential role in acids. But it’s quite another thing to say that our current theories

---

<sup>14</sup>See Cooke (1889), 741 and Bensaude-Vincent (1983), 69–70 for the identification of chlorine and the muriatic radical.

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

of acidity allow any room for the notion of a unique principle of acidity, in Lavoisier's sense of a substance which renders something an acid, just by means of being one of its components. It's clear that they do not, and it's worth pointing out that acids needn't contain hydrogen. Boron trifluoride ( $\text{BF}_3$ ), for example, is an acid that contains no hydrogen—it is a Lewis acid, since it is an electron pair acceptor. Moreover, it's highly implausible to understand Lavoisier's use of 'principle of acidity' as referring to hydrogen, and not to oxygen. He chose the name 'oxygen' because of the purported acid-generating properties of that substance (Lavoisier (1802), 101). And so, if Lavoisier's 'oxygen' refers to oxygen, and not to hydrogen, then the same is true of Lavoisier's 'principle of acidity.'<sup>15</sup> In that case, it's also difficult to support the claim that 'muriatic radical' refers to chlorine.

There is a more general route by which we can reach essentially the same point. If one wants to claim that the term 'muriatic radical' refers to an empirical entity, it would have to refer either to hydrogen, or to chlorine. In order to determine which entity the term refers to, one would have to answer the question: In muriatic acid, which component is the base or radical, and which is the acidifying principle? But as Hasok Chang has pointed out, this is a meaningless question from the point of view of contemporary chemistry (Chang (2011), 417). Moreover, there are other constraints on the muriatic radical that neither hydrogen nor chlorine can satisfy. For example, according to Lavoisier, the muriatic radical was supposed to unite with oxygen to form oxymuriatic acid (i.e., chlorine) as well. In this case, the most natural conclusion to draw is that today, 'muriatic radical' refers neither to

---

<sup>15</sup>But see Chang (2011), 417 for some considerations in favor of the claim that Lavoisier's 'oxygen' actually referred to nothing at all.

hydrogen, nor to chlorine.<sup>16</sup>

But this still leaves open the possibility that ‘muriatic radical’ ‘kind of’ refers to either hydrogen, or to chlorine, or to both. In order to conclude that ‘muriatic radical’ lacks a referent altogether, we first need a reason to eliminate this possibility.

Psillos’ notion of approximate reference (Psillos (1996), S313) and Hartry Field’s notion of partial denotation (Field (1973), 474) are two such views of ‘kind of’ reference.<sup>17</sup> According to Psillos’ notion, a term that was originally used with the intention of referring to a posit that has been abandoned, may approximately refer to a current posit. This would be the case if the abandoned posit and the current posit share some attributes, and are meant to play similar roles, in the sense that the current posit is responsible for causing the phenomena that the abandoned posit was initially thought to have caused. Field, on the other hand, motivates his notion by appeal to the example of mass. According to Field, “Newton’s word ‘mass’ partially denoted proper mass and partially denoted relativistic mass; since it *partially* denoted each of them, it didn’t *fully* (or *determinately*) denote *either*” (Field (1973), 474). More generally, a term that fails to fully denote some entity may, in fact, partially denote that entity.

With regard to the case of the muriatic radical, then, one might think that, even if ‘muriatic radical’ doesn’t fully refer to some empirical entity, it may be the case that it has an approximate referent or a partial denotation. In order for such a possibility to be plausible, it would have to be the case that ‘muriatic radical’ approximately refers to and/or partially

---

<sup>16</sup>Although he isn’t explicit about it, it is presumably reasoning like this that leads Chang to deny that the muriatic radical exists (Chang (2011), 417; Chang (2012a), 692).

<sup>17</sup>See chapter 1 for more discussion of these two views.

denotes *both* hydrogen and chlorine. This is because, based on my discussion above, there is no reason to prefer one of these entities to the other. Both of these entities share some of the properties ascribed to the muriatic radical. For example, both have the property that they enter into the composition of muriatic acid. Moreover, both are causally responsible for the phenomena that the muriatic radical was supposed to be causally responsible for, namely the behavior of muriatic acid. In short, if the reasons for thinking that ‘muriatic radical’ refers to hydrogen are just as good as those for thinking that it refers to chlorine, then one might conclude that ‘muriatic radical’ approximately refers to and/or partially denotes both.<sup>18</sup>

But taking this option is problematic for a number of reasons. To see this, we can first consider Psillos’ notion of approximate reference. This notion is clearly not as worked-out as Field’s partial denotation, and Psillos admits that approximate reference is in need of further development (Psillos (1996), S313). While it could certainly be the case that approximate reference proves to be a fruitful notion, working out the details will involve some serious challenges. In particular, it’s difficult to see how it could be the case that a term for an abandoned posit could ever fail to approximately refer to the current posit. After all, the two posits will presumably always share some attributes. And insofar as the current posit is a replacement for the abandoned posit, the two will always be thought to play similar roles, in the sense that the current posit is responsible for causing the phenomena that the abandoned posit was initially thought to have caused. If the notion is trivialized in this way, it’s not clear that it can prove to be fruitful at all. In any case, when it comes to

---

<sup>18</sup>Similarly, Field claims that “there is no basis for choosing between” the hypothesis that Newton’s ‘mass’ referred to proper mass, and the hypothesis that it referred to relativistic mass (Field (1973), 467).

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

the muriatic radical, it's clear that the approximate reference account would have to mirror the partial denotation account, in the sense that the term 'muriatic radical' would have to approximately refer to and/or partially denote both hydrogen and chlorine. It is to this latter notion that I now turn.

In order to get clearer on partial denotation, it will help to have some paradigm cases in mind. John Earman and Arthur Fine have pointed out that Field's central example of Newton's 'mass' is, in fact, poorly chosen (Earman and Fine (1977)). Fortunately, other philosophers have pointed to some examples that may work better. Field himself suggests that 'gene' partially denoted cistrons, mutons, and recons (Field (1973), 477).<sup>19</sup> Richard Boyd mentions two such examples (Boyd (2010), 216). The first is the term 'element,' as used before chemists drew a distinction between elements and isotopes. Before this distinction was drawn, the term partially denoted both. The second is the term 'species.' Given the plurality of definitions of 'species,' and the different kinds that these definitions pick out, the term 'species' partially denotes all of these kinds. Finally, P. Kyle Stanford and Philip Kitcher have claimed that 'acid,' as used by chemists before the twentieth century, partially denoted Arrhenius acids, Brønsted-Lowry acids, and Lewis acids (Stanford and Kitcher (2000), 119–120).

The common thread among these three examples is that they are all examples of kind splitting—what was once thought to be a single kind turns out, upon closer examination, to be two or more distinct kinds.<sup>20</sup> If the muriatic radical case is one that involves kind

---

<sup>19</sup>Field actually uses the gene example to extend his account of partial denotation from singular terms and quantity terms to general terms, of which 'gene' is his example. He introduces the terminology "partially signify" for the latter, but since the account that he gives is analogous to that for partial denotation, I will ignore this difference in terminology.

<sup>20</sup>These kinds needn't be natural. They can be functional kinds, or merely kinds that are scientifically

## CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

splitting, then we would have good reason to claim that ‘muriatic radical’ partially denotes both hydrogen and chlorine. However, this case is most naturally understood as one that involves the elimination of one or more kinds, as opposed to one that involves kind splitting. On Lavoisier’s theory, the radical of an acid is that which combines with oxygen to form that acid. When it comes to muriatic acid, it’s not the case that there are, in some sense, two radicals, whereas chemists once thought there was one. It’s that the whole theory, according to which acids have radicals, is off the mark. The notion of a radical hasn’t been split—it’s been eliminated. Therefore, the muriatic radical case is not a good candidate for a treatment in terms of partial denotation.

At this point, we have eliminated the possibility that ‘muriatic radical’ ‘kind of’ refers to an empirical entity. We can therefore draw the conclusion that today, ‘muriatic radical’ altogether fails to refer to an empirical entity.

### 4.3.2 COUNTERFACTUAL HISTORIES: SUCCESSFUL REFERENCE

However, I will now examine two counterfactual courses of events in order to argue that it didn’t have to be this way. These histories are *not* counterfactual in the sense that the outcomes of experiments are different from what they are in the actual history. In fact, I’ll stipulate that end-state chemistry would be the same in the actual and counterfactual histories. These histories are counterfactual, then, in the sense that chemists made decisions about the use of theoretical terms that are different from the decisions that chemists made in the actual history. More specifically, on these counterfactual histories, ‘muriatic radical’  

---

important for one reason or another.

would refer to an empirical entity today. In the first counterfactual history, it would refer to hydrogen; and in the second, it would refer to chlorine.

Before moving on to these counterfactual histories, it will be necessary to emphasize some of the actual historical details surrounding the muriatic radical case. The first set of details involves the notion of a base. Within the context of the oxygen theory of acidity, Lavoisier and Louis Bernard Guyton de Morveau held that ‘base’ and ‘radical’ are synonyms, and that they are the substances that are left over once one subtracts the oxygen from an acid (Guyton de Morveau et al. (1788), 33–34; Lavoisier (1802), 121–122). But chemists working in the late eighteenth and early nineteenth centuries often employed the notion of a base or basis without any commitment to the oxygen theory of acidity. Lavoisier, for example, employs the notion in his theory of gases, according to which gases are compounded of a base with caloric. Oxygen, then, is the base of oxygen gas (Lavoisier (1802), 63). To take another example, Davy at one point entertains the idea that the metals, charcoal, sulphur, phosphorus, and nitrogen are composed of hydrogen combined with an unknown base (Davy (1836), 405–406). Regarding theories of acidity in particular, Richard Kirwan employs the notion of a basis in his phlogiston theory of acidity (Kirwan (1789), 126). And Davy at one point entertains the idea that acids, along with the metallic oxides and alkalies, are composed of a base plus water (Davy (1808a), 33). To be sure, Davy’s ideas were admittedly mere speculation that did not pan out, and Kirwan’s and Lavoisier’s ideas turned out to be wrong as well. But this still shows that the notion of a base had a life outside of the oxygen theory of acidity. And if terms like ‘base’ can be divorced from theories like the oxygen and phlogiston theories of acidity, then they may, in fact, successfully

refer to empirical entities.

The second set of details shows that terms like ‘base’ can, in fact, successfully refer. It involves the case of boracic acid ( $\text{H}_3\text{BO}_3$ ). This acid had not been decomposed in Lavoisier’s day, and so he hypothesized the existence of the boracic radical, which was supposed to combine with oxygen to form the acid.<sup>21</sup> Lavoisier’s hypothesis was not very far off the mark, and towards the end of 1808, both Davy, on the one hand, and Gay-Lussac and Thénard, on the other, succeeded in decomposing boracic acid, and thereby isolating boron for the first time. Davy proposes the name “boracium” for what he previously called the “boracic basis” (Davy (1809a), 85), while Gay-Lussac and Thénard propose the name “bore” for what they previously called the “radical boracique” (Gay-Lussac and Thénard (1808), 173). It seems clear that these chemists succeeded in referring to boron with these newly proposed terms. But in this case, we have to admit that after they isolated boron, they could use older terms like ‘boracic radical’ and ‘boracic basis’ to refer to the same substance. This admission does not imply that chemists who used these older terms before 1808 could successfully refer to boron before it was isolated. Rather, it merely reflects what I take to be a truism, namely, that once Gay-Lussac, Thénard, and Davy succeeded in isolating the substance that we call ‘boron,’ they could refer to it with whatever term they preferred, whether it be the older terminology or some new term. And when these chemists introduce new terminology in the papers referenced above, ‘boracic basis’ and ‘radical boracique’ are best read as referring to the same substance that ‘boracium’ and ‘bore’ refer

---

<sup>21</sup>See Lavoisier (1802), 291 for Lavoisier’s inclusion of the boracic radical in his table of simple substances.

to, namely, the substance that they isolated—boron.<sup>22</sup> In this case, ‘boracic radical’ is now a successfully referring, though antiquated, term, and it refers to boron.

These actual historical details suggest some ways in which the history of the muriatic radical could have played out, so as to make it the case that ‘muriatic radical’ now refers to an empirical entity. More specifically, if Davy, or perhaps some other chemist, were to identify the muriatic radical with either hydrogen or chlorine, then ‘muriatic radical’ would have an empirical referent. The history of the boracic radical gives us some precedent for this kind of identification, and in the very same lecture in which Davy reports the isolation of the boracic basis, he reports his failure to isolate the muriatic basis (Davy (1809a), 99). The identification of this basis with either hydrogen or chlorine in subsequent work was thus a possibility. And while Davy may have been led to avoid making such an identification by the fact that muriatic acid defied the expectations of the oxygen theory of acidity, the above discussion of bases shows that one can simultaneously use the language of bases while rejecting that theory. In that case, it may be an accident of history, due to an unactualized possibility, that ‘muriatic radical’ now lacks an empirical referent. These unactualized possibilities have to do with what Davy says once he determines that muriatic acid is composed of hydrogen and chlorine.

We can easily imagine one counterfactual history, according to which ‘muriatic radical’ refers to hydrogen. As a matter of fact, Davy notes some analogies between oxygen and chlorine, and speculates that “muriatic acid may be considered as having hydrogen for

---

<sup>22</sup>This may tell against Chang’s claim that the boracic radical is a non-existent entity (Chang (2012b), 54). Though Chang’s claim may be consistent with the claim that a term for a non-existent entity can end up referring to an empirical entity.

its basis, and oxymuriatic acid for its acidifying principle” (Davy (1810b), 243). Davy is best understood as consciously speculating in this passage, and as I discussed above, this speculation is not sufficient to establish that ‘muriatic radical’ refers to hydrogen. But if Davy had been more confident, he may well have claimed that hydrogen *is* the basis of muriatic acid. We can even imagine him adopting Kerr’s suggestion, and pushing for a change in terminology, according to which the term ‘hydrogen’ should be replaced by ‘muriogen.’ To be sure, Davy would have been wrong to hold that chlorine is a principle of acidity. But there is no question that he was able to refer to both hydrogen and chlorine at this point, and he certainly could have chosen to refer to the former with terms like ‘muriatic radical,’ ‘muriatic basis,’ or ‘muriogen.’

We can also imagine a second counterfactual history, according to which ‘muriatic radical’ refers to chlorine. It could have been the case that Davy was impressed by another feature of the muriatic radical, namely, Lavoisier’s inclusion of that hypothetical substance on his table of simple substances (Lavoisier (1802), 291). Davy certainly thought that chlorine should be considered an element, as it had not yet been decomposed (Davy (1811a), 1). And so, based on his work on muriatic acid, he might have claimed that chlorine should occupy the spot on Lavoisier’s table that the muriatic radical once occupied. Although this would involve ignoring much of what Lavoisier hypothesized regarding the muriatic radical, it is within the realm of possibilities that Davy identify the muriatic radical with chlorine. Hydrogen already occupied a spot on the table, and so the most straightforward way in which to revise the table would be to substitute chlorine for the muriatic radical.

We can further imagine, that on each of these counterfactual histories, Davy’s identifica-

tion was recognized by the community of chemists with whom he worked. They may even have adopted his terminology. But at the very least, we can imagine that they recognized that he was using terms like ‘muriatic radical’ to refer to an empirical entity. Moreover, we can imagine that the story of Davy’s identification of the muriatic radical with either hydrogen or chlorine was eventually codified in the brief histories that often find their way into science textbooks.

It should be emphasized that all of the details of these two counterfactual histories are consistent with the eventual rejection of the language of bases and radicals. If chemists were to subsequently give up the terminology of ‘muriatic radical’ and ‘muriatic base,’ this would not be for the reason that Davy, according to these counterfactual histories, failed to refer to anything. Clearly, he could successfully refer to both hydrogen and chlorine. Instead, it would be for the reason that this kind of language is connected to theories that have since been rejected, like the oxygen and phlogiston theories of acidity and the caloric theory of heat.

There are many cases in the history of science, in which scientists reject a bit of terminology, but not because that terminology fails to refer. Davy himself opted for a change in terminology, from ‘oxymuriatic acid’ to ‘chlorine,’ not on the grounds that the former failed to refer, but on the grounds that it was inappropriate, given that the substance in question contains neither oxygen nor muriatic acid (Davy (1811a), 32). ‘Phlogiston,’ as used in the late eighteenth and early nineteenth centuries, is another example, though admittedly a more controversial one given what most philosophers have claimed about this term. Chemists like Kirwan and Joseph Priestley identified phlogiston with hydrogen (Kirwan

CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS  
AND A DEFENSE OF SUPPOSITIONALISM

---

(1789), 4–5; Priestley (1892), 33). And since they could refer to hydrogen, their tokens of ‘phlogiston’ are best understood as referring to hydrogen. Most philosophers have held ‘phlogiston’ to be a paradigm case of a non-referring term, but a look at the history of chemistry suffices to show that this is not the case, at least not by the end of the eighteenth century.<sup>23</sup> When Davy discusses Scheele’s view of chlorine, he makes this same point. Davy claims that “SCHEELE’S view, (though obscured by terms derived from a vague and unfounded general theory,) of the nature of the oxymuriatic and muriatic acids, may be considered as an expression of facts” (Davy (1810b), 237). Scheele held that oxymuriatic acid results from subtracting the phlogiston from muriatic acid. Davy rejects Scheele’s terminology, but still holds that it is an expression of facts because Davy understands ‘phlogiston’ as referring to hydrogen. Other examples might include the use of J. J. Thomson’s ‘corpuscle’ to refer to electrons and the use of Dmitri Mendeleev’s ‘ekasilicon’ to refer to germanium.

The important point for my purposes is that on the counterfactual histories I’ve discussed, subsequent chemists might have taken issue with Davy’s mode of expression, but without implying that he failed to refer. To be sure, the term ‘muriatic radical’ is not one that would be used very often in, say, 2013 to refer to either hydrogen or chlorine. But that does not imply that it would fail to refer. There are, in fact, cases of abandoned theoretical terms that succeeded in referring, and have been abandoned for reasons other than failure of reference. Just as the term ‘oxymuriatic acid’ refers to chlorine today, on these counterfactual histories, the term ‘muriatic radical,’ understood as a term that has putative

---

<sup>23</sup>Kitcher is an exception here. See Kitcher (1978), 534.

reference to an empirical entity, would refer either to hydrogen or to chlorine today.

Counterfactual history of science is controversial, and I won't engage in any kind of extended defense of it here.<sup>24</sup> But I will deal with two possible objections to the counterfactual histories that I've constructed. The first goes as follows. In the actual history, the match between the muriatic radical, on the one hand, and either hydrogen or chlorine, on the other, was insufficient for Davy and the rest of the community of chemists to identify the former with either of the latter. In that case, the match couldn't have been sufficient in either of the counterfactual histories. Moreover, the same issue comes up if we restrict our attention to the counterfactual histories. If the match was sufficient for chemists to identify the muriatic radical with hydrogen, then it could not have been sufficient for them to identify the muriatic radical with chlorine. The objection, then, is that the same match cannot be sufficient in one history and insufficient in another history, and that I have therefore done something illicit in constructing the counterfactual histories.

I grant that the match is the same in all three histories, but deny that this poses a problem. In the course of constructing the counterfactual histories, I've simply stipulated that different individuals in different histories may react to the same match in different ways. One individual (Davy in the actual history) did not associate the muriatic radical with any empirical entity. Another individual (Davy in the first counterfactual history) associated it with hydrogen. And a third individual (Davy in the second counterfactual history) associated it with chlorine. The objection assumes that the same match will prompt different individuals to make the same judgments. But this seems to me to be false, and, in any case,

---

<sup>24</sup>For some defense, see Chang (2012b), 62–65.

there is no problem in conceiving the counterfactual histories that I've discussed.

It may be further objected that this shows that in all except for one history, Davy and the rest of the community were acting irrationally, and this is the second objection. This objection, however, presupposes the claim that there is exactly one rational response to a given set of empirical evidence. This claim is, at the very least, contentious, and is incompatible with many accounts of rationality, such as those that explain rationality in terms of coherent sets of beliefs. But for the sake of argument, we can grant this claim. At this point, we can see that the objection rests on a further claim, namely, that a history in which a scientist acts irrationally is an illicit history, in some sense. This, however, would rule out, not just counterfactual histories, but also the actual history of science. In this case, the objection does not pose a problem for the counterfactual histories that I've discussed.

Examination of these counterfactual histories has an important consequence. At least part of the reason why 'muriatic radical' does not, in fact, refer to an empirical entity is that these counterfactual courses of events are, after all, counterfactual. Moreover, it's worth pointing out that the modifications that I've made to the actual history, in order to yield these counterfactual histories, are really quite minimal. In the counterfactual histories, I've left all of the details of the natural world intact, except for the decisions of Davy and other chemists. My conclusion about the reference of 'muriatic radical' in the actual history therefore, in a sense, depends, not just on what did happen, but also on what very easily could have but did not happen.

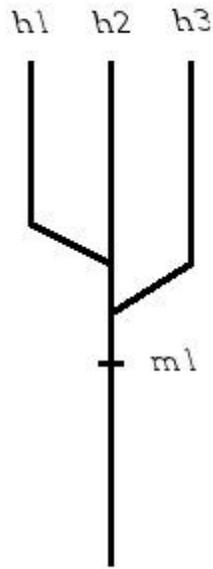


Figure 4.1: A diagram of the actual and counterfactual histories of the muriatic radical.

#### 4.4 THE PROBLEM WITH THE EXTANT VIEWS

I've just argued that, today, 'muriatic radical' fails to refer to an empirical entity, but that there are counterfactual histories according to which it would refer to an empirical entity today. We can picture the actual and counterfactual histories as they are represented in Figure 4.1. Let  $m_1$  be some moment in 1789, after the publication of Lavoisier's *Elements*. Time increases as one goes further up the diagram, and the branches represent the actual and counterfactual histories subsequent to  $m_1$ . Let  $h_1$  be the actual history, according to which, today, 'muriatic radical' doesn't refer to an empirical entity. Let  $h_2$  be the first counterfactual history, according to which, today, 'muriatic radical' refers to hydrogen. And finally, let  $h_3$  be the second counterfactual history, according to which, today, 'muriatic radical' refers to chlorine.

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

These actual and counterfactual histories are meant as test cases for the extant views of hypothetical entities and the suppositionalist view. In light of these histories, we can ask proponents of each of these views a further question about reference: At  $m_1$ , to what did the HE term ‘muriatic radical’ refer? It’s worth noting that this question is different from the question as to what ‘muriatic radical’ refers to today. The extant views give us some possible responses. It could have referred to an empirical entity. It could have ‘kind of’ referred to such an entity. Or it could have failed to refer altogether. The extant views, or at least a subset of them, would seem to be in good shape, so long as one of these possibilities is actually the case. And it would seem to be a straightforward matter of looking at Lavoisier’s work, along with the subsequent history of chemistry, in order to discover which, if any, of these possibilities is actually the case.

My aim in this section is to use these actual and counterfactual histories in order to raise a problem for the extant views. I will begin by raising the problem for realism. In short, I’ll use the realist trichotomy in order to raise a trilemma for the realist. After I do so, I will show that the other extant views face the problem as well.

First of all, there is the claim that ‘muriatic radical’ referred to an empirical entity at  $m_1$ . I will argue that this claim requires a commitment to temporal externalism, a view developed and defended by Henry Jackman, according to which the content of an utterance can depend on future linguistic usage (Jackman (1999)). I will argue that temporal externalists are forced to practice whig history of science, and that this horn of the trilemma should therefore be avoided.

Secondly, there is the claim that ‘muriatic radical’ ‘kind of’ referred to an empirical

entity at  $m_1$ . I will focus on Field's partial denotation, and develop this claim accordingly. I will argue that this view of 'kind of' reference also commits one to practice whig history of science, and that this horn of the trilemma is problematic for the same reason that the first horn is.

Thirdly, and finally, there is the claim that 'muriatic radical' failed to refer to an empirical entity at  $m_1$ . This is actually a claim that I agree with. But the realist is committed to the claim that the HE term 'muriatic radical' is a putatively referring expression that has putative reference to an empirical entity. So for the realist, all there is to the semantics of the HE term 'muriatic radical' is that it is an expression that fails to refer to an empirical entity. I will argue that this horn of the trilemma commits the realist to give an incomplete history of the muriatic radical.

Since the other extant views of hypothetical entities are quite similar to the realist view, they face the same problem. In the remaining section of the chapter, I will argue that suppositionalism is to be preferred, since it gives us a history of the muriatic radical that is neither a whig history, nor an incomplete history.

#### 4.4.1 'MURIATIC RADICAL' AS A SUCCESSFULLY REFERRING TERM

How would a realist respond to the question: At  $m_1$ , to what did the HE term 'muriatic radical' refer? The first response that I want to consider is that it referred to an empirical entity. But given that I have argued that, in the actual history  $h_1$ , 'muriatic radical' fails to refer to an empirical entity today, it would be implausible for realists to claim that 'muriatic radical' actually referred to an empirical entity at  $m_1$ . As a result, my argument will

make heavy use of the counterfactual histories  $h_2$  and  $h_3$ . In particular, I will consider the following response on behalf of the realist:

As a matter of fact, at  $m_1$ , the term ‘muriatic radical’ failed to refer to an empirical entity, and this is what we see in  $h_1$ . But if  $h_2$  had been the actual history, then at  $m_1$ , ‘muriatic radical’ would have referred to hydrogen. And if  $h_3$  had been the actual history, then at  $m_1$ , ‘muriatic radical’ would have referred to chlorine.

Such a response involves successful reference, but it is counterfactual, rather than actual. I will now argue that this response requires a commitment to temporal externalism. I’ll begin by summarizing the details of temporal externalism.

Jackman introduces temporal externalism as a further development of the externalist views developed by Saul Kripke (1980), Hilary Putnam (1975a), and Tyler Burge (1979). Jackman characterizes these views as, very roughly, committed to the claims that linguistic practices extend into the past, are located within a certain physical environment, and are shared by a society, respectively. Jackman’s innovation comes with the claim that these practices also extend into the future, and the content of an utterance can depend on future usage (Jackman (1999), 157–158).<sup>25</sup>

Jackman supports this view by appeal to two cases, one of which he draws from the history of zebra classification (Jackman (1999), 159–160).<sup>26</sup> According to the actual his-

---

<sup>25</sup>Jackman develops and defends temporal externalism in a number of papers, including Jackman (2004), Jackman (2005), and Jackman (2006). Here, I focus on his introduction of the position in Jackman (1999). For some criticism of temporal externalism, see Brown (2000).

<sup>26</sup>The other case is the druid case discussed in Wilson (1982), from which Jackman also draws the zebra case.

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

torical details, in 1820, the term ‘Grant’s zebra’ was introduced to name a kind of zebra which was found in Kenya. Later, the term ‘Chapman’s zebra’ was introduced to name a morphologically distinct kind of zebra, which was found in Zimbabwe. Once both terms were introduced, taxonomists discovered that both types interbred around the Zambezi river, and that one type gradually faded into the other. Grant’s and Chapman’s zebras, then, were two races of the species *Equus burchilli*. But it didn’t have to be this way. The taxonomists might have hit upon the zebras at the Zambezi river before reaching Zimbabwe. In this case, ‘Grant’s zebra’ would denote all of *Equus burchilli*, as opposed to the single race that we call ‘Grant’s zebra.’

Jackman draws the following moral from this case: In 1820, ‘Grant’s zebra’ referred to a specific race of *Equus burchilli* because of events that took place subsequent to 1820; and that it is a mere “historical accident” that ‘Grant’s zebra’ did not refer to all of *Equus burchilli*, as it would if taxonomists had reached the Zambezi river immediately after the introduction of the term ‘Grant’s zebra’ (Jackman (1999), 159). For Jackman, this is a clear case of future usage affecting the content of an utterance.

In order to generalize this moral, Jackman introduces some terminology. To begin with, he defines a *moment* as “a spatially complete but instantaneous event,” and introduces a causal ordering relation between moments ( $\leq$ ) (Jackman (1999), 162). A *history*, for Jackman, is “a maximal chain of moments” (Jackman (1999), 162). In order to evaluate the content of an utterance, one must specify both a moment and a history—content is undefined unless one specifies both.

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

Once both are specified, the content of an utterance can be either (1) *determinate* or *indeterminate*, and either (2) *settled* or *unsettled*. We can start by getting clear on distinction (1). According to the way in which Jackman uses this terminology, a term's extension is indeterminate if there is no fact of the matter about what the term's extension is (Jackman (1999), 168). Otherwise, the term's extension is determinate. 'Grant's zebra,' for example, would have an indeterminate extension if there were no fact of the matter as to whether the term picks out the whole of *Equus burchilli*, as opposed to some subset. Jackman, however, holds that the extension of 'Grant's zebra' is determinate, and the term picks out a particular race of *Equus burchilli* (Jackman (1999), 163–164).

Now we can turn to distinction (2), namely, the distinction between settled and unsettled content. This distinction is somewhat easier to understand with regard to truth-value. Jackman's example is the following hypothetical future-tensed utterance by Grant in 1820: "I will never be discussed in a philosophy paper" (Jackman (1999), 164). This utterance is false now, and it was false in 1820, since Jackman himself discusses Grant. But in 1820, it was not *settled* that the utterance is false—it was only settled once Jackman published his own paper in 1999.<sup>27</sup> Before 1999, then, Grant's utterance was unsettled false; since 1999, it has been settled false. More generally, Jackman uses this terminology in such a way that, "[r]oughly, an assertion is *settled true* at a moment if it is true in every history passing through that moment" (Jackman (1999), 163). He goes on to propose that "[i]n much the same way, a term's extension can be said to be 'settled' at a moment if the term has the same extension in every history passing through that moment" (Jackman (1999),

---

<sup>27</sup>Assuming, of course, the false claim that Jackman is the first philosopher to discuss Grant.

163). Going back to the Grant's zebra case, the extension of 'Grant's zebra' was unsettled in 1820. The events that made it the case that 'Grant's zebra' refers to Grant's zebra, and not to the whole of *Equus burchilli*, took place subsequent to 1820, and it is these events that settled the extension of 'Grant's zebra.'

Given these two distinctions, we have four possibilities when it comes to the content of an utterance (Jackman (1999), 172). First of all, the content of an utterance may be *unsettled determinate*, as utterances of 'Grant's zebra' were in 1820. On Jackman's understanding of the Grant's zebra case, 'Grant's zebra' always had a determinate extension, even in 1820. It refers to Grant's zebra, and not to the whole of *Equus burchilli*. The term has this extension because of events that took place subsequent to 1820, involving the study of zebras in Zimbabwe and around the Zambezi river. To be sure, it was a historical accident that these events took place, as opposed to the counterfactual ones that Jackman discusses. But they suffice to settle the extension of 'Grant's zebra.' However, since they had not yet taken place in 1820, the extension of the term in 1820, while determinate, was unsettled.

Secondly, the content of an utterance may be *settled determinate*. Today, the extension of 'Grant's zebra' is determinate, as it was in 1820. Moreover, the extension has been settled by events that took place in the past, subsequent to 1820.

Thirdly, the content of an utterance may be *unsettled indeterminate* (Jackman (1999), 172). Utterances of 'Grant's zebra' in 1820, while in fact unsettled determinate, may have been unsettled indeterminate. This would be the case if the linguistic practice that the taxonomists were engaged in were to die off, say, immediately after the term 'Grant's

zebra' was introduced. For example, one could imagine the world ending sometime in 1820. In that case, the subsequent history would not indicate a determinate extension for the term—it would always be indeterminate between Grant's zebra and the whole of *Equus burchilli* (Jackman (1999), 168). The actual events subsequent to 1820 made it the case that 'Grant's zebra' refers to the former, while the counterfactual events that Jackman discusses would have made it that case that 'Grant's zebra' refers to the latter. But if there is no usage subsequent to 1820, then nothing could settle the extension one way or the other. Hence, in addition to being indeterminate, the content would be unsettled.

Fourthly, and finally, Jackman admits cases in which utterances are *settled indeterminate*, or alternatively, "ontologically confused"; examples of "ontological confusion" include cases in which a speaker's "usage [is] unknowingly prompted by two independent sources of information ... [as] when, say, a speaker mistakenly uses a single name for a pair of twins, each of whom encounters the speaker equally often (though never together)" (Jackman (1999), 172).

These two distinctions can interact with one another in some subtle ways, and so it's worth being clear about how they differ from one another. It can be helpful to think of the determinate-indeterminate distinction as a *metaphysical* distinction that has to do with *what the extension of a term is*. A term has the extension that it does regardless of whether anyone knows that it does. The settled-unsettled distinction is also metaphysical, insofar as it has to do with *what makes it the case that a term has the extension that it does*. But this distinction can be used to address an *epistemological* issue. In particular, if the content of an utterance at some moment is settled, then at that moment, we can, in principle, know

what the extension of the term is, whether determinate or indeterminate. And if the content of an utterance at some moment is unsettled, then we cannot know what the extension of the term is. As Jackman puts it, “[c]laiming that the reference of a term is unsettled is . . . to admit that one does not have access to what is required to interpret it properly” (Jackman (1999), 170). To say that the reference of a term is settled is to say that we do have such access, at least in principle.

At this point, I hope to have illustrated the details of Jackman’s temporal externalism by appeal to the Grant’s zebra case. For the sake of this illustration, I have simply assumed that Jackman’s intuitions regarding the Grant’s zebra case are correct. But it’s questionable whether his intuitions about that case are widely shared, and I am not endorsing them here. Jackman seems to take it as given that future usage can determine the content of past utterances, and his intuition that the extension of ‘Grant’s zebra’ in 1820 was unsettled determinate accords with this position. But one who does not accept Jackman’s temporal externalism may have a different intuition, namely, that the extension of ‘Grant’s zebra’ in 1820 was indeterminate, and made determinate by subsequent usage.<sup>28</sup> Jackman does appeal to intuitions in order to argue for his position (Jackman (1999), 168), and if we do not share his intuitions, this is *prima facie* evidence against his position. Temporal externalism, then, is at the very least a controversial position, though I’ll assume for the sake of argument that Jackman’s intuitions are correct.

---

<sup>28</sup>Jackman admits that one may have that intuition, and give an account in terms of Field’s partial denotation in order to accommodate it. Although he gives some reasons for preferring temporal externalism to partial denotation, he holds that “there may be no way to give conclusive reasons for preferring either a [temporal] externalistic or a partial reference semantics over the other” (Jackman (1999), 170).

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

The realist response that I am now considering requires a commitment to temporal externalism. According to that response, in the actual history  $h_1$ , ‘muriatic radical’ failed to refer to an empirical entity at  $m_1$ . But it would have referred to an empirical entity at  $m_1$  if either  $h_2$  or  $h_3$  were the actual history. Without a commitment to temporal externalism, ‘muriatic radical’ at  $m_1$  either ‘kind of’ referred to an empirical entity, or failed to refer to an empirical entity, and this is the case whether we are talking about the actual history or the two counterfactual histories.

To begin with, we can see how temporal externalism provides a straightforward account of how the realist could go about arguing in favor of the response I am considering. On the actual history  $h_1$ , ‘muriatic radical’ does not refer to an empirical entity today. In the terminology of temporal externalism, its extension today is settled determinate. It has the determinate extension of the empty set. And this extension has been settled by events that occurred in the past, subsequent to  $m_1$ . In that case, at  $m_1$ , the extension of ‘muriatic radical’ was unsettled determinate. Because of events that occurred subsequent to  $m_1$ , it had the determinate extension of the empty set. But this determinate extension was not settled until some time after  $m_1$ . On the counterfactual histories  $h_2$  and  $h_3$ , today, ‘muriatic radical’ would have referred to hydrogen and chlorine, respectively. In both cases, its extension would be settled determinate. In  $h_2$ , it would have the determinate extension of hydrogen, and in  $h_3$  it would have the determinate extension of chlorine. And these extensions would be settled by events that took place in the past, subsequent to  $m_1$ . In that case, at  $m_1$ , the extension of ‘muriatic radical’ would have been unsettled determinate. In each of the two histories, ‘muriatic radical’ has the same extension that it would have

today—hydrogen in  $h_2$  and chlorine in  $h_3$ . But since events subsequent to  $m_1$  settle these extensions, at  $m_1$ , the extensions would be unsettled.

It's worth being clear that temporal externalists are not committed to the claim that 'muriatic radical' at  $m_1$  had three different determinate extensions—the empty set, hydrogen, and chlorine. Instead, they claim that at  $m_1$ , the extension was, in fact, the empty set. On the assumption that  $h_2$  is the actual history, the extension at  $m_1$  would have been hydrogen. And on the assumption that  $h_3$  is the actual history, the extension at  $m_1$  would have been chlorine. Temporal externalism requires us to specify, not just a moment, but also a history, in order to evaluate the content of an utterance. In this case, the content of 'muriatic radical' at  $m_1$  is undefined until we specify which history is actual. Once we do so, 'muriatic radical' at  $m_1$  will have exactly one extension. According to the temporal externalist, this is the way usage subsequent to  $m_1$  can affect the content of utterances at  $m_1$ .

I will now examine what the realist is left with if temporal externalism is not an option. The realist would be committed to the claim that 'muriatic radical' at  $m_1$  has the extension that it does regardless of whether  $h_1$ ,  $h_2$ , or  $h_3$  is the actual history. In this case, at  $m_1$ , 'muriatic radical' would have the same extension in all three histories, and one needn't specify a history in order to evaluate the content of an utterance. This is a consequence of denying that future usage can affect past content. If the realist takes this option, she would owe us a plausible account of the extension of 'muriatic radical' at  $m_1$ . Moreover, the realist would owe us some coherent story that explains the transition from the term's extension at  $m_1$  to its actual determinate extension today on  $h_1$  (the empty set), and to its counterfactual

determinate extensions today on  $h_2$  and  $h_3$  (hydrogen and chlorine, respectively).

If the realist rejects temporal externalism, then ‘muriatic radical’ did not successfully refer to either hydrogen or chlorine at  $m_1$ . Suppose it had the determinate extension of hydrogen. The realist would then owe us a story of why it did not refer to chlorine, or to no empirical entity at all. Moreover, the realist would have to tell us what happened, in the actual history, to make it the case that today the term altogether fails to refer to an empirical entity. Analogous points can be made regarding the possibility that, at  $m_1$ , it had the determinate extension of chlorine. I take it that any story that the realist could tell in these cases would be quite implausible. So, given that hydrogen and chlorine are the only options for empirical entities that ‘muriatic radical’ could have successfully referred to at  $m_1$ , I conclude that, without temporal externalism, successful reference is not an option, not even if it is counterfactual successful reference.

Some claims that some realists make show that they are, in fact, committed to temporal externalism. In particular, Putnam’s use of the Principle of Charity (or, alternatively, Principle of Benefit of Doubt), in order to make “retrospective reference assignments” from the standpoint of our current theories, seems sufficient to commit him to temporal externalism (Putnam (1978), 22). To begin with, Putnam has this to say about the principle:

when speakers specify a referent for a term they use by *description* and, because of mistaken factual beliefs that these speakers have, that description fails to refer, we should assume that they would accept reasonable reformulations of their description (in cases where it is clear, given our knowledge, how their description should be reformulated so as to refer, and there is no ambiguity

about how to do it in the practical context). (Putnam (1978), 23–24)

Putnam applies this principle to Newton, Dalton, and Mendel in the following passage:

Yet it is a fact that we can assign a referent to ‘gravitational field’ in Newtonian theory *from the standpoint of* relativity theory (though not to ‘ether’ or ‘phlogiston’); a referent to Mendel’s ‘gene’ from the standpoint of present-day molecular biology; and a referent to Dalton’s ‘atom’ from the standpoint of quantum mechanics. These retrospective reference assignments depend on a principle that has been called the ‘principle of benefit of the doubt’ or the ‘principle of charity’, but not on *unreasonable* ‘charity’. Surely the ‘gene’ discussed in molecular biology is the gene (or rather ‘factor’) Mendel *intended* to talk about; it is certainly what he should have intended to talk about! (Putnam (1978), 22)

Such reference assignments, if they are indeed retrospective, would make it the case that usage subsequent to the utterances of Newton, Dalton, and Mendel affects the content of those utterances. And this is enough to commit Putnam to temporal externalism. Now it’s possible that the reason that these past scientists would accept reformulations of their descriptions is that they are causally connected to the referents that we would assign to their terms from the standpoint of our current theories. But if these scientists were, in fact, willing to let twentieth-century scientists affect the content of their pre-twentieth-century utterances, then these past scientists are just as committed to temporal externalism as Putnam seems to be. Perhaps this link between the principle and temporal externalism

is to be expected, given that Jackman views temporal externalism as a natural development of externalist views like Putnam's (Jackman (1999), 157–158).

Even if temporal externalism is a natural choice for some realists, it is also possible for a realist to reject this position. Such a realist is therefore left with two options. First of all, she could claim that 'muriatic radical,' at  $m_1$ , 'kind of' referred to an empirical entity. Secondly, she could claim that 'muriatic radical,' at  $m_1$ , failed altogether to refer to an empirical entity. I will consider both of these options in turn.

#### 4.4.2 'MURIATIC RADICAL' AS A TERM THAT 'KIND OF' REFERS

How would a realist respond to the question: At  $m_1$ , to what did the HE term 'muriatic radical' refer? The second response that I want to consider is that it 'kind of' referred to an empirical entity at  $m_1$ . I have already argued that today, 'muriatic radical' does not 'kind of' refer. In  $h_1$ , it fails to refer today. And if  $h_2$  or  $h_3$  were actual, then today, it would refer to hydrogen or chlorine, respectively. But strictly speaking, this is not inconsistent with the claim that, at  $m_1$ , 'muriatic radical' 'kind of' referred to an empirical entity. In order to argue for such a claim, the realist needs to give an account of what it 'kind of' referred to, in such a way that the transitions that take place in the actual and counterfactual histories are rendered intelligible. More specifically, the realist would owe us an account of the transition from 'kind of' reference to reference failure in  $h_1$ , and accounts of the transition from 'kind of' reference to successful reference in  $h_2$  and  $h_3$ .

I've discussed two forms of 'kind of' reference, namely, Psillos' approximate reference, and Field's partial denotation. I've already argued that the latter is the more developed

view. Furthermore, I've argued that any view of 'kind of' reference, as applied to 'muriatic radical' today, must mirror the partial denotation account, in the sense that it must take the form that 'muriatic radical' 'kind of' referred to both hydrogen and chlorine. I take it that this also shows that at  $m_1$ , if 'muriatic radical' 'kind of' referred to an empirical entity, it must have 'kind of' referred to both hydrogen and chlorine. In that case, in the course of developing this second realist response, I will focus exclusively on Field's partial denotation, with the understanding that if it fails to yield an adequate account of the semantics of 'muriatic radical' at  $m_1$ , then all accounts of 'kind of' reference must fail for similar reasons.

Although it may be the case that one can simultaneously make use of both temporal externalism and partial denotation, I will assume here that these are rival accounts. Jackman holds that this is the case (Jackman (1999), 170–172). Moreover, temporal externalism may undercut the need for partial denotation. For example, in the Grant's zebra case, the temporal externalist can offer an account that gives 'Grant's zebra' a determinate extension in 1820. On the other hand, a partial denotation account will imply that the term's extension in 1820 was indeterminate between Grant's zebra and the whole of *Equus burchilli*. But if partial denotation is a tool that we can use to make sense of indeterminacy, then it is a tool that the temporal externalist will not need to make use of in cases like the Grant's zebra case. That said, if a proponent of partial denotation accepts temporal externalism, then my objections to temporal externalism will apply to her as well. So, for the sake of argument, I will make the simplifying assumption that a proponent of partial denotation will reject temporal externalism.

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

At this point, we can see how a realist could make use of partial denotation in order to answer the above question. I've already argued that if one rejects temporal externalism, then at  $m_1$ , 'muriatic radical' would have the same extension regardless of which history is actual. In this case, the realist can say that, at  $m_1$ , 'muriatic radical' partially denoted both hydrogen and chlorine.

As for the transitions to the present day, we can begin with the counterfactual histories. In  $h_2$ , chemists came to use this term to fully denote hydrogen at some time subsequent to  $m_1$ . And in  $h_3$ , chemists came to use this term to fully denote chlorine at some time subsequent to  $m_1$ . In both cases, then, we have a transition from partial to full denotation. These cases would be analogous to Boyd's 'element' case, according to which 'element' once partially denoted both elements and isotopes, whereas now it fully denotes elements.

Finally, coming back to the actual history  $h_1$ , the proponent of partial denotation may claim that 'muriatic radical' eventually fully denoted the empty set because chemists, in fact, did not make either of the decisions that they made in  $h_2$  and  $h_3$ . Now it may be objected that a term that previously partially denoted two extensions cannot fail altogether to refer at a later date. But it's not clear that this is the case. One possible view of the term 'gene' is that in the past, it partially denoted cistrans, mutons, and recons, but today fails to refer altogether.<sup>29</sup> In any case, if this objection goes through, then partial denotation cannot be used to make sense of the muriatic radical case, and I needn't consider it any further. So, for the sake of argument, I will assume that the objection can be answered. That is to say, I

---

<sup>29</sup>This would be to combine Field's suggestion that the term partially denoted three extensions in the past (Field (1973), 477), with the suggestion of Clyde L. Hardin and Alexander Rosenberg that 'gene' fails to refer today (Hardin and Rosenberg (1982), 607).

will assume that a realist can use partial denotation to give a plausible-sounding account of the extension of ‘muriatic radical’ at  $m_1$ , and of the transitions that take place subsequent to  $m_1$  in the actual and counterfactual histories.

#### 4.4.3 WHIG HISTORY

At this point, I will argue that both temporal externalism and partial denotation force one to give a whig history of the muriatic radical. In this case, the realist should abandon the claim that ‘muriatic radical’ at  $m_1$  successfully referred to an empirical entity, as well as the claim that it ‘kind of’ referred to an empirical entity.

It will first be necessary to say a word about what whig history is. As Herbert Butterfield puts the matter, “[i]t is part and parcel of the whig interpretation of history that it studies the past with reference to the present” (Butterfield (1965), 11). But there is more to the whig interpretation of history than this. C. B. Wilde provides a useful characterization of whig history. Speaking of the actual Whig historians, Wilde accuses them of “devoting attention to seemingly modern ideas and movements regardless of their importance in their own time, refusing historical understanding to all opposing tendencies,” and “mak[ing] the present the absolute judge of past controversies and the sole criterion for the selection of episodes of historical importance” (Wilde (1981), 445). Whig history eventually took hold in the history of science, at least in part because historians of science before the 1950s were sufficiently impressed with the progress of science. They used currently accepted scientific theories in order to tell a history that involved heroes (those who were right according to those current theories) and villains (those who were wrong). Wilde goes on to list a

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

number of reasons for the demise of whig history of science (Wilde (1981), 445–446). One reason is that historians eventually came to see the superiority of approaches that seek to reconstruct the problem situations that scientists actually faced in their historical context, as opposed to looking at the past through the lens of our best theories. Another reason involves historical work which has shown that ideas that have since been rejected were actually of great importance for focusing the attention of scientists on various phenomena and the problems that they pose. Even today, whig history is seen as a pejorative term for these and other reasons.

The quotation from Butterfield captures the sense in which temporal externalism is whiggish—it studies the past with reference to the present. More specifically, the semantic properties of terms and utterances as used in the past are inextricably linked to future usage, and, indeed, cannot be fully understood without reference to future usage. An essential part of the history of science obviously involves understanding the meanings of past speakers' terms and utterances. Anyone who wants to make use of temporal externalism in the course of doing history of science will thereby end up doing whig history of science.

While I take it that any application of temporal externalism to the history of science will yield a whig history, I admit the possibility that partial denotation can function as part of a non-whig history. That said, its application to the case of the muriatic radical is bound to be whiggish. If we look at Lavoisier's *Elements* from the standpoint of our current knowledge of acids, then hydrogen and chlorine are the only possible empirical referents for the term 'muriatic radical' at  $m_1$ . If 'muriatic radical' partially denoted both hydrogen and chlorine at  $m_1$ , then one of its partial extensions would include a substance that Lavoisier simply

would have denied talking about at all, namely, chlorine. According to Lavoisier, chlorine, which he called ‘oxygenated muriatic acid,’ contains muriatic acid, and therefore can’t possibly be one of its constituents. Partial denotation therefore fails to give a picture of Lavoisier’s work on acidity that is true to the historical context in which he worked, and yields a whig history of the muriatic radical.

At this point, I’ve argued that the successful reference of ‘muriatic radical’ at  $m_1$  requires a commitment to temporal externalism, and that the ‘kind of’ reference of ‘muriatic radical’ at  $m_1$  requires the use of partial denotation. I’ve argued that in general, temporal externalism leads to whig history, and that a partial denotation account of ‘muriatic radical’ in particular leads to whig history. Since whig history of science has been much maligned in the literature already, this may already be reason enough to reject temporal externalism in general, along with any position that implies that ‘muriatic radical’ successfully referred at  $m_1$ . And it may already be reason enough to reject the application of partial denotation to the case of the muriatic radical, and any account of the ‘kind of’ reference of ‘muriatic radical.’

However, it may be possible to make a case for whig history. As I discussed in chapter 1, two of the arguments that proponents of the extant views address are the no-miracles argument and the pessimistic induction. According to the former, the explanatory and predictive success of our best scientific theories would be miraculous if those theories weren’t at least approximately true. According to the latter, we have no reason to think that those theories will avoid the fate of their predecessors, and eventually be rejected as being not even approximately true. The challenge for the extant views is to give an account of what

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

is and is not preserved over theory change, so as to account for the intuitions that drive these two seemingly opposed arguments.<sup>30</sup> The extant views face this challenge in various ways. Realists claim that predecessor theories will approximate their successors in such a way that the former will still look approximately true in light of the latter.<sup>31</sup> Proponents of ESR claim that structure, but not nature, is preserved over theory change.<sup>32</sup> Proponents of OSR claim that structure is all there is, and that it is preserved over theory change.<sup>33</sup> And constructive empiricists claim that empirical content is all that is preserved over theory change.<sup>34</sup> These ways of meeting the challenge all imply the seemingly whiggish claim that, to some extent, the history of science is an inevitable march toward progress. And since the extant views owe us an account of what is and is not preserved over theory change, it would seem that some element of whiggishness is necessary to reconcile these two seemingly opposed arguments.

But there are two responses to this argument that purports to show the benefits of whig history. First of all, not every view of the progress of science needs to be whiggish. It seems undeniable that science has made progress on the empirical level. The sciences have been able to accommodate more and more phenomena over time, and they have been able to give more and more accurate descriptions of those phenomena. To be sure, some philosophers have attempted to defend various examples of so-called ‘Kuhn loss,’ according to which a phenomenon that was explained by a predecessor theory fails to be explained by its successor. But, as Worrall has argued, such examples are either highly theoretical or highly

---

<sup>30</sup>Worrall (1989) poses this as the central challenge in the realism debate.

<sup>31</sup>See the discussion of realism in Laudan (1981).

<sup>32</sup>See Worrall (1989), 117.

<sup>33</sup>See Ladyman and Ross (2007), 157.

<sup>34</sup>See van Fraassen (2001), 163 and van Fraassen (2008), 110–111.

vague (Worrall (1989), 108). The consensus, then, seems to be that science does display empirical progress. Moreover, one needn't adopt the standpoint of our current theories in order to characterize progress on the empirical level, since such progress is empirical, as opposed to theoretical. A commitment to progress on the empirical level, then, does not imply a commitment to whig history.

Secondly, there is a problem even assuming that one can make a good case for whig history. This is because the commitment to temporal externalism and/or partial denotation *forces* one to give a whig history of the muriatic radical—no other kind of history is even an option. Presumably, it would be desirable to allow some flexibility here, especially when it comes to adopting a kind of history as controversial as whig history. But if my argument is on point, neither temporal externalism nor partial denotation allow any such flexibility in the case of the muriatic radical. This gives us reason to reject the first two realist responses, namely, that 'muriatic radical' at  $m_1$  either successfully referred or 'kind of' referred. In that case, we are left with one option, namely, that 'muriatic radical' at  $m_1$  failed to refer.

#### 4.4.4 'MURIATIC RADICAL' AS A TERM THAT FAILS TO REFER

How would a realist respond to the question: At  $m_1$ , to what did the HE term 'muriatic radical' refer? The third response that I want to consider is that it failed altogether to refer to an empirical entity. More specifically, at  $m_1$ , 'muriatic radical' failed to refer in the actual history  $h_1$ , and it would have failed to refer if either  $h_2$  or  $h_3$  were the actual history. If successful reference and 'kind of' reference commit the realist to giving a whig history of the muriatic radical, then reference failure is the realist's last hope. Indeed, I agree with

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

a realist who pursues this option, insofar as I agree that, at  $m_1$ , ‘muriatic radical’ did not refer to an empirical entity in any of the histories that I’ve considered.

A realist who pursues this response can give the following account of the transitions that take place in the actual and counterfactual histories. Starting with the counterfactual histories, she can explain the transition from reference failure at  $m_1$  to successful reference today in terms of the decisions that Davy and other chemists might have made. In  $h_2$ , Davy identified the muriatic radical with hydrogen, and chemists followed him in this decision. And in  $h_3$ , Davy identified the muriatic radical with chlorine, and once again chemists followed him in this decision. To use the language of the causal theory of reference, the realist might explain the transition from reference failure to successful reference in terms of a dubbing ceremony, in which Davy baptized hydrogen and chlorine, respectively.<sup>35</sup> Coming back to the actual history, she can explain why ‘muriatic radical’ never referred by appeal to the fact that such dubbings, if they occurred at all, never caught on in the community of chemists.

I think that this explanation is good, as far as it goes. But I want to argue that it doesn’t go far enough. In particular, it leaves some important questions unaddressed. In particular, in  $h_1$ , why did the community of chemists collectively decide not to use the term ‘muriatic radical’ to refer to an empirical entity? In  $h_2$ , why did Davy decide to use the ‘muriatic radical’ to refer to hydrogen? And in  $h_3$ , why did he decide to use it to refer to chlorine? The realist can give the details of the actual and counterfactual histories as I have presented them, and can tell us *that* the term ‘muriatic radical’ came to be a successfully referring

---

<sup>35</sup>See chapter 1 for some discussion of this terminology.

term in the counterfactual histories, but not in the actual history. But for the realist, there is nothing in the semantics of the HE term ‘muriatic radical’ other than the fact that it purports to refer to an empirical entity, and, in fact, fails to refer. In that case, there’s nothing in the semantics of the term that the realist can point to in order to explain *why* the term ‘muriatic radical’ came to be a successfully referring term in the counterfactual histories, but not in the actual history. I take it that a position that can answer such why-questions by appealing to the semantics of the terms in question will fare better than a position that cannot do this. Since the realist view of hypothetical entities falls in the latter category, I conclude that realists are able to give, at most, an incomplete history of the muriatic radical if they opt for the response that, at  $m_1$ , ‘muriatic radical’ failed to refer to an empirical entity in all three histories.

Now it may be objected that I am assuming that the realist is committed to a direct reference account of the semantics of HE terms, according to which the semantic content of a term is exhausted by its referent, and that this assumption is unfair to the realist.<sup>36</sup> I agree that realists needn’t adopt such a view of the semantics of HE terms. But it’s not clear that a view that admits senses will fare any better. Senses can be quite minimal. In order to explain the transition that takes place in  $h_3$  by appeal to the sense of the HE term ‘muriatic radical,’ the realist would need to show that the sense is sufficiently rich to include such things as the fact that Lavoisier listed the radical in his table of simple substances. And while a realist may opt for a semantic position that trades in such sufficiently rich senses, she would have to do some work to defend the combination of realism and this semantic

---

<sup>36</sup>See, for example, Salmon (1986) and Soames (1987).

position. Moreover, she would have to position herself against much of what has gone on in the development of realism, which often makes heavy use of the causal theory of reference so as to dispense with the need for such rich senses. And if senses are sufficiently rich, the realist may be left with the consequence that some terms, which she took to be referring (like Dalton's 'atom' and Mendel's 'gene'), in fact, fail to refer. The burden of proof, then, would be on the realist to argue for such a view of senses.

Until the realist is able to show otherwise, then, I conclude that the realist view of hypothetical entities is inadequate when it comes to giving a history of the muriatic radical, since realists face an unpalatable choice between whig history and incomplete history.

#### 4.4.5 THE OTHER EXTANT VIEWS REVISITED

And as we've seen in chapter 1, the upshot of my survey of the scientific realism debate was that the extant views share an important similarity, namely, that HE terms are putatively referring expressions that have putative reference to empirical entities. In one way or another, proponents of realism, constructive empiricism, ESR, and OSR all hold this idea. In that case, my arguments concerning realism apply rather straightforwardly to the other extant views as well. But since there are some differences between the realist view and the other extant views, some brief remarks are in order.

To begin with, ESR differs from realism only in the sense that it admits one possibility where realism admits three. More specifically, according to the proponent of ESR, the term 'muriatic radical' referred to an empirical entity at  $m_1$ . I've already argued that, at most, we can say that 'muriatic radical' referred counterfactually to an empirical entity at  $m_1$ .

So if proponents of ESR are committed to the claim that it actually referred, I take it as a strike against their position. Moreover, even the counterfactual successful reference of ‘muriatic radical’ at  $m_1$  depends on temporal externalism, in which case proponents of ESR are committed to giving a whig history of the muriatic radical. Hence, my argument against the realist view of hypothetical entities applies to ESR as well.

OSR differs from realism only in the sense that proponents of OSR replace talk of empirical entities and terms with talk of real patterns and locators. But so long as there is a way for proponents of OSR to map the former onto the later, there is a way to apply my arguments concerning realism to OSR. Proponents of the latter face the same unpalatable choice between whig history and incomplete history.

Finally, constructive empiricism differs from realism only in the sense that it admits the possibility of reference to mathematical entities. But in the case of the muriatic radical, constructive empiricism and realism are actually in complete agreement. This is because we can eliminate the possibility that ‘muriatic radical’ refers to a mathematical entity. As I discussed in chapter 1, on the constructive empiricist account of HE terms, the possibility of referring to a mathematical entity arises only in the case of terms that name unobservable entities. For the sake of argument, we can assume that ‘muriatic radical’ is such a term.<sup>37</sup>

Van Fraassen raises the possibility that a statement asserting the existence of forces may simultaneously be literally construed, and be construed as asserting the existence of certain functions (van Fraassen (1980), 11). It’s clear that van Fraassen has in mind theories that

---

<sup>37</sup>Alan Chalmers has claimed that the gases that chemists in the late eighteenth and early nineteenth centuries examined are examples of unobservable entities (Chalmers (2009), 4; Chalmers (2011), 712). One might certainly question this claim. But if Chalmers’ examples of oxygen, nitrogen, and hydrogen were candidates for the muriatic radical, then Chalmers’ claim lends some support to the view that ‘muriatic radical’ is a term that names an unobservable entity.

can be mathematized to a large extent, and Lavoisier's oxygen theory of acidity is not a theory of this kind. Mathematics does play some role in Lavoisier's theory, as when he concerns himself with weight relations before and after a reaction has taken place. But this is a long way off from the theories of mathematical physics that van Fraassen has in mind, and there seems to be no sense in which 'muriatic radical' can simultaneously be literally construed, and construed as referring to a mathematical entity. In that case, if 'muriatic radical' has a referent or partial denotation at  $m_1$ , it refers to or partially denotes an empirical entity. And so, once again, constructive empiricists face the same unpalatable choice between whig history and incomplete history.

#### 4.5 SUPPOSITIONALISM AND THE MURIATIC RADICAL

As we've seen in chapter 2, the extant views of hypothetical entities and HE terms are not compulsory, since there exists another view, namely, suppositionalism. Suppositionalists are committed to  $S_{HE}$ , according to which HE terms refer to hypothetical entities, and hypothetical entities are objects of supposition. Moreover, they accept the following three claims:

- (S<sub>1</sub>) Theoretical terms are either putatively referring expressions that have putative reference to empirical entities, or else they refer to objects of supposition.
- (S<sub>2</sub>) Representations of suppositions are theories capable of being vindicated or discredited, but they are not best understood as candidates for being true or false *simpliciter*.

(S<sub>3</sub>) Homonymous theories contain truth-evaluable sentences that are best understood as capable of being true or false *simpliciter*.

How, then, would a suppositionalist answer the question: At  $m_1$ , to what did the HE term ‘muriatic radical’ refer? Since ‘muriatic radical’ was an HE term at  $m_1$ , the suppositionalist would answer that at  $m_1$ , ‘muriatic radical’ referred to a hypothetical entity, understood as an object of supposition.

My goal in this section is to defend this answer, and show that it is superior to the answers that the extant views are committed to. I’ll begin by explaining what it means for the muriatic radical to be an object of supposition. I will then apply the appropriation model to the actual and counterfactual histories. Finally, I will argue that suppositionalism escapes the charge of whiggishness, and the charge of giving an incomplete history, in which case it is preferable to the extant views.

#### 4.5.1 THE MURIATIC RADICAL AS AN OBJECT OF SUPPOSITION

According to the suppositionalist, the HE term ‘muriatic radical’ refers to a hypothetical entity, understood as an object of supposition. So understood, the muriatic radical has both conformal and normal properties. If we restrict ourselves to what Lavoisier tells us in his *Elements*, we know that it has at least the conformal properties of:

- being an acidifiable base or radical,
- not having been isolated yet,
- combining with oxygen to form muriatic acid,

- combining with even more oxygen to form oxygenated muriatic acid, and
- being a simple substance.

It also has a number of normal properties, including the properties of:

- playing a role in Lavoisier's theory of acidity, and
- appearing in Lavoisier's table of simple substances.

At  $m_1$ , the reference of 'muriatic radical' was quite determinate for the suppositionalist. The term referred to the object of supposition that has all and only the conformal properties that the muriatic radical has according to the supposition generated by Lavoisier's *Elements*.

This claim needs some clarification and some defense. To begin with, this way of understanding the reference of 'muriatic radical' may be very different from the way in which Lavoisier himself intended. But the semantics that suppositionalism is committed to is independent of the intentions of scientists, and suppositionalism does not entail anything about what Lavoisier intended. Instead, it amounts to the claim that using speculative works (like the *Elements*) as props to generate suppositions is the best way in which to understand those works. So, when the suppositionalist claims that 'muriatic radical' refers to an object of supposition, this is to say that it refers to an object which comes to be as a result of such a supposition.<sup>38</sup>

One might object to suppositionalism on the grounds that it is susceptible to the same objection that I raised against the extant views. I've argued that the successful reference

---

<sup>38</sup>See the end of chapter 2 for a fuller defense of suppositionalism against the charge that it is flawed because it ignores the intentions of scientists.

## CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

of ‘muriatic radical’ at  $m_1$  requires a commitment to temporal externalism, and that the latter forces us to engage in whig history. But suppositionalism is not susceptible to this objection. The extant views face the objection only because all of them are, in some way, committed to the claim that HE terms are putatively referring expressions that have putative reference to empirical entities. The suppositionalist escapes the dilemma by treating HE terms as terms that refer to objects of supposition, as opposed to terms that putatively refer to empirical entities. If ‘muriatic radical’ at  $m_1$  is understood as a term that refers to an object of supposition, as opposed to a putatively referring expression, then the reference of that term can be determinate even if one rejects temporal externalism.

### 4.5.2 THE ACTUAL AND COUNTERFACTUAL HISTORIES AND THE APPROPRIATION MODEL

My task at this point is to explain how the appropriation model can account for the actual and counterfactual transitions from the extension of ‘muriatic radical’ at  $m_1$  to its extension today. The suppositionalist claims that at  $m_1$ , the HE term ‘muriatic radical’ referred to an object of supposition. But ‘muriatic radical’ is no longer an HE term today. In the actual history, it fails to refer to an empirical entity today. In that case, it must have failed to be appropriated to refer to an empirical entity after  $m_1$ , and so we say that Lavoisier’s theory of acidity was discredited. And in the counterfactual histories, it would refer to an empirical entity today. In that case, it would have been appropriated to refer to an empirical entity sometime after  $m_1$ , and so we would say that Lavoisier’s theory was vindicated.<sup>39</sup>

---

<sup>39</sup>As I mentioned in chapter 2, vindication and discreditation come in degrees, and the same theory may be both vindicated and discredited if it has an HE term that has been appropriated and an HE term that has failed to be appropriated. When I speak of Lavoisier’s theory being vindicated and discredited in what follows, I

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

We can explain these transitions in terms of representations of suppositions and their homonymous theories. The representation of a supposition is a set of sentences that are true according to the supposition that one generates by using some work as a prop. The representation of a supposition will therefore have terms that name objects of supposition (including HE terms), and we can obtain the homonymous theory of such a representation by replacing, as much as possible, terms that name objects of supposition with terms that have putative reference to empirical entities.<sup>40</sup>

For simplicity's sake, we can restrict ourselves to the representation of a supposition generated by using Lavoisier's *Elements* as a prop. The first representation,  $RS_1$ , treats 'muriatic radical' as an HE term that has neither succeeded nor failed to be appropriated. Its homonymous theory is  $HT_1$ . After Lavoisier's theory was either vindicated or discredited, we have another representation of a supposition generated by using Lavoisier's *Elements* as a prop—call it  $RS_2$ . The homonymous theory of  $RS_2$  is  $HT_2$ , and we obtain the latter by replacing the HE term 'muriatic radical' with a homonymous term that has putative reference to an empirical entity. We can visualize this in terms of the figure from chapter 2, which I reproduce here for the reader's convenience. Once again, the single arrows indicate the mapping that takes one from a representation of a supposition to a homonymous theory by means of replacing various terms, as I described in chapter 2. In the actual history, we must compare the representations, and their homonymous theories, before and after Lavoisier's theory was discredited. And in the counterfactual histories, we must compare the representations, and their homonymous theories, before and after Lavoisier's theory

---

am restricting my attention to the term 'muriatic radical.'

<sup>40</sup>See my discussion in chapter 2 for a more detailed explanation.

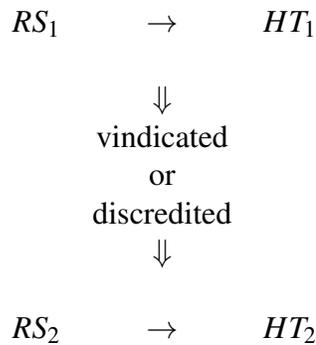


Figure 4.2: Vindicated and discredited theories, and their relation to the world.

was vindicated.

We can now see how the appropriation model can account for the transitions that take place in these histories, starting with the counterfactual ones. In the first counterfactual history  $h_2$ , Davy was so impressed with the analogies between oxygen and chlorine that he associated hydrogen with the muriatic radical. In the second counterfactual history  $h_3$ , Davy associated the muriatic radical with chlorine, on the grounds that the most straightforward way of revising Lavoisier's table of simple substances would involve replacing the former with the latter. Moreover, in both histories, the community of chemists at the time recognized this association. In each history, the match between the conformal properties of the muriatic radical and the normal properties of an empirical entity was sufficient to give Davy and the rest of the community cause to associate the muriatic radical with an empirical entity. We can say that the HE term 'muriatic radical<sub>1</sub>' referred to an object of supposition at  $m_1$ . And two homonymous terms, 'muriatic radical<sub>2</sub>' in  $h_2$  and 'muriatic radical<sub>3</sub>' in  $h_3$ , were introduced subsequent to  $m_1$ . These terms, neither of which is an HE term, have putative reference to empirical entities, and refer to hydrogen and to chlorine, respectively. Hence, in each of these histories, the term 'muriatic radical' was appropriated,

and Lavoisier's theory was vindicated. And so, in each history,  $HT_2$  is more truth-evaluable than  $HT_1$ . This is of interest even if Lavoisier's theory of acidity was no longer in use by the time these counterfactual vindications were complete. This is because after vindication, we can evaluate the truth and falsity of certain statements involving the term 'muriatic radical,' regardless of whether the theory is being used, whereas before, these statements were not best understood as candidates for truth and falsity.

Now we can turn to the actual history  $h_1$ . One of the upshots of my discussion of the counterfactual histories was that the term 'muriatic radical' fails to refer to an empirical entity today, at least in part because the events in the counterfactual histories were, after all, counterfactual. In actuality, the matches between the conformal properties of the muriatic radical, *qua* object of supposition, and the normal properties of any empirical entity, including hydrogen and chlorine, were not sufficient to give Davy cause to associate the muriatic radical with an empirical entity. Nor were they sufficient for the community of chemists as a whole. So once again, we have the HE term 'muriatic radical<sub>1</sub>,' and a homonymous term 'muriatic radical<sub>4</sub>,' which was introduced subsequent to  $m_1$  in  $h_1$ . This latter term is not an HE term—it has putative reference to an empirical entity, and, in fact, fails to refer to an empirical entity. Hence, in the actual history, appropriation failed, and Lavoisier's theory was discredited. Once again,  $HT_2$  is more truth-evaluable than  $HT_1$ . This is of interest even if Lavoisier's theory of acidity was no longer in use by the time it had been discredited. This is because after the theory had been discredited, we can evaluate the truth and falsity of certain statements involving the term 'muriatic radical,' regardless of whether the theory is being used, whereas before, these statements were not best understood as candidates for

truth and falsity.

Moreover, in the actual history, today we have a dead HE term, namely, ‘muriatic radical<sub>5</sub>.’ This term refers to a dead hypothetical entity, namely, object of supposition that has the conformal properties that the muriatic radical has according to Lavoisier’s *Elements*. This dead HE term, however, is no longer a term that can be appropriated to refer to an empirical entity, and hence, it differs from the HE term ‘muriatic radical<sub>1</sub>.’ Since the dead HE term is a term that refers to an object of supposition, it fails to refer to an empirical entity. And this is consistent with what I argued earlier in the chapter, namely, that as a matter of fact, ‘muriatic radical’ does not refer to an empirical entity today.

It’s worth pointing out that in my discussion of the actual and counterfactual histories earlier in the chapter, I spoke of the *identification* of the muriatic radical with an empirical entity. But in applying the appropriation model, I have spoken of the *association* of the muriatic radical with an empirical entity. This is because, strictly speaking, a hypothetical entity, *qua* object of supposition, cannot be identical to an empirical entity. So a suppositionalist will construe such identification claims non-literally, in terms of association.

Before moving on, a remark about vindicating and discrediting theories is in order. It is difficult to point to a specific time at which a theory has been vindicated or discredited, since the appropriation of an HE term, or its failure to be appropriated, must have the backing of the scientific community. We can begin by considering the failure of appropriation in the actual history. Davy’s work on muriatic acid culminated in his 1810 Bakerian Lecture. But it’s difficult to point to 1810 as the time when Lavoisier’s theory was discredited. In the decade that followed, Jöns Jacob Berzelius and John Murray continued to employ the

muriatic radical, so as to defend some version of Lavoisier's theory of acidity. And as late as 1858, Christian Friedrich Schönbein held that chlorine contains murium, a hypothetical entity that bears a striking resemblance to the muriatic radical.<sup>41</sup> Lavoisier's theory had surely been discredited by the late nineteenth century. But it's difficult to point to a specific time at which this happened.

Similar remarks apply to the successful appropriations in the counterfactual histories. Attempted appropriations by individual scientists can be dated in a relatively unproblematic way, by pointing to specific passages that contain those attempted appropriations. But successful appropriation requires, not just the actions of a single scientist like Davy, but the recognition of a scientific community. And it can be difficult to point to a specific time at which such recognition is sufficient for a successful appropriation, in which case dating the vindication of a theory is just as difficult as dating its discreditation. Kuhn argued that scientific discovery is a temporally-extended process (Kuhn (1962); Kuhn (1996), §VI). If Kuhn is correct, then insofar as the appropriation model is a model for understanding the 'discovery' of hypothetical entities, we might think that appropriation must be temporally extended as well. In  $h_2$ , we can speak of the discovery that the muriatic radical is hydrogen, and in  $h_3$ , we can speak of the discovery that it is oxymuriatic acid, or alternatively, chlorine. But in neither case were these discoveries instantaneous, since claims that the muriatic radical is something-or-other depend on the successful appropriation of the term 'muriatic radical,' which in turn depends on the backing of the scientific community. In this case, we have good reason to believe that the processes by which theories are vindicated

---

<sup>41</sup>See my discussion at the end of chapter 3 for more details.

and discredited are temporally extended.

Given this, the semantics of theoretical terms can be quite messy while the process is going on. But suppositionalism is equipped to handle this messiness. In the case of the muriatic radical, we can talk about distinct objects of supposition. For example, Lavoisier's muriatic radical is, strictly speaking, distinct from Berzelius', since they differ in conformational properties. Muriatic acid contains hydrogen for Berzelius, but not for Lavoisier. And Berzelius may be best understood as working with an object of supposition, and a corresponding HE term, even if chemists like Davy were already working with a homonymous term that was not an HE term, but instead a term that putatively refers to an empirical entity or a dead HE term. More generally, the individuals that make up the scientific community may have been working with distinct terms and distinct entities before the processes of vindicating and discrediting theories began. These processes might therefore be seen as a kind of clean-up job. The recognition of the community as a whole amounts to the collective adoption of the same terms, which putatively refer to empirical entities as opposed to objects of supposition. But it takes time for this collective adoption to take effect, and the details of such a process can only be understood by closely examining the work of the scientists involved. Suppositionalism provides a framework for engaging with such work, as messy as it can be.

#### 4.5.3 SUPPOSITIONALISM, WHIG HISTORY, AND INCOMPLETE HISTORY

My final goal in this section is to argue that suppositionalists, unlike proponents of the extant views, can escape the charge of whiggishness, but that they are free to practice whig

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

history if they like. Moreover, I will argue that suppositionalists can also escape the charge of giving an incomplete history. In this case, they are able to account for the history of science in a satisfactory way, whereas the extant views are not.

To begin with, there is the charge of giving a whig history. The suppositionalist, unlike the proponents of the extant views, needn't understand the past in terms of the present. In order to understand what a hypothetical entity like the muriatic radical is, and what an HE term like 'muriatic radical' referred to, all that is necessary is that one use some speculative scientific work as a prop, and generate a supposition. An individual living in 1789 and one living in 2013 are just as capable of doing this. To be sure, as I discussed in chapter 2, individuals will sometimes employ interpretive engines that differ from one another, and will, to that extent, generate distinct suppositions. It's likely that an individual living in 2013 will employ an interpretive engine that imports background knowledge that would be unavailable to an individual living in 1789. But it's surely possible for an individual living in 2013 to employ an interpretive engine that would be available to an individual living in 1789. I take it that those historians and philosophers who are especially critical of whig history aim to do something like this when they engage with speculative scientific works. They aim to understand such a work on its own terms, within its historical context. Suppositionalism provides a framework for understanding this kind of engagement with the history of science. Suppositionalists, then, are not forced to practice whig history.

But one might wonder whether suppositionalism allows any place for whig history. I rejected the extant views, not just because whig history is held in low repute, but because the extant views do not allow for any other kind of history. If suppositionalism makes

#### CHAPTER 4. AN OBJECTION TO THE EXTANT VIEWS AND A DEFENSE OF SUPPOSITIONALISM

---

whig history impossible, then one might object that it should be rejected for being just as inflexible as the extant views.

Suppositionalists, however, can also be whiggish, and we can see this by contrasting whiggish and non-whiggish suppositionalists. These two kinds of suppositionalists will differ when it comes to the ways in which they emphasize representations of suppositions and homonymous theories. Suppose that both suppositionalists use Lavoisier's *Elements* as a prop to generate a supposition. A non-whiggish suppositionalist will primarily be concerned with using the representation of the supposition to understand the *Elements* on its own terms. The homonymous theory will give her a way to connect Lavoisier's work to the world, but it will not be her primary means for understanding that work. A whiggish suppositionalist will be more concerned with mapping the representation of a supposition to its homonymous theory, and will take care to replace all of the old HE terms with the corresponding terms that putatively refer to empirical entities. She will want to understand the *Elements* by reference to the present, and the homonymous theory will allow her to do that. However, even the whiggish suppositionalist will stop short of claiming that 'muriatic radical' putatively referred to an empirical entity in 1789. She agrees with the non-whiggish suppositionalist that 'muriatic radical,' in 1789, referred to an object of supposition.

These whiggish and non-whiggish suppositionalists may embody extreme positions, and they may be caricatures to some extent. But their existence shows that suppositionalism allows for these two extreme positions, as well as a variety of positions in between which depend on the relative emphases on representations of suppositions and homonymous theories. Hence, suppositionalism should not be rejected for being too inflexible.

Finally, suppositionalism allows one to give a history that is more complete than the history that the extant views can give. This is because a hypothetical entity has the conformal properties that it does *necessarily*, and if, for some object of supposition  $\tau$  and some property  $\varphi$ ,  $\mathcal{S}(\varphi(\tau))$  is true, then it is analytically true. So in the case of the object of supposition that is Lavoisier's muriatic radical, it necessarily has the conformal property of being a simple substance, and it is analytically true that it is a simple substance according to the supposition. Earlier in the chapter, I posed a series of questions to the realist concerning the transitions that take place in the actual and counterfactual histories of the muriatic radical. I claimed that answers to these questions that appeal to the semantics of the HE term in question will be better than answers that do not appeal to the semantics of the HE term. Suppositionalism gives us a way to point to the semantics of the HE term 'muriatic radical' in order to answer these questions. In  $h_1$ , chemists collectively decided not to use the HE term 'muriatic radical' to refer to an empirical entity because it is part of the meaning of the HE term that the muriatic radical combines with the principle of acidity to form muriatic acid, and no empirical entity fits the bill. In  $h_2$ , Davy decided to use the HE term 'muriatic radical' to refer to hydrogen based on the mistaken idea that chlorine was another principle of acidity. It is part of the meaning of the HE term that the muriatic radical combines with a principle of acidity to form muriatic acid, and Davy mistakenly thought that hydrogen has the normal property of combining with a principle of acidity to form muriatic acid. Finally, in  $h_3$ , Davy decided to use the HE term to refer to chlorine because he viewed chlorine to have the normal property of being a simple substance, and it is part of the meaning of the HE term that the muriatic radical is a simple substance. The

suppositionalist, then, can point to the semantics of the HE term ‘muriatic radical’ in order to give a history of the muriatic radical that is more complete than what the extant views can offer, and is therefore preferable to the extant views as a view of hypothetical entities and HE terms.

#### 4.6 CONCLUSION

This concludes my defense of the suppositionalist view of hypothetical entities. Suppositionalism yields a way in which to understand scientific discourse involving hypothetical entities that is better than any of the extant views. I’ve attempted to show this by considering, in detail, the specific case of the muriatic radical. The extant views face a choice between giving a whig history of the muriatic radical, and giving an incomplete history of that entity. The suppositionalist view does not encounter this problem, and provides a framework that is both fruitful and flexible. If I am right about this, then suppositionalism deserves consideration. It should be applied to other cases in the history of science. It should be further elaborated, developed, and defended. And, in general, it merits further attention.

---

## *Bibliography*

- Barwise, Jon and Etchemendy, John (2002). *Language, Proof and Logic*. CSLI Publications, Stanford, Calif.
- Bayne, Tim and Montague, Michelle (2011). Cognitive Phenomenology: An Introduction. In Tim Bayne and Michelle Montague (Eds.), *Cognitive Phenomenology*, (pp. 1–34). Oxford University Press, Oxford; New York.
- Bensaude-Vincent, Bernadette (1983). A Founder Myth in the History of Sciences? The Lavoisier Case. In Wolf Lepenies, Loren R. Graham, and Peter Weingart (Eds.), *Functions and Uses of Disciplinary Histories*, volume 7 of *Sociology of the Sciences*, (pp. 53–78). D. Reidel Publishing Company, Dordrecht.
- Berthollet, Claude Louis (1788). Mémoire sur l'Acide marin déphlogistiqué. *Histoire de l'Académie Royale des Sciences*, Année 1785:pp. 276–295.
- (1905). Memoir on Dephlogisticated Marine Acid. In *The Early History of Chlorine*, number 13 in Alembic Club Reprints, (pp. 11–31). The Alembic Club, Edinburgh. Translated from "Memoires de l'Académie Royale" for the year 1785. Paris, 1788.

- Berzelius, Jöns Jacob (1813). On the Nature of Muriatic Acid. *Annals of Philosophy*, 2(10):pp. 254–259.
- (1816). Comparison of the Old and New Theories respecting the Nature of Oxymuriatic Acid, to enable us to judge which of the two deserves the Preference. *Annals of Philosophy*, 7-8(40, 42; 45, 46):pp. 272–280, 429–441, 200–209, 256–264.
- Boyd, Richard N. (1981). Scientific Realism and Naturalistic Epistemology. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1980:pp. 613–662.
- (1983). On the Current Status of the Issue of Scientific Realism. *Erkenntnis*, 19(1):pp. 45–90.
- (2010). Realism, Natural Kinds, and Philosophical Methods. In *The Semantics and Metaphysics of Natural Kinds*, number 1 in Routledge Studies in Metaphysics, (pp. 212–234). Routledge, New York.
- Brooke, John Hedley (1980). Davy’s Chemical Outlook: The Acid Test. In Sophie Forgan (Ed.), *Science and the Sons of Genius: Studies on Humphry Davy*, (pp. 121–175). Science Reviews, London.
- Brown, James Robert (1982). Realism, Miracles, and the Common Cause. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1982:pp. 98–106.
- (1985). Explaining the Success of Science. *Ratio*, 27:pp. 49–66.
- Brown, Jessica (2000). Against Temporal Externalism. *Analysis*, 60(266):pp. 178–188.

- Bueno, Otávio (1997). Empirical Adequacy: A Partial Structures Approach. *Studies in History and Philosophy of Science*, 28(4):pp. 585–610.
- (1999). What Is Structural Empiricism? Scientific Change in an Empiricist Setting. *Erkenntnis*, 50:pp. 59–85.
- Bueno, Otávio, French, Steven, and Ladyman, James (2002). On Representing the Relationship between the Mathematical and the Empirical. *Philosophy of Science*, 69(3):pp. 452–473.
- Burge, Tyler (1979). Individualism and the Mental. In Peter A. French, Theodore E. Uehling, and Howard K. Wettstein (Eds.), *Midwest Studies in Philosophy IV: Studies in Metaphysics*, (pp. 73—121). University of Minnesota Press, Minneapolis.
- Butterfield, Herbert (1965). *The Whig Interpretation of History*. Norton, New York. Originally published: London: G. Bell, 1931.
- Caneva, Kenneth L. (2001). *The Form and Function of Scientific Discoveries*. Smithsonian Institution Libraries, Washington, D. C. Dibner Library Lecture, November 16, 2000.
- (2005). ‘Discovery’ as a Site for the Collective Construction of Scientific Knowledge. *Historical Studies in the Physical and Biological Sciences*, 35(2):pp. 175–291.
- Cassirer, Ernst (1956). *Determinism and Indeterminism in Modern Physics; Historical and Systematic Studies of the Problem of Causality*. Yale University Press, New Haven. Originally published in 1936 as *Determinismus und Indeterminismus in der modernen Physik*.

- Cavendish, Henry (1784). Experiments on Air. *Philosophical Transactions of the Royal Society of London*, 74:pp. 119–153.
- Chakravartty, Anjan (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge University Press, Cambridge.
- (2011). Scientific Realism. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2011/entries/scientific-realism>, summer 2011 edition.
- Chalmers, Alan (2008). Atom and Aether in Nineteenth-Century Physical Science. *Foundations of Chemistry*, 10(3):pp. 157–166.
- (2009). *The Scientist's Atom and the Philosopher's Stone How Science Succeeded and Philosophy Failed to Gain Knowledge of Atoms*. Number v. 279 in Boston studies in the philosophy of science. Springer, Dordrecht.
- (2011). Drawing Philosophical Lessons from Perrin's Experiments on Brownian Motion: A Response to van Fraassen. *The British Journal for the Philosophy of Science*, 62(4):pp. 711–732.
- Chang, Hasok (2009). We Have Never Been Whiggish (About Phlogiston). *Centaurus*, 51(4):pp. 239–264.
- (2010). The Hidden History of Phlogiston: How Philosophical Failure Can Generate Historiographical Refinement. *HYLE – International Journal for Philosophy of Chemistry*, 16(2):pp. 47–79.

- (2011). The Persistence of Epistemic Objects Through Scientific Change. *Erkenntnis*, 75(3):pp. 413–429.
- (2012a). Acidity: The Persistence of the Everyday in the Scientific. *Philosophy of Science*, 79(5):pp. 690–700.
- (2012b). *Is Water H<sub>2</sub>O? Evidence, Realism and Pluralism*, volume 293 of *Boston Studies in the Philosophy and History of Science*. Springer, Dordrecht; New York.
- Cooke, Josiah Parsons (1889). The Chemical Elements: History of the Conception Which This Term Involves. *Popular Science*, (pp. 733–750).
- Creath, Richard (2011). Logical Empiricism. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2011/entries/logical-empiricism>, summer 2011 edition.
- Crosland, Maurice P. (1980). Davy and Gay-Lussac: Competition and Contrast. In Sophie Forgan (Ed.), *Science and the Sons of Genius: Studies on Humphry Davy*, (pp. 95–120). Science Reviews, London.
- Currie, Gregory and Ravenscroft, Ian (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Clarendon Press ; Oxford University Press, Oxford : Oxford ; New York.
- da Costa, Newton C. A. and French, Steven (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford Studies in Philosophy of Science. Oxford University Press, Oxford.

Davy, Humphry (1801). Letter to James Watt.

— (1802). *A Syllabus of a Course of Lectures on Chemistry, Delivered at the Royal Institution of Great Britain*. The Press of the Royal Institution of Great Britain, London.

— (1807). The Bakerian Lecture: On Some Chemical Agencies of Electricity. *Philosophical Transactions of the Royal Society of London*, 97:pp. 1–56.

— (1808a). The Bakerian Lecture: On Some New Phenomena of Chemical Changes Produced by Electricity, Particularly the Decomposition of the Fixed Alkalies, and the Exhibition of the New Substances which Constitute Their Bases; and on the General Nature of Alkaline Bodies. *Philosophical Transactions of the Royal Society of London*, 98:pp. 1–44.

— (1808b). Electro-Chemical Researches, on the Decomposition of the Earths; with Observations on the Metals Obtained from the Alkaline Earths, and on the Amalgam Procured from Ammonia. *Philosophical Transactions of the Royal Society of London*, 98:pp. 333–370.

— (1809a). The Bakerian Lecture: An Account of Some New Analytical Researches on the Nature of Certain Bodies, Particularly the Alkalies, Phosphorus, Sulphur, Carbonaceous Matter, and the Acids Hitherto Undecomposed; with Some General Observations on Chemical Theory. *Philosophical Transactions of the Royal Society of London*, 99:pp. 39–104.

— (1809b). Letter to John Dalton.

- (1809c). New Analytical Researches on the Nature of Certain Bodies, Being an Appendix to the Bakerian Lecture for 1808. *Philosophical Transactions of the Royal Society of London*, 99:pp. 450–470.
- (1810a). The Bakerian Lecture for 1809. On Some New Electrochemical Researches, on Various Objects, Particularly the Metallic Bodies, from the Alkalies, and Earths, and on Some Combinations of Hydrogene. *Philosophical Transactions of the Royal Society of London*, 100:pp. 16–74.
- (1810b). Researches on the Oxymuriatic acid, Its Nature and Combinations; and on the Elements of the Muriatic Acid. with Some Experiments on Sulphur and Phosphorus, Made in the Laboratory of the Royal Institution. *Philosophical Transactions of the Royal Society of London*, 100:pp. 231–257.
- (1811a). The Bakerian Lecture: On Some of the Combinations of Oxymuriatic Gas and Oxygene, and on the Chemical Relations of These Principles, to Inflammable Bodies. *Philosophical Transactions of the Royal Society of London*, 101:pp. 1–35.
- (1812a). *Elements of Chemical Philosophy*. Bradford and Inskcep, Philadelphia.
- (1814). An Account of Some New Experiments on the Fluoric Compounds; with Some Observations on Other Objects of Chemical Inquiry. *Philosophical Transactions of the Royal Society of London*, 104:pp. 62–73.
- (1840). *The Collected Works of Sir Humphry Davy*, volume VIII. Smith, Elder and Co. Cornhill, London. Agricultural lectures, part II, and other lectures.

- Davy, John (1811b). Some Remarks on the Observations and Experiments of Mr. Murray on the Nature of Oximuriatic Acid, and its Relation to Muriatic Acid. *A Journal of Natural Philosophy, Chemistry, and the Arts*, by William Nicholson, 28:pp. 193–206.
- (1812b). On the Nature of Oximuriatic and Muriatic Acid Gas, in Reply to Mr. Murray. *A Journal of Natural Philosophy, Chemistry, and the Arts*, by William Nicholson, 31:pp. 310–316.
- (1836). *Memoirs of the Life of Sir Humphry Davy*, volume 1. Printed for Longman, Rees, Orme, Brown, Green, & Longman, London.
- (1839). *The Collected Works of Sir Humphry Davy*, volume I. Smith, Elder and Co. Cornhill, London. *Memoirs of the Life of Sir Humphry Davy*.
- Demopoulos, William and Friedman, Michael (1985). Bertrand Russell's The Analysis of Matter: Its Historical Context and Contemporary Interest. *Philosophy of Science*, 52(4):pp. 621–639.
- Denham, Alison E (2000). *Metaphor and Moral Experience*. Clarendon Press ; Oxford University Press, Oxford; Oxford; New York.
- Dennett, Daniel C. (1991). Real Patterns. *The Journal of Philosophy*, 88(1):pp. 27–51.
- Descartes, René (1988). *Descartes: Selected Philosophical Writings*. Cambridge University Press, Cambridge ; New York.
- Doggett, Tyler and Egan, Andy (2007). Wanting Things You Don't Want: The Case for an Imaginative Analogue of Desire. *Philosopher's Imprint*, 7(9):pp. 1–17.

- Duhem, Pierre (1991). *The Aim and Structure of Physical Theory*. Princeton Science Library. Princeton University Press, Princeton. Originally published as *La Théorie Physique: Son Objet et sa Structure* in 1906.
- Earman, John and Fine, Arthur (1977). Against Indeterminacy. *The Journal of Philosophy*, 74(9):pp. 535–538.
- Field, Hartry (1973). Theory Change and the Indeterminacy of Reference. *The Journal of Philosophy*, 70(14):pp. 462–481.
- Fisher, Alec (1989). Suppositions in Argumentation. *Argumentation*, 3(4):pp. 401–413.
- Frank, Philipp (1941). Physical Theories of the Twentieth Century and School Philosophy. In *Between Physics and Philosophy*, (pp. 55–103). Harvard University Press, Cambridge, Mass. Originally published in 1930.
- French, Steven and Ladyman, James (1999). Reinflating the Semantic Approach. *International Studies in the Philosophy of Science*, 13(2):pp. 103–121.
- (2003). Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure. *Synthese*, 136(1):pp. 31–56.
- Frigg, Roman and Votsis, Ioannis (2011). Everything You Always Wanted To Know About Structural Realism But Were Afraid To Ask. *European Journal for Philosophy of Science*, 1(2):pp. 227–276.
- Frost-Arnold, Greg (2011). From the Pessimistic Induction to Semantic Antirealism. *Philosophy of Science*, 78(5):pp. 1131–1142.

- Gay-Lussac, Joseph Louis and Thénard, Louis Jacques (1808). Sur la décomposition et la recombinaison de l'acide boracique. *Annales de Chimie*, 68:pp. 169—174.
- (1809). Extrait Des Mémoires lus à l'Institut national, depuis le 7 mars 1808 jusqu'au 27 février 1809. *Mémoires de physique et de chimie de la Société d'Arcueil*, 2:pp. 295–358.
- (1905). Abstract of the Memoirs read at the National Institute, from the 7th March 1808 to the 27th February 1809. In *The Early History of Chlorine*, number 13 in Alembic Club Reprints, (pp. 34–48). The Alembic Club, Edinburgh. Translated from "Mémoires de Physique et de Chimie, de la Société d'Arcueil," tome ii. Paris, 1809.
- Gendler, Tamar Szabó (2000). The Puzzle of Imaginative Resistance. *The Journal of Philosophy*, 97(2):pp. 55–81.
- (2011). Imagination. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2011/entries/imagination>, fall 2011 edition.
- Giere, Ronald N. (1985). Constructive Realism. In Paul M. Churchland and Clifford A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*, Science and Its Conceptual Foundations, (pp. 75–98). University of Chicago Press, Chicago.
- Goldman, Alvin I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Philosophy of Mind. Oxford University Press, Oxford ; New York.

- Golinski, Jan (1992). *Science as Public Culture: Chemistry and the Enlightenment in Britain, 1760–1820*. Cambridge University Press, Cambridge.
- Green, Mitchell S. (2000). The Status of Supposition. *Noûs*, 34(3):pp. 376–399.
- Gross, Steven (2011). Does ‘True’s Expressive Role Preclude Deflationary Davidsonian Semantics? Forthcoming in a volume on minimalism, pragmatism, and expressivism, edited by Steven Gross, Nick Tebben, and Michael Williams.
- Guyton de Morveau, Louis Bernard, Lavoisier, Antoine Laurent, Berthollet, Claude Louis, and Fourcroy, Antoine François comte de (1787). *Méthode de Nomenclature Chimique*. Chez Cuchet, Paris.
- (1788). *Method of Chymical Nomenclature*. G. Kearsley, London.
- Hacking, Ian (1982). Experimentation and Scientific Realism. *Philosophical Topics*, 13:pp. 71–87.
- (1985). Do We See Through a Microscope? In Paul M. Churchland and Clifford A. Hooker (Eds.), *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. van Fraassen*, Science and Its Conceptual Foundations, (pp. 132–152). University of Chicago Press, Chicago.
- (2007). Natural Kinds: Rosy Dawn, Scholastic Twilight. *Royal Institute of Philosophy Supplements*, 61:pp. 203–239.
- Halvorson, Hans (2012). What Scientific Theories Could Not Be. *Philosophy of Science*, 79(2):pp. 183–206.

- Hardin, Clyde L. and Rosenberg, Alexander (1982). In Defense of Convergent Realism. *Philosophy of Science*, 49(4):pp. 604–615.
- Heim, Irene (1983). File Change Semantics and the Familiarity Theory of Definiteness. In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow (Eds.), *Meaning, Use, and Interpretation of Language*, (pp. 164–189). Walter de Gruyter, Berlin.
- Hilgevoord, Jan (Ed.) (1994). *Physics and Our View of the World*. Cambridge University Press, Cambridge ; New York.
- Horisk, Claire (2007). The Expressive Role of Truth in Truth-conditional Semantics. *The Philosophical Quarterly*, 57(229):pp. 535–557.
- Ichikawa, Jonathan and Jarvis, Benjamin (2012). Rational Imagination and Modal Knowledge. *Noûs*, 46(1):pp. 127–158.
- Jackman, Henry (1999). We Live Forwards But Understand Backwards: Linguistic Practices and Future Behavior. *Pacific Philosophical Quarterly*, 80:pp. 157–177.
- (2004). Temporal Externalism and Epistemic Theories of Vagueness. *Philosophical Studies*, 117:pp. 79–94.
- (2005). Temporal Externalism, Deference, and Our Ordinary Linguistic Practice. *Pacific Philosophical Quarterly*, 86:pp. 365–380.
- (2006). Temporal Externalism, Constitutive Norms, and Theories of Vagueness. In Tomas Marvan (Ed.), *What Determines Content? The Internalism/Externalism Dispute*, (pp. 221–242). Cambridge Scholars Press, Newcastle.

- Kirwan, Richard (1789). *An Essay on Phlogiston and the Constitution of Acids*. J. Johnson, London, 2nd edition.
- Kitcher, Philip (1978). Theories, Theorists and Theoretical Change. *The Philosophical Review*, 87(4):pp. 519–547.
- (1993). *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press, New York.
- Knight, David M. (1978). *The Transcendental Part of Chemistry*. Dawson, Folkestone.
- (1992). *Humphry Davy: Science and Power*. Blackwell, Cambridge, Mass.
- Kripke, Saul A. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, Mass.
- Kuhn, Thomas S. (1962). Historical Structure of Scientific Discovery. *Science*, 136(3518):pp. 760–764.
- (1996). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, 3rd edition.
- Ladyman, James (1998). What Is Structural Realism? *Studies in History and Philosophy of Science*, 29(3):pp. 409–424.
- (2000). What’s Really Wrong with Constructive Empiricism? van Fraassen and the Metaphysics of Modality. *The British Journal for the Philosophy of Science*, 51(4):pp. 837–856.

- (2009). Structural Realism. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2009/entries/structural-realism>, summer 2009 edition.
- (2011). Structural Realism Versus Standard Scientific Realism: The Case of Phlogiston and Dephlogisticated Air. *Synthese*, 180(2):pp. 87–101.
- Ladyman, James and Ross, Don (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press, Oxford. With John Collier and David Spurrett.
- Laudan, Larry (1981). A Confutation of Convergent Realism. *Philosophy of Science*, 48(1):pp. 19–49.
- Lavoisier, Antoine Laurent (1776). Mémoire sur l'existence de l'air dans l'acide nitreux, et sur les moyens de décomposer et de recomposer cet acide. *Mémoires de l'Académie des sciences*, (pp. 671–680).
- (1778). Considérations générales sur la nature des acides et sur les principes dont ils sont composés. *Mémoires de l'Académie des sciences*, (pp. 535–547).
- (1783). *Essays on the Effects Produced by Various Processes on Atmospheric Air: With a Particular View to an Investigation of the Constitution of the Acids*. W. Eyers, London.
- (1801). *Traité élémentaire de chimie présenté dans un ordre nouveau, et d'après les découvertes modernes*, volume II. Chez Deterville, Paris, 3rd edition.
- (1802). *Elements of Chemistry*, volume I. G. and J. Robinson, and T. Kay, Edinburgh, 5th edition.

- Le Grand, H.E. (1974). Ideas on the Composition of Muriatic Acid and Their Relevance to the Oxygen Theory of Acidity. *Annals of Science*, 31(3):pp. 213–225.
- Leslie, Alan M. (1994). Pretending and Believing: Issues in the Theory of ToMM. *Cognition*, 50(1–3):pp. 211–238.
- Lewis, David (1970). How to Define Theoretical Terms. *The Journal of Philosophy*, 67(13):pp. 427–446.
- (1978). Truth in Fiction. *American Philosophical Quarterly*, 15(1):pp. 37–46.
- (1984). Putnam’s Paradox. *Australasian Journal of Philosophy*, 62(3):pp. 221–236.
- Linsky, Bernard and Zalta, Edward N. (1995). Naturalized Platonism versus Platonized Naturalism. *The Journal of Philosophy*, 92(10):pp. 525–555.
- Lutz, Sebastian (2012). On a Straw Man in the Philosophy of Science: A Defense of the Received View. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 2(1):pp. 77–120.
- Mackie, J. L. (1973). *Truth, Probability, and Paradox: Studies in Philosophical Logic*. Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford.
- Maxwell, Grover (1962). The Ontological Status of Theoretical Entities. In Herbert Feigl and Grover Maxwell (Eds.), *Scientific Explanation, Space and Time*, volume 3 of *Minnesota Studies in the Philosophy of Science*, (pp. 3–27). University of Minnesota Press, Minneapolis.

- (1970). Structural Realism and the Meaning of Theoretical Terms. In Stephen Winokur and Michael Radner (Eds.), *Analyses of Theories and Methods of Physics and Psychology*, volume 4 of *Minnesota Studies in the Philosophy of Science*, (pp. 181–192). University of Minnesota Press, Minneapolis.
- Melnyk, Andrew (2003). *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge studies in philosophy. Cambridge University Press, Cambridge, UK.
- Miller, Arthur I. (1981). *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905-1911)*. Addison-Wesley, Reading Mass.
- Monton, Bradley and van Fraassen, Bas C. (2003). Constructive Empiricism and Modal Nominalism. *The British Journal for the Philosophy of Science*, 54(3):pp. 405–422.
- Murray, John (1811a). Observations and Experiments on the Nature of Oximuriatic Acid, and its Relation to Muriatic Acid. *A Journal of Natural Philosophy, Chemistry, and the Arts, by William Nicholson*, 28:pp. 132–152.
- (1811b). On the Nature of Oximuriatic Acid, in Reply to Mr. John Davy. *A Journal of Natural Philosophy, Chemistry, and the Arts, by William Nicholson*, 28:pp. 294–310.
- Musgrave, Alan (1988). The Ultimate Argument for Scientific Realism. In Robert Nola (Ed.), *Relativism and Realism in Science*, volume 6 of *Australasian Studies in History and Philosophy of Science*, (pp. 229–252). Kluwer Academic Publishers, Dordrecht.
- (2006). The 'Miracle Argument' For Scientific Realism. *The Rutherford Journal*, 2.

- Newman, M. H. A. (1928). Mr. Russell's "Causal Theory of Perception". *Mind*, 37(146):pp. 137–148.
- Nichols, Shaun (2006). Imaginative Blocks and Impossibility: An Essay in Modal Psychology. In Shaun Nichols (Ed.), *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction*, (pp. 237–256). Clarendon Press, Oxford.
- Nicholson, William (1795). *A Dictionary of Chemistry*, volume 1. G. G. and J. Robonson, Paternoster Row, London.
- Odling, William (1871). *On Improvements in the Manufacture of Chlorine*. Royal Institution of Great Britain, London.
- Papineau, David (1996). Theory-Dependent Terms. *Philosophy of Science*, 63(1):pp. 1–20.
- (2010). Realism, Ramsey sentences and the pessimistic meta-induction. *Studies In History and Philosophy of Science Part A*, 41(4):pp. 375–385.
- Parsons, Terence (1980). *Nonexistent Objects*. Yale University Press, New Haven.
- Partington, J. R (1961). *A History of Chemistry*, volume 3. Macmillan, London.
- Peacocke, Christopher (1985). Imagination, Experience, and Possibility: A Berkeleyian View Defended. In John Foster and Howard Robinson (Eds.), *Essays on Berkeley: A Tercentennial Celebration*, (pp. 19–35). Clarendon Press, Oxford.
- Perrin, Jean (1916). *Atoms*. D. van Nostrand Company, New York. Translated from the 4th revised edition of *Les atomes*.

- Poincaré, Henri (1962). *Science and Hypothesis*. Dover Publications, New York. Originally published as *La Science et l'Hypothèse* in 1902.
- Priestley, Joseph (1892). *Scientific Correspondence of Joseph Priestley: Ninety-seven Letters Addressed to Josiah Wedgwood, Sir Joseph Banks, Capt. James Keir, James Watt, Dr. William Withering, Dr. Benjamin Rush, and Others*. Collins Printing House, New York.
- Psillos, Stathis (1996). Scientific Realism and the 'Pessimistic Induction'. *Philosophy of Science*, 63:pp. S306–S314.
- (1999). *Scientific Realism: How Science Tracks Truth*. Philosophical Issues in Science. Routledge, London.
- Putnam, Hilary (1975a). The Meaning of 'Meaning'. In *Mind, Language, and Reality: Philosophical Papers*, volume 2, (pp. 215–271). Cambridge University Press, Cambridge.
- (1975b). What is Mathematical Truth? In *Mathematics, Matter, and Method: Philosophical Papers*, volume 1, (pp. 60–78). Cambridge University Press, Cambridge.
- (1975c). What Theories are Not. In *Mathematics, Matter, and Method: Philosophical Papers*, volume 1, (pp. 215–227). Cambridge University Press, Cambridge.
- (1978). *Meaning and the Moral Sciences*. Routledge & K. Paul, London.
- Quine, W. V. O. (1948). On What There Is. *Review of Metaphysics*, 2:pp. 21–38.

Remsen, Ira (1901). *A College Text-book of Chemistry*. H. Holt and company, New York, first edition.

Russell, Bertrand (1905). On Denoting. *Mind*, 14(56):pp. 479–493.

— (1927). *The Analysis of Matter*. International Library of Psychology, Philosophy and Scientific Method. Harcourt, Brace ; K. Paul, Trench, Trubner, New York : London.

Rynasiewicz, Robert (2009). Science, Supposition and Reference: The New Program. A talk given at the University of Pittsburgh, December 4, 2009.

Rynasiewicz, Robert, Steinert-Threlkeld, Shane, and Suri, Vivek (2010). Mathematical Existence De-Platonized: Introducing Objects of Supposition in the Arts and Sciences. Preprint on PhilSci Archive. <http://philsci-archive.pitt.edu/archive/00005345>.

Salmon, Nathan (1986). *Frege's Puzzle*. MIT Press, Cambridge, Mass.

Sankey, Howard (2008). *Scientific Realism and the Rationality of Science*. Ashgate, Aldershot, England.

Scheele, Carl Wilhelm (1786). On Manganese, Manganese, or Magnesia Vitrariorum. In *The chemical essays of Charles-William Scheele. Translated from the transactions of the Academy of Sciences at Stockholm. With additions*. printed for J. Murray; W. Gordon and C. Elliot, Edinburgh, London. Translated from "Om Brunsten, eller Magnesia, och dess Egenskaper," Kong. Vetenskaps Academiens Handlingar, xxxv. Stockhom, 1774.

- Schönbein, Christian Friedrich (1857). On the Connexion of Catalytic Phænomena with Allotropy. *Philosophical Magazine*, 13(86):pp. 248–261.
- (1858). Further Observations on the Allotropic Modifications of Oxygen, and on the Compound Nature of Chlorine, Bromine. *Philosophical Magazine*, 16(106):pp. 178–182.
- Siegfried, Robert (1959). The Chemical Philosophy of Humphry Davy. *Chymia*, 5:pp. 193–201.
- (1964). The Phlogistic Conjectures of Humphry Davy. *Chymia*, 9:pp. 117–124.
- Soames, Scott (1987). Direct Reference, Propositional Attitudes, and Semantic Content. *Philosophical Topics*, 15(1):pp. 47–87.
- Stalnaker, Robert (1998). On the Representation of Context. *Journal of Logic, Language and Information*, 7(1):pp. 3–19.
- Stanford, P. Kyle and Kitcher, Philip (2000). Refining the Causal Theory of Reference for Natural Kind Terms. *Philosophical Studies*, 97(1):pp. 97–127.
- Strawson, P. F. (1950). On Referring. *Mind*, 59(235):pp. 320–344.
- Suppe, Frederick (Ed.) (1977). *The Structure of Scientific Theories*. University of Illinois Press, Urbana, 2d ed edition.
- Suppes, Patrick (1957). *Introduction to logic*. The University series in undergraduate mathematics. D. Van Nostrand Co, Princeton, N.J.

- Uebel, Thomas (2011). Vienna Circle. In Edward N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2011/entries/vienna-circle>, summer 2011 edition.
- van Fraassen, Bas C. (1966). Singular Terms, Truth-Value Gaps, and Free Logic. *The Journal of Philosophy*, 63(17):pp. 481–495.
- (1975). Platonism’s Pyrrhic Victory. In Alan Ross Anderson, Ruth Barcan Marcus, and R. M. Martin (Eds.), *The Logical Enterprise*, (pp. 39–50). Yale University Press, New Haven.
- (1980). *The Scientific Image*. Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford.
- (1989). *Laws and Symmetry*. Clarendon Press, Oxford.
- (1992). From Vicious Circle to Infinite Regress, and Back Again. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2:pp. 6–29.
- (2001). Constructive Empiricism Now. *Philosophical Studies*, 106:pp. 151–170.
- (2006). Structure: Its Shadow and Substance. *The British Journal for the Philosophy of Science*, 57(2):pp. 275–307.
- (2008). *Scientific Representation: Paradoxes of Perspective*. Clarendon Press, Oxford.
- van Inwagen, Peter (1977). Creatures of Fiction. *American Philosophical Quarterly*, 14(4):pp. 299–308.

- Vickers, Peter (2010). Can Partial Structures Accommodate Inconsistent Science? *Principia: an international journal of epistemology*, 13(2):pp. 233–250.
- Votsis, Ioannis (2004). *The Epistemological Status of Scientific Theories: An Investigation of the Structural Realist Account*. PhD thesis, University of London, London School of Economics.
- (2005). The Upward Path to Structural Realism. *Philosophy of Science*, 72(5):pp. 1361–1372.
- (2007). Uninterpreted Equations and the Structure-Nature Distinction. *Philosophical Inquiry*, 29(1-2):pp. 57–71.
- Walton, Kendall L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.
- Weatherson, Brian (2004). Morality, Fiction, and Possibility. *Philosopher's Imprint*, 4(3):pp. 1–27.
- Webster, John W. (1826). *A Manual of Chemistry, on the Basis of Professor Brande's*. Richardson & Lord, Boston.
- Weinberg, Jonathan M. and Meskin, Aaron (2006). Puzzling Over the Imagination: Philosophical Problems, Architectural Solutions. In Shaun Nichols (Ed.), *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction*, (pp. 175–202). Clarendon Press, Oxford.

- White, W. Jeffrey (1903). Note on the Philosophy of a Supposition. *Mind*, 12(48):pp. 502–506.
- Wilde, C. B. (1981). Whig History. In William F. Bynum, E. Janet Browne, and Roy Porter (Eds.), *Dictionary of the history of science*, (pp. 445–446). Princeton University Press, Princeton, N.J.
- Wilson, Mark (1982). Predicate Meets Property. *The Philosophical Review*, 91(4):pp. 549–589.
- Wimsatt Jr., William K. and Beardsley, Monroe C. (1946). The Intentional Fallacy. *The Sewanee Review*, 54(3):pp. 468–488.
- Worrall, John (1989). Structural Realism: The Best of Both Worlds? *Dialectica*, 43(1–2):pp. 99–124.
- (2007). Miracles and Models: Why Reports of the Death of Structural Realism May be Exaggerated. *Royal Institute of Philosophy Supplements*, 61:pp. 125–154.
- (2008). Why the Structural Realist has Nothing to Fear from "Underdetermination". From Theoretical Frameworks and Empirical Underdetermination Workshop.
- (2011). Underdetermination, Realism and Empirical Equivalence. *Synthese*, 180(2):pp. 157–172.
- Worrall, John and Zahar, Elie (2001). Ramseyfication and Structural Realism. In *Poincaré's Philosophy: From Conventionalism to Phenomenology*, (pp. 236–251). Open Court, Peru, Illinois.

Yablo, Stephen (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, 53(1):pp. 1–42.

— (2002). Coulda, Woulda, Shoulda. In Tamar Szabó Gendler and John Hawthorne (Eds.), *Conceivability and possibility*, (pp. 441–492). Clarendon Press, Oxford ; New York.

---

## *Curriculum Vita*

Jonathon Hricko was born on May 15, 1982, in Pittsburgh, Pennsylvania. In 2004, he received his B.A. in philosophy from the University of Pittsburgh, along with minors in Japanese and Asian Studies. He began his graduate work in philosophy at the Johns Hopkins University in 2005, and has interests in the history and philosophy of science, late-eighteenth and early-nineteenth century chemistry, language, logic, mind, and moral psychology. During his time at Hopkins, he was a Dean's Teaching Fellow for two semesters, and spent six semesters as an instructor in the Expository Writing Program. He also spent three years as a Writing Center tutor, and served as a graduate adviser to Prometheus, the undergraduate philosophy organization at Hopkins, for four years. He took part in many reading groups with his fellow graduate students, and has presented his work at a number of conferences.