I, A. Leland Morton, hereby submit this work as part of the requirements for the degree of:

Doctor of Philosophy

in:

Philosophy

It is entitled:

Making It Intelligible: An Historical Approach to Understanding Intelligibility in the Assessment of Scientific Theories.

This work and its defense approved by:

Chair: Robert C. Richardson, Ph. D.
      John Bickle, Ph. D.
      William Jensen, Ph. D.
      John McEvoy, Ph. D.
      Thomas Polger, Ph. D.
      Robert Skipper, Ph. D.
Making It Intelligible:  
An Historical Approach to Understanding Intelligibility in the Assessment of Scientific Theories.

A Dissertation submitted to the 
Division of Research and Advanced Studies 
Of the University of Cincinnati

in partial fulfillment of the 
requirements for the degree of

Doctor of Philosophy (Ph. D.)

in the Department of Philosophy 
of the McMicken College of Arts and Sciences

May 19, 2006

by

A. Leland Morton, M.A.

Committee

Chair: Robert C. Richardson, Ph. D.  
Member: John Bickle, Ph. D.  
Member: William B. Jensen, Ph. D.  
Member: John McEvoy, Ph. D.  
Member: Thomas Polger, Ph. D.  
Member: Robert Skipper, Ph. D.
Abstract

The principal goal of this dissertation is to develop and defend a criterion of intelligibility adequate to the task of explaining the variety of historical appeals to this notion in philosophy and, more importantly, science. The thesis I defend holds that any account capable of explaining the historical use of intelligibility in criticizing scientific explanations must begin with the fact that charges generally encapsulated concerns with theoretical terms and their associated representational contents. For individuals like Leibniz and Boyle as well as latter atomists and other nineteenth century scientists like Maxwell the pertinent question was whether the putative referents of the theoretical terms of explanations could be conceived of. “Conceiving of” for early modern thinkers had principally to do with associating terms with ideas that are either rationally grounded or derived from experience. Later theorists like Maxwell, while remaining faithful to the traditional use of intelligibility, implicitly extend the idea by associating theoretical abstractions with models that are accessible to the imagination. After establishing this core thesis I proceed to develop an alternative account of intelligibility that can accommodate the historical worries while at the same time addressing some potential philosophical challenges to the use of this notion in the assessment of scientific theories. The thesis defended in this regard is that terms are intelligible if and only if they can be associated with observationally based representations that figure in a broader network of such representations whose relations are suitably constrained by the theory.
# Table of Contents

Preface: 3

Chapter I:
- 1.0.0. Introduction 16
- 1.1.0. Leibniz's Metaphysical Naturalism 20
  - 1.1.1. A Tale of Two Kingdoms 24
- 1.2.0. Intelligibility, Miracles, & Occult Qualities 31
- 1.3.0. Conclusion 43

Chapter II:
- 2.0.0. Introduction 55
- 2.1.0. Sine Mente Soni 56
- 2.2.0. Elaborating the Theory of Ideas 59
- 2.3.0. 'Attraction' vs. Action-at-a-distance 64
- 2.4.0. Buchdahl and the Role of Intelligibility in Theory Assessment 71
  - 2.4.1. Evaluating Buchdahl 80
- 2.5.0. Concluding Remarks 85

Chapter III:
- 3.0.0. Introduction 87
- 3.1.0. Humean Intelligibility 90
  - 3.1.1. Articulating the Problem 95
  - 3.1.2. Loeb’s Solution 97
  - 3.1.3. Problems with Loeb’s Interpretation 101
  - 3.1.4. Hume’s Solution 105
- 3.2.0. HC Reconsidered 110
  - 3.2.1. Inferentialism 112
  - 3.2.2. Cognitive Significance 115

Chapter IV:
- 4.0.0. Introduction 120
- 4.1.0. Semantic Agnosticism, Observation, & Cognitive Significance 122
  - 4.1.1. Observation 124
  - 4.1.2. Miscellany 126
- 4.2.0. Constraining Representations 129
- 4.3.0. Boyle, Models, & Theories 137
- 4.4.0. Pragmatic Virtues 148
  - 4.4.1. Criticism, Rejoinder, & Resolution 150

Chapter V:
- 5.0.0. Introduction 157
<table>
<thead>
<tr>
<th>Chapter VI:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0.0. Introduction</td>
<td>186</td>
</tr>
<tr>
<td>6.1.0. Elaborating De regt’s and Dieks’ Model (DND)</td>
<td>192</td>
</tr>
<tr>
<td>6.2.0. Understanding: Concerns &amp; Confusions</td>
<td>197</td>
</tr>
<tr>
<td>6.3.0. Criticisms of CIT</td>
<td>206</td>
</tr>
<tr>
<td>6.3.1. CIT’s Radical Subjectivism</td>
<td>206</td>
</tr>
<tr>
<td>6.3.2. CIT’s commitment to SST</td>
<td>212</td>
</tr>
<tr>
<td>6.3.3. CIT’s Holistic Character</td>
<td>222</td>
</tr>
<tr>
<td>6.3.4. CIT’s Extrinsic Character</td>
<td>223</td>
</tr>
</tbody>
</table>

Chapter VI: 226

Bibliography: 231
Preface

Introduction

The principal goal of my dissertation is to develop a criterion of intelligibility adequate to the task of explaining the variety of historical appeals to this notion in philosophy and, more importantly, science. Questions regarding intelligibility arise as part of a broader concern with the nature of understanding. And often the two notions are treated interchangeably. Thus, we might seek to understand or render intelligible the process of combustion, or we may attempt to understand or find intelligible an explanation of combustion. While ‘understanding x’ and ‘finding x intelligible’ might reasonably be treated as synonymous in each of the questions, the two questions presuppose different objects of understanding.

An interest in the nature of what is understandable or intelligible may arise in the context of attempting to answer either of the following more general questions:

1)  How do scientific theories (or other explanatory strategies) generate understanding?

2)  What is involved in our understanding of scientific theories?

Nicholas Rescher expresses the former as a question concerning how “the majestic lawful order of nature is intelligible to us”. Hempel provides a potential answer, holding that scientific theories generate understanding when they subsume phenomena under general regularities. On the other hand, sensitive to the concern expressed in the second question, James Clerk Maxwell adopted the method of physical analogy, which he saw as

---

1 Rescher, Nicholas (??). Nature and Understanding.
a necessary aid in our understanding of formal mathematical explanations. In doing so he sought a compromise between the positivist and structuralist philosophies that characterized 19th century science. Unlike Hempel, Maxwell’s concern was not simply with how theories enable us to explain and hence understand the “majestic lawful order” of nature but with how we are able to understand the theories and laws themselves. For Maxwell, the use of physical analogy allowed one to understand the mathematical explanation, “to present the mathematical ideas to the mind in an embodied form, as a systems of lines and surfaces, and not as mere symbols.”

Whereas 1) takes nature to be the pertinent object of understanding, the object of understanding at issue in 2) is language, the theoretical language of science in particular. One potential general answer to the first question—though not one I aim to defend here—is that theories or explanations generate understanding of their target phenomena by tying the phenomena to a set of processes or an underlying theoretical structure which, if sufficient, would account for the occurrence of the phenomena. Roughly speaking then, theories may be thought of as sets of sentences or even more loosely as stories concerning the organization and structure of nature in which propositions describing the phenomena may be intelligibly situated.

Again, I don’t propose this as a complete or even as an accurate first step toward an account of what a theory or explanation is. Nor do I wish to defend it as the right sort of response to 1). The principle reasons for mentioning it here are two-fold. First, something approximating this account plays a prominent role in the development of the views discussed in the final two chapters of the dissertation. Henk De regt and Denis

---

Dieks (discussed in chapter VI), for instance, maintain that we understand a phenomenon only when we have some such story of the underlying processes and we are able to see how the phenomenon arises from them as a consequence (though not necessarily deductively.) Twentieth century logical positivists, on the other hand, see understanding of phenomena to come about by exhibiting the statements describing them as logical consequences of the statements comprising the theoretical framework.

The second—and for present purposes more important—reason for mentioning the above response to 1) is that it makes for a fairly straightforward connection between the two questions. If the task of a theory is to generate understanding of its target phenomena and it achieves this by providing us with a story of the underlying structure or organization of nature within which propositions describing the phenomena may be inferentially situated, it seems reasonable that these stories should themselves be comprehensible. The language employed in revealing “the majestic lawful order of nature” should not itself be mysterious or opaque. The theory should be couched in terms that can be understood. Though understanding a given theory may require effort, it ought to at least be possible. The question taken up here thus concerns what is required in order for us to be able to comprehend the theoretical language of science.

While 1) and 2) are related, there are good reasons to treat them as distinct questions. In the first place, understanding a theory certainly does not entail understanding of the phenomena which the theory purports to explain—this for the simple reason that we can understand theories which are false, even radically so. Second, even if Hempel’s (or some alternative) response to the first question is on target, Maxwell’s worry would remain. Having exhibited a proposition as rationally or logically
connected with a set of statements, the question persists regarding whether and how we are able to understand this new set of propositions and, equally, the terms employed therein. So while the two questions are no doubt naturally related, they define distinct conceptual spaces.

Setting aside (for now) the question of their relation the focus of this dissertation is on the second question. That is, the focus is on how we understand theories (and language more generally), rather than on the manner in which successful theories generate an understanding of the phenomena they propose to explain. Broadly, the issue taken up here regards the nature of linguistic comprehension. More narrowly, the issue concerns the nature and extent of our apprehension of the theoretical and frequently abstract language of scientific explanations.

As stated previously my principal interest is in the development of a criterion adequate to the task of explaining the historical data; the historical data being those paradigmatic historical episodes wherein theories were criticized on the grounds that they were incomprehensible or unintelligible. At a minimum, I take it that an adequate response to 2) ought to be capable of explaining those episodes.

Concrete examples of such episodes are not difficult to find in the historical literature. Boyle characterized many of the theories proposed by his contemporaries as variously “dark”, “odd”, “ambiguous”, & “unclear.” He clearly marked their attempts to explain chemical phenomena as unintelligible. In the same period Leibniz characterized Newton’s appeal to ‘attractive’ force as “unintelligible”, “inexplicable”, “mysterious”

---

and so on. The ongoing controversy between adherents of the structuralist (e.g. Atomists) and positivist philosophies in the nineteenth century constitute a further episode. Atomists like Kekulé held that without some account of the underlying structure of matter we have no idea what the terms of our chemical formula might refer to. In each of these cases; Boyle’s charge against the alchemists and peripatetic (Aristotelian) chemists as well as Leibniz’s critique of Newton, and the structuralists’ antipathy toward the formalistic explanations of positivists, the characterization of aspects these views as incomprehensible was made explicit. The central task I take up here is to construct an account that comprehends the presumed common threads of these criticisms.

The project so presented is to develop an account tailored to explaining the above historical data. Given this, some preliminary concerns should to be addressed. First, am I entitled to presume that the above remarks constitute a coherent and univocal set of worries? If indeed they do not frame a single homogeneous problem space, then it seems likely that any attempt to articulate the worry will necessarily miss the mark.

A second worry pertains less to my method than to the scope of my conclusion. It is certainly true that I have not responded to every account of linguistic understanding that can be plausibly inferred from the extant literature. This is a potential problem. Granting that one cannot hope to have canvassed every remark on a given matter, my claim is for an account of a particular issue that surfaces historically and principally with the occurrence of this issue in the history of science. Moreover, it is a consequence of my account that much of what philosophers may have to say with regard to our comprehension of language is, even if correct, irrelevant to the specific issue taken up

---

here. There is further work to be done in order to develop the thesis a part of a more
general account of our comprehension of language. I do not undertake that work here.

To the first question, whether I am entitled to the presumption that these historical
episodes present a univocal worry, I grant that this requires an argument. In chapter II I
attempt to make the case that the comments made by Boyle and Leibniz, as well as others
in their time, can reasonably be situated as expressions of a common attitude concerning
the relationship between language and thought. Of course that discussion doesn’t settle
the issue concerning the later nineteenth century structuralist opposition to positivistic
prescriptions for science. I acknowledge this. Moreover, I concede that the primary case
from which I extract my account is that of Leibniz’s critique of Newton’s appeal to
‘attraction’.

There is no doubt that Leibniz’s concern with the cognitive status of ‘attraction’
constitutes the focal point for understanding the historical concern with language—and
there are good reasons for this that extend beyond my thesis. It is, of course, possible that
we may be misled by the investigation of such a case but it is nonetheless a necessary
first step in accounting for the general worry implied there. Beyond this the view I extract
is indeed intended to provide some understanding of the discussion concerning the
appropriate methodology for the generation of scientific understanding that came about in
the nineteenth century. Whether or not my account of intelligibility can meet that
challenge is a fair test of the view. In short then, I don’t presume that the historical
concerns I have mentioned constitute a homogeneous set of worries. Rather, I endeavor
to show that they do and moreover that it is possible to develop a unified account capable
of explaining these episodes.
Of course this doesn’t mean that we ought to presume that these characterizations were accurate. As with others, the claims made by Boyle and Leibniz may have been mistaken. But to assess whether they were, we need to first appreciate the sense in which they held these theories to be incomprehensible. Moreover, we should not presume that the inability to understand a theory (or language generally) involves in every instance a failure of a common cognitive mechanism.

*Prima facie* there seem to be a variety of ways that we might fail to understand a theory. At least two are directly relevant for developing an understanding of the aforementioned episodes. On the one hand, we might find the theory confused, be unable to see clearly what follows from it. We may fail to be able to grasp the use of certain terms, how they are integrated in the logic of the theory. A second sort of failure may be realized when the terms of the theory fail to be conceivable in which case we are unable, in the language of one theorist, to “frame a picture” of their reference.7

Hume’s critique of the metaphysicians’ talk of ‘substance’ focused on the latter sort of failure. He understood well enough the process of abstraction, as an operation on ideas, which the metaphysicians employed in order to infer their putative notion. Moreover, he understood the sort of explanatory work that the proposed notion was thought to do. The problem of course was that the process of abstraction carried to this end stripped the term of any associated cognitive content. Similarly, Leibniz did not

---

7 The expression comes from the following passage in, Buchdahl, Gerd (1970a). “Gravity and Intelligibility: Newton to Kant,” in *The Methodological Heritage of Newton*, ed. Robert E. Butts & John W. Davis (Toronto: University of Toronto Press), 1970, pp. 74-102. Here Buchdahl list a number of other potential sources: *Certainly, we meet with a whole spectrum of cases, from one extreme of straight logical impossibility, via Hume’s approach which proceeds from his contention that actio in distans is contrary to one of his criteria for causal action, spatial contiguity, via a supposed lack of “familiarity,” and “inability to frame a picture,” via the supposed clash with predominant physical theories or conceptual schemes, all the way to the other extreme of straight muddleheadedness or even prejudice* (p. 85).
complain that he did not understand the consequences of Newton’s theory. On the contrary he conceded that it was predictive of certain celestial phenomena.

Whether or not we agree with Hume’s critique of the metaphysicians or with Leibniz’s critique of Newton is beside the present point. It is clear enough that these criticisms are distinct from concerns over the logic or use of terms like ‘substance’ or ‘attraction’. In order to better frame my project I construe worries over the use of terms or expressions as concerns respecting the terms’ articulation. For the issue at stake in the latter worries, those of Leibniz and Hume for example, I employ the term intelligibility. The principal goal of this dissertation then is to develop an account of intelligibility in this sense.

**Thesis and Outline**

The thesis I defend holds that any account capable of explaining the historical use of intelligibility in criticizing scientific explanations must begin with the fact that charges generally encapsulated concerns with theoretical terms and their associated representational contents. For individuals like Leibniz and Boyle as well as latter atomists and other nineteenth century scientists like Maxwell the pertinent question was whether the putative referents of the theoretical terms of explanations could be conceived of. “Conceiving of” for early modern thinkers had principally to do with associating terms with ideas that are either rationally grounded or derived from experience. Later theorists like Maxwell, while remaining faithful to the traditional use of intelligibility, implicitly extend the idea by associating theoretical abstractions with models that are accessible to the imagination. After establishing this core thesis I proceed to develop an alternative account of intelligibility that can accommodate the historical worries while at the same
time addressing some potential philosophical challenges to the use of this notion in the assessment of scientific theories.

The dissertation begins with a reconsideration of Leibniz’s criticism of ‘attraction’. The standard reading of the criticism treats ‘attraction’ and ‘action at a distance’ to be interchangeable and correspondingly views Leibniz’s use of locutions like “unintelligible”, “inexplicable”, “miraculous”, “non-mechanical” and “occult” to be roughly synonymous. Philosophers with this view in mind will generally see Leibniz’s critique as a univocal one, according to which action at a distance and attraction are held to be incomprehensible primarily because such notions were seen to be incommensurate with the mechanical philosophy. Against this view I argue that Leibniz’s criticism is more complex.

I hold that Leibniz’s criticism in fact took the form of a tri-lemma. Specifically, I demonstrate that “action at a distance” was not viewed by Leibniz to be unintelligible. Nor was it assumed by him to be synonymous with “attraction”. That identification was but one option available to Newton and his exponents like Clarke. As far as Leibniz was concerned if Newton and his advocates wanted to explain gravitational phenomena in terms of actions at a distance, then they faced a dilemma. If true such actions could only be accounted for in terms of God’s continual intervention in the course of nature; a consequence that Leibniz objected to on both theological and rational grounds. Obviously, such an explanation was not mechanical and arguably was not one that could be made consistent with the demand to provide naturalistic explanations in science. If not miraculous then the notion of an action at a distance involved an absurdity and was thus false. Neither of these alternatives was acceptable to Newton or his advocates like Clarke
and consequently Leibniz argued they were forced into accepting a third alternative; they must be appealing to some unintelligible notion. Leibniz held terms to be true, false, or unintelligible. The claim that ‘attraction’ was unintelligible clearly did not stem from its being construed as action at a distance. On the contrary, I show that it was precisely because Newton and Clarke did not want to associate it with action at a distance that Leibniz thought that the charge of speaking unintelligibly was licensed. Given the inability to associate ‘attraction’ with any definite cognitive content, which might be judged true or false, the term was thought to be empty or vacuous.

After clarifying Leibniz’s broader criticism I turn to developing the specific charge that ‘attraction’, in Newton’s use of it, is unintelligible. I argue that at base Leibniz’s claim appealed to a “core conception” of intelligibility widely endorsed during the 17th and 18th centuries. According to this core conception terms were thought to be intelligible if and only if there was some associated idea. In part the second chapter of the dissertation is devoted to illustrating the manner in which this core conception had bipartisan support from both empiricists and rationalists. I illustrate the manner in which, among others, Descartes, Hobbes, Arnauld, Locke, and later Hume all made appeals to such a standard.

The core conception, however, does not provide a complete account of intelligibility, but a general constraint on terms. From the point of view of an individual committed to the theory of ideas it is nearly tautological. What is needed is not only an account of the nature of ideas but clarification of the relationship between theoretical terms and their associated ideas. Many contemporary philosophers tend to see the early
modern picture of the relationship between terms and ideas to be, at base, a misguided attempt at a theory of reference or meaning. I contend here that this view is mistaken.

Short of claiming that none of the thinkers of this period articulated anything that may justifiably be labeled a theory of meaning I argue that when seen as an account of a term’s intelligibility, the early modern picture of the relationship between terms and ideas is not incompatible with contemporary semantic theories. I maintain that intelligibility is an epistemic issue concerning the cognitive status of theoretical terms. Moreover, it is not a worry respecting the role played by theoretical abstractions in the theory or, consequently, the predictiveness of the theory. Nor is it an issue concerned with commensurability or the conformity of theoretical concepts to accepted regulative or explanatory norms.

While the later issues were raised in the context of the scientific episodes under consideration they were conceptually distinct from the concern over intelligibility. Leibniz, for instance, certainly raised the specter of incommensurability with regard to action at a distance but as I show the challenge here did not have to do with the intelligibility of action at a distance but with the idea that this notion, action at a distance, was incompatible with the mechanism supposed inherent in the operations of nature. Equally, Leibniz, in maintaining that ‘attraction’ was unintelligible, was not objecting to the constitution, logic, or predictiveness of the theory. He had conceded the predictive value of the appeal to “attraction” even while questioning its intelligibility.

Based on this examination of its historical significance I proceed to develop an account of intelligibility. There are, to be sure philosophical problems with adopting straightforwardly the early modern conception of intelligibility. One of the central
problems has to do with specifying which ideas or representational contents are the relevant ones to associate with a given theoretical term. An account provided simply in terms of ideas and their associations provides no direct or explicit answer to this question. Additionally, if the account of intelligibility in terms of ideas is taken as an exhaustive of a term’s cognitive significance, it seemingly leaves something out. This latter worry can perhaps be best explained by seeing how Hume employ’s a criterion of intelligibility in critique of the metaphysicians’ talk of ‘substance’.

My aim is thus to develop a criterion of intelligibility that is responsive to the philosophical issues involved in the application of a criterion of intelligibility in assessing theories and explanations and one which equally accounts for the historical cases. The account I develop construes the *intelligibility* and *articulation* of terms as two distinct aspects of a terms cognitive significance. Intelligibility, I maintain, depends on our ability to associate terms with appropriate cognitive constructs. Articulation I construe as a worry over the theoretical role of terms or expressions.

The cognitive constructs that serve to render a term intelligible need not on my account be, and often are not, literal representations of the terms referent. Rather, in my view such constructs serve as models or even analogies with which we may better appreciate theories or explanations, in which the terms figure, and their implications. The thesis defended in this regard is that *terms are intelligible if and only if they can be associated with observationally based representations that figure in a broader network of such representations whose relations are suitably constrained by the theory*. This account of intelligibility is elaborated as part of a more sophisticated model of a terms cognitive
significance—the other dimension consisting, as mentioned previously, of a terms articulation.

After developing this account I contrast it with two more or less recent accounts of intelligibility. The first of these alternatives is the verificationist principle proposed by logical positivists in the early part of the twentieth century. The second account is a more recent alternative developed by Henk De regt and Denis Dieks. Briefly, I argue that my account better explains the historical cases than either of these alternatives. Additionally, I argue that my account is preferable on independent grounds. With respect to the former I argue that the intelligibility of theoretical terms cannot simply be associated with the observational statements that they are logically connected with. Talk of valency properties and attractive forces could readily be associated with observational statements. Insofar as these terms are meant to refer to some material property of things, specifying their effects and the circumstances of their occurrence is not sufficient to afford someone a cognitive grasp of the property itself. De regt and Dieks model essentially takes the intelligibility of a theory to be a function of its having characteristic consequences. However, I point out that even unintelligible theories may have this feature.
Chapter I

1.0.0. Background

Shortly after Newton published the Principia in 1687, Leibniz wrote to Huygens to express his dissatisfaction with Newton’s appeal to attraction:

[regarding Newton] I do not understand how he conceives of gravity, that is, attraction. It seems that, according to him, it is only a certain incorporeal and inexplicable power, whereas you explained it very plausibly by the laws of mechanics.\(^8\)

Despite his confusion Leibniz went on in the letter to concede that Newton’s system was predictive of certain celestial phenomena.\(^9\) Still, throughout the last decade of the 17\(^{th}\) century Leibniz continued to disagree with “certain important mathematicians” who accepted the notions of void and attraction into their systems. Yet, as Alexandre Koyré has noted, it was not until he wrote the *New Essays* between 1703 & 1705 that Leibniz’s dissatisfaction with the notion of attraction began to reach the tenor it would come to have in his *Correspondence* with Clarke circa 1715—1716.

Prior to 1703 Leibniz gives little hint of the nature of his dissatisfaction with attraction beyond suggesting that it might be incorporeal, inexplicable and non-mechanical. The language of “occult qualities” and “miracles” is, however, entirely


\(^9\) Leibniz was not alone in maintaining that Newton’s system was predictive. Alexandre Koyré suggests that this is generally true of Cartesians; Koyré, Alexandre, (1965). *Newtonian Studies*, (Cambridge: Harvard University Press), p. 56.
absent from his remarks on Newton prior to 1703. Also absent is any specific demarcation of ‘attraction’ as unintelligible. But by the time of the New Essays Leibniz had begun to take a stronger position claiming that, “with the aid of ‘what God can do’, we give too much leeway to bad philosophy by admitting these ‘centripetal powers’ and ‘immediate attractions’ at a distance without being able to make them intelligible.”  

Moreover, Leibniz maintains, the appeal to attraction, “amounts to a return to qualities which are occult and, what is more, inexplicable,” adding that while, “we must not deny what we do not understand…we are entitled to deny (within the natural order at least) whatever is absolutely unintelligible and inexplicable.”

However, due to Locke’s death in 1704 Leibniz chose not to publish the New Essays and they were not made public until 1765. Nonetheless, Leibniz continued to push his criticism of Newton both in letters to Hartsoeker between 1710-1712, and, more notably, in the Correspondence with Clarke published in 1717—shortly after Leibniz’s death. Indeed, the most synoptic statement of his criticism of attraction is provided in the fifth paper to Clarke:

118. I objected, that an attraction, properly so called, or in the scholastic sense, would be an operation at a distance, without any means intervening. The author [Clarke] answers here, that an attraction without any means intervening, would be indeed a contradiction. Very well! But then what does he mean, when he will have the sun to attract the globe of the earth through an empty space? Is it God himself that performs it? But this would

---

11 Ibid, p. 65.
be a miracle, if ever there was any. This would surely exceed the powers of creatures.

119. Or, are perhaps some immaterial substances, or some spiritual rays, or some accident without a substance, or some kind of *species intentionalis*, or some other I know not what, the means by which this is pretended to be performed? Of which sort of things, the author seems to have still a good stock of things in his head without explaining himself sufficiently.

120. That means of communication (says he) is invisible, intangible, not mechanical. He might as well have added, inexplicable, unintelligible, precarious, groundless, and unexampled.

121. But it is regular, (says the author,) it is constant, and consequently natural. I answer; it cannot be regular, without being reasonable; nor natural, unless it can be explained by the natures of creatures.

122. If the means, which causes an attraction properly so called, be constant, and at the same time inexplicable by the powers of creatures, and yet be true; it must be a perpetual miracle: and if it is not a miraculous, it is false. ‘Tis a chimerical thing, a scholastic occult quality.12

In fact, Leibniz potentially raises several criticisms here. Are the two charges a) that attraction is simply an occult quality or b) that it’s miraculous, simply alternative ways of expressing the same point or are they different, though perhaps related, criticisms?

---

Furthermore, how are they connected with Leibniz’s charge that attraction is unintelligible? What is it to say that it’s inexplicable or unnatural, and so on?

Certainly, there is room to disagree here. Koyré, for instance suggests that the issue as it is taken up in the Correspondence is over whether ‘attraction’ is merely an occult quality and miraculous. Of Cartesians more generally, Koyré characterizes them as rejecting, “outright Newtonian attraction, in which they persisted in seeing immediate action-at-a-distance, that is, an occult quality, or, even worse, magic or a miracle….” Yet, Leibniz’s appeal to ‘occult qualities’ in the above passage seems to suggest, pace Koyré, that he views it to be more serious than the charge that it is miraculous. In the former case attraction would be false: a mere chimera. Indeed, Leibniz seems to have had something closer to this second reading in mind when he remarked to the Abbé Conti in a letter written late in 1715 that, “If every body is heavy, it follows…that gravity will be a scholastic occult quality or else the effect of a miracle.”

On the other hand Koyré makes little out of the charge that attraction may simply be unintelligible or inexplicable. If we are entitled to take Leibniz at his word, then we can’t simply assume that the claim that Newton is appealing to an occult quality is tantamount to the claim that attraction is inexplicable for as I noted previously Leibniz clearly suggests that the latter is something still worse: “qualities which are occult and,” he says, “what is more, inexplicable.” I want to suggest all of these are part of a deeper concern for Leibniz and that this is a concern over whether Newton had succeeded in explaining the phenomena that he had purported to explain. Thus, Leibniz’s various

---

13 Koyré (1965), pp. 139-148.
14 Ibid, p. 57 (italics added).
claims concerning attraction; that it is either miraculous, occult, or unintelligible, or inexplicable, are different ways in which it might fail to be adequate as an explanatory concept. Moreover, some of these appeared, to Leibniz, to constitute more serious failures than others. If we want, then, to understand Leibniz’s individual criticisms of the Newtonian notion of attraction, and more specifically in the end the claim that it is unintelligible, then we need to begin by first understanding what Leibniz thinks is required of an adequate explanation.

1.1.0. Leibniz’s Metaphysical Naturalism

Leibniz repeatedly stresses that adequate explanations must account for phenomena via the nature of things. Consider, for instance, passage 121 from his fifth paper to Clarke cited above: What is natural is what can be explained by the natures of creatures. Leibniz expresses a similar point in the letter to the Abbé Conti mentioned previously: “It is not sufficient to say that: God has made such a law of Nature, therefore the thing is natural. It is necessary that the law should be capable of being fulfilled by the nature of created things”. These last statements made by Leibniz are indeed not statements of an isolated or esoteric stance taken up late in life. In “A Specimen of Dynamics”—a paper published in the Acta Eruditorium in 1695 and constituting a mature statement of his physics—Leibniz rather emphatically maintains the belief that, “there is no natural truth in things whose explanation [ratio] ought to be sought directly from divine action or will, but that God has always endowed things themselves with something from which all of their predicates are to be explained (italics added).”16 Directly following this he asserts:

even though I admit an active and, so to speak, vital principle superior to material notions everywhere in bodies, I do not agree with *Henry More* and other gentlemen distinguished in piety and ability, who use an Archaeus (unintelligible to me) or hylarchic principle even for dealing with the phenomena, as if not everything in nature can be explained mechanically, and as if those who try to explain everything mechanically are thought to eliminate incorporeal things, not without the suspicion of impiety, or as if it were necessary, with Aristotle, to attach intelligences to the rotating spheres, or as if one ought to say that the elements rise or fall by virtue of form, a concise, but useless doctrine.\(^\text{17}\)

Leibniz’s later statements in the fifth paper and in the letter to the Abbé Conti reiterate the view he expresses here. And the view expressed here represents a clear statement of what I will call Leibniz’s *Naturalism*. That view is, straightforwardly, that the activity of physical bodies is to be explained naturalistically or through the intrinsic nature of the bodies.

This, of course, doesn’t tell us much unless we know something of “the nature of created things.” For Leibniz, this consists of “shape and motion.” However, Leibniz rejected the Cartesian concept of quantity of motion (size x speed) and in its stead utilized a notion of force, together with shape or extension, in order to explain natural phenomena. In particular Leibniz finds a notion of force necessary in order to render motion intelligible. As Daniel Garber has noted,

\(^{17}\) Ibid, p. 126.
For Leibniz all real properties of things in the world reside in genuine individuals. If that is the case, it is evident why mere motion, the mere change of place, just won’t do; for it to be intelligible that there is motion at all, there must be something nonrelativistic, something that is an absolute and nonarbitrary property of some individual thing, that is the cause and ground of motion.\textsuperscript{18}

*Force*, for Leibniz, is this underlying cause of motion. It follows, moreover, that motion itself is not a basic property of things if as Leibniz suggests it must be explained by something further; namely, force. Yet, as Garber goes on to suggest, it becomes an important question for Leibniz to distinguish the individuals in which forces reside. It is true enough that they are in Leibniz’s system properties and that they are indeed basic properties. But for Leibniz, as Garber notes, properties must reside in a substance of some sort. And here there are only two alternatives: either force is a basic property of corporeal extended substance, or it is a property of a substance that is incorporeal. Leibniz’s answer seems to be that force is a basic property of both kinds of substance. He announces in a letter to Des Billettes written in 1696 that he “believes,”

That everything really happens mechanically in nature, and can be explained by efficient causes, but that, at the same time, every thing also takes place morally, so to speak, and can be explained by final causes. These two kingdoms, the moral one of minds and souls and the mechanical one of bodies, penetrate each other and are of perfect accord

through the agency of the Author of things, who is at the same time the first efficient cause and the last end.\(^{19}\)

At this point, it is worth considering matters more carefully. As I have just presented it, Leibniz construes corporeal substance to be that which immediately underlies bodies or matter and consequently to be that of which the properties of (derivative) force and extension are properly predicated. On the other hand, I have at least implicitly suggested that Leibniz contrasts this with incorporeal substance, with soul or mind. Garber appears to see it a bit differently:

Leibniz holds that there are living things everywhere in his world. All body is grounded in corporeal substances that are composed of a unity of primitive active and primitive passive force. Or, since Leibniz identified the primitive active force with form or soul, and the primitive passive force with matter or body, these corporeal substances that ground the physical world can be thought of as Aristotelian substances, unities of form and matter….\(^{20}\)

For Garber then, Leibnizian corporeal substance includes soul. Yet, it is difficult to see clearly how this fits with Leibniz’s remark to Des Billettes, quoted above, to the effect that body and soul constitute two separate kingdoms or his claim in “Against Barbaric Physics” that “all organic bodies in Nature are animated, but neither souls nor bodies change one another’s laws.”\(^{21}\) It seems rather, insofar as Leibniz’s physics is concerned

\(^{19}\) Ibid, p. 331.
the interaction of bodies is what we aim to explain through developing “the principles of mechanism” and that both extension and force are basic properties of matter.

1.1.1. A Tale of Two Kingdoms

The problem is that Garber’s picture runs together Leibniz’s two “kingdoms.” Leibniz makes several distinctions in an effort to clarify what he has in mind. He first distinguishes between the physical doctrine and the doctrine of monads. It is this distinction he aims at with his comments in “Against Barbaric Physics,” as well as with his remarks to Des Billettes. That is, between the “kingdom of efficient causes” or what I shall call the “kingdom of means” and the “kingdom of ends” respectively. The former kingdom consists of physical objects though these are not for Leibniz ultimately real. The latter kingdom, on the other hand, consists of what is ultimately real (i.e. monads or corporeal substance). This is summarized in figure 1.1.0.

<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Principles</th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>Causal-Mechanical</td>
<td>Natural Bodies</td>
</tr>
<tr>
<td>Ends</td>
<td>Metaphysical</td>
<td>Monads</td>
</tr>
</tbody>
</table>

The distinction is crucial both for Leibniz himself, as well as for our understanding of his system. That this is a crucial distinction follows from a number of considerations. First, Leibniz clearly intends with this distinction to exclude, “all influx
of body into soul or of soul into body." Moreover, Leibniz repeatedly stresses that everything in souls takes place through appetite and perception. This then supports the view that monads are the proper subjects of the kingdom of ends, and that, on the other hand, bodies are the proper subjects of the kingdom of means. Second, it is essential to Leibniz’s system that “extension” is not a predicate of any monad or aggregation of monads nor could it be since monads are introduced post facto by Leibniz in his effort to explain extension, of which shape is but a mode. Third, given that shape is but a mode of, or modification of, extended matter Leibniz views ‘extension’ as a mere abstraction or modal entity: “extension is merely something modal like number and time”. It is as such an idealization. But such idealizations are not real or actual. They are as it were abstractions of phenomenal qualities. In Leibniz’s system monads constitute the only real substances:

Accurately speaking, however, matter is not composed of these constitutive unities but results from them, since matter or extended mass is nothing but a phenomenon grounded in things, like a rainbow or the mock-sun, and all reality belongs only to unities. Phenomena can therefore always be divided into lesser phenomena which could be observed by other, more subtle, animals and we can never arrive at smallest phenomena. Substantial unities are not parts but foundations of phenomena.

---

22 Ibid.
With this passage we can begin to see a bit more clearly what Garber means when he suggests that for Leibniz “all body is grounded in corporeal substance”. Insofar as there is any substance, and for Leibniz there must be, corporeal or otherwise, it is indivisible. And Leibniz identifies this substantial unity with monads, which, as he makes clear in the *Monadology* (sec. 19), are more than merely analogous to souls—rather he reserves the term soul for the dominant monad (esp. secs. 61-62 & 70). Thus, the only real substances are those constituted by monads. Leibniz’s argument here is easy enough to see (perhaps because he repeats it so frequently): Substance must be simple or indivisible. There must be substance for properties to adhere in. Yet this substance cannot be a physical substance as this latter is infinitely divisible. Therefore it is composed of a more fundamental unity that grounds the physical phenomena.

It is clear then that Leibniz directly equates substance with the monads of his metaphysics. But at least two questions remain concerning Garber’s reconstruction. First, in what sense are monads “grounding” physical phenomena given that the two kingdoms operate “by their own laws” and independently of one another? Second, how can body in this sense be a constituted of monads if bodies operate by their own laws. The answer to the latter question is that Leibniz uses the term bodies equivocally, or perhaps what is preferable, analogously, when discussing the two kingdoms. That is, if we wish to provide an explanation of phenomena in accordance with Leibniz’s physics we will not appeal to form or soul together with a body in the sense in which it is part of what constitutes the unextended unity of a monad. Rather, we appeal to abstract notion of a shaped thing. This latter is of course not a real entity. If instead we wish to pursue the

---
ultimate explanation of the world we will necessarily be pursuing explanations in the kingdom of ends. I take it then that the sense in which such explanations “ground” all body is, from Leibniz’s point of view, the sense in which all contingent facts are grounded by metaphysical necessity. 25 Put another way as Leibniz sees it, it is one thing to ask how a certain phenomenon or type of phenomena is brought about, or what the physical laws are that govern the phenomena. It is quite another to ask why God should have chosen those particular laws. In responding to the latter question we will of course be providing reasons that are as such the ultimate reasons but that will nonetheless play no role in the mechanical explanations themselves. On this view Garber’s explication is better suited to an account of Leibniz’s metaphysics than it is to an account of his physics. As to the first question, I will postpone discussion of it for it now.

Of course, this is a reconstruction of Leibniz’s system is incomplete. It doesn’t yet make clear the relation between Leibniz’s metaphysics and his physics. To that end we might be inclined to agree with Russell when, in one of his characteristically polemical moments, he claims,

The relation of Leibniz’s Dynamics to his Metaphysics is hopelessly confused… Leibniz has acquired much credit for the vaunted interconnection of his views in these two departments, and few seem to have perceived how false his boast really is. As a matter of fact, the want of connection is… one of the weakest points in his system. 26

The present model does, however, go somewhat further than what might be expected from Russell’s pessimistic attitude. Russell’s concern was with Leibniz’s attempt to

---

25 See esp. Leibniz (1714), secs. 33-45.
26 Russell (1900), p. 89.
reduce ‘motion’ which was a relational concept to a metaphysical quality of things—that is, to monadic predicates. His clear implication is that the “interconnection” is undeservedly vaunted. I contend that the connection is more one of providing rational foundations than of causal reductions. For, as Leibniz continually reminds us, the two kingdoms can have no influence on one another’s laws. As we suggested in our response to the question of how corporeal substance might ground physical phenomena the interconnection could not be causal in either direction.

Garber’s claim is that there is a deeper level that is the level at which corporeal substances are constituted for Leibniz. At this level, corporeal substances are constituted by form and “primary” matter. These are then, of course, Leibnizian monads that form the basis for Leibniz’s metaphysical explanation of the principles of mechanics. I think we can agree with this while nonetheless separating Leibnizian physical explanations from metaphysical ones. Monads play no role themselves in mechanical explanations of natural phenomena, belonging as they do to a separate kingdom, and it is mechanical explanations that Leibniz thinks we should appeal to in constructing our physics.

This picture doesn’t yet capture all of the important features of Leibniz’s mechanism. True, Leibniz was a mechanist as I have repeatedly stressed but he adopted a mechanism different from that of Descartes and his exponents. Unlike Descartes who identified matter with extension, Leibniz added force to matter as an inherent property of it and moreover as that property which is, in matter, responsible for its motion. Like Descartes, however, Leibniz viewed impulse or contact to be the only means by which forces are altered or modified. It is in this sense that Leibniz adopts a mechanical view of nature. Consequently, when in a late and unpublished essay Leibniz claims, “that
everything in corporeal nature must be explained mechanically,” he is claiming not only that all natural phenomena must be explained in terms of their extension and force, but also that these explanations must provide the contiguous causal antecedents. But, as I have already suggested above, this is not the end of the story as, from Leibniz’s point of view there is something more to be explained; namely, we still need to recognize, “the true metaphysical principles or the explanations of motion and laws of nature that derive from them.”

In the “Specimen” he states this view a bit more clearly: “we acknowledge that all corporeal phenomena can be derived from efficient and mechanical causes, but we understand that these very mechanical laws as a whole are derived from higher reasons.” Thus, when we offer explanations of what we observe in nature we can account for the phenomena in terms of their constituents. That is, we can do so in terms of the shape and motion (or extension and force) of their constituents. If we consider the way in which Leibniz himself goes about his physical explanations as for example in the “Specimen”, we won’t see him appealing to metaphysical principles. Rather, he adopts a causal-mechanical picture of the phenomena and sets aside questions of why those particular causal laws are operative the world as a question of reason and metaphysics.

It is in this sense that Leibniz adopts a naturalistic attitude toward explanations in physics while at the same time deriving his naturalism from his metaphysics. For Leibniz, then, there are two levels of explanation. First, the level at which we offer explanations of natural phenomena. At this level we are confined to providing laws—Leibniz here has in mind mathematical laws—governing the shape and motion of objects. “Indeed,” says

---

27 Leibniz, (1710-16).
Leibniz, “in this way, sensible effects have been explained by causes which fall within the senses, something most conducive to practical applications and to imitating and improving upon nature.” It is, however, also important to recognize that Leibniz is strongly committed to the view that the nature of the physical world, the corporeal realm, is contingent. Why “created things” have the natures that they do, and why they are ordered according to whatever laws they are, is not empirically accessible. In order to understand why the physical universe is the way it is we need to move to a second, perhaps deeper, level of explanation.

It is only at this further level that we can begin to understand the reasons for the contingent structure of nature. It is at this level that Leibniz appeals to the various metaphysical principles that buttress his natural philosophy (a rather prominent example is the principle of sufficient reason). God, being, as he is, the author of nature, could have, should he have so chosen, written the book of nature otherwise. But he didn’t!

The principle distinction for the purposes of understanding Leibniz’s criticism of Newton is the distinction between the “kingdom of ends” and the “kingdom of means”. The coexistence of these two kingdoms is essential to an appreciation of the criticism. It helps to explain one of the main features of the criticism that I shall elaborate below. For Leibniz, to say that the explanation of phenomena should be given in terms of the natures of bodies is tantamount to saying that they require a mechanistic explanation. As I see it this idea, that it is in the nature of bodies to operate by mechanical principles, is a contingent fact for which there is a metaphysical explanation. I take naturalism and mechanism to be coextensive and therefore only contingently identical. Moreover, the coexistence of these two kingdoms helps to explain why Leibniz was pushing for a
naturalistic explanation of gravity and the corresponding focus on a mechanical account of it, as well as why we should think of this criticism as simply synonymous with that of its being “miraculous.” Miracles are violations of the mechanistic laws by which bodies naturally operate; yet it is not true that phenomena are never to be explained in terms of them.

1.2.0. Intelligibility, Miracles, & Occult Qualities

With the relationship between Leibniz’s physics and metaphysics in view we can begin to better appreciate what in general he is claiming when he claims that a given term is unintelligible or that it designates an occult quality or that by appealing to it in our explanations is “to have recourse to a miracle.” Understanding the character of these criticisms is important for two reasons. First, and most immediately, it will allow us to distinguish them as distinct claims against the Newtonian appeal to attraction. Second, and perhaps more important for the larger project, it will allow us to begin to develop an understanding of the so-called pre-Newtonian notion of intelligibility—a notion left unarticulated by proponents of the view that this episode constitutes an example of shifting standards of intelligibility.29 Consequently, addressing this latter project constitutes a necessary preliminary for the argument I will offer in chapter VI concerning the shifting standards thesis. In this section then I will be looking at each of these criticisms in greater detail beginning with the notion of intelligibility.

As I have already pointed out, some commentators, like Koyré, have maintained that a notion’s being unintelligible is simply synonymous with the claim that it designates

29 The idea that standards of intelligibility shift over time will be taken up more thoroughly in the final chapter when I discuss the criterion of intelligibility offered recently by Henk De regt and Denis Dieks.
an occult quality. Others go still further claiming that it is equally synonymous with being non-mechanical, mysterious, etc. This, I will argue obscures the actual character of Leibniz’s various criticisms of attraction. It is, for instance, simply not true that Leibniz held that whatever is non-mechanical is unintelligible. Rather, as we shall see, the question of intelligibility with which Leibniz was concerned pertained to the cognitive significance of ‘attraction’.

Larry Laudan has claimed that the charges that “Action at distance is occult, unintelligible, mysterious, and unmechanical” are “All more or less synonymous.” I claim they are even more than less synonymous. A number of these notions are not in the least synonymous for Leibniz. Laudan’s only evidence to the contrary comes from a few lines extracted from the long passage cited in section 1.0.0 above:

That these terms were seen as more or less identical in meaning is confirmed by that famous passage in the Leibniz-Clarke Correspondence where Leibniz chides Clarke (and, by implication, Newton) as follows: “That means of communication (says he) is invisible, intangible, not mechanical. He might as well have added, inexplicable, unintelligible, precarious, groundless, and unexampled...‘Tis a chimerical thing, a scholastic occult quality.”

Laudan goes on to say that Leibniz “was not saying that he did not know the meaning of action-at-a-distance.” The charge against Newton, according to Laudan, was not meant to highlight empirical difficulties in Newton’s explanation of gravitational phenomena (e.g.

---

failure to predict, etc.) but intellectual and conceptual difficulties involved in the notion. These conceptual difficulties resulted from Leibniz’s inability to assimilate the notion of action-at-a-distance into the ontological framework of his mechanical philosophy.

While it is true that Leibniz found attraction troubling because it was not a mechanical notion or one that fit readily with his mechanical philosophy (indeed, it did not fit at all), this was not simply equivalent to the charge that it was unintelligible. As I have already noted, Leibniz did not view every non-mechanical notion to be unintelligible. To take an obvious example, neither souls nor minds operate by mechanical principles and they are not, in Leibniz’s view, thereby rendered unintelligible. To claim that attraction was unintelligible was indeed to claim that we have no idea of it whatsoever and thus, pace Laudan, to question the meaning or conceptual significance of attraction.

Shortly after the passage from the Correspondence cited above, Leibniz goes on to clarify what he means by the notion of intelligibility. He charges that to admit attraction is to deny the principle of sufficient reason, and as well to accept that there are intelligible expressions that are neither true nor false. Rather than being merely a confused idea as occult qualities are or even contradictory as some occult qualities are or merely non-mechanical as are the notions that comprise his metaphysical doctrine, ‘attraction’ fails to have any significance under this criticism. It is, I argue, not even on par with the concept of the fastest motion, a concept Leibniz claims we understand though we have no idea of it. We have no idea of it precisely because the expression involves a contradiction and is hence impossible. Newtonian attraction on the other hand

---

was not even susceptible of being contradictory. If I am right then Leibniz’s criticism of attraction amounted to a many-horned dilemma. Yet, in order to understand each horn we must first appreciate Leibniz’s basic epistemological framework.32

At the most fundamental level, thinking involves entertaining notions that we may grasp either clearly or obscurely. We grasp a notion obscurely when it is not sufficient for us to recognize its referent. Leibniz seems to have had two kinds of examples in mind here. On the one hand he considers empirical cases of obscure notions wherein we remember having seen some individual but in which the notion that we have is insufficient for us to distinguish that individual from others when we encounter them subsequently. The idea I have of the orangutan I saw at the zoo, insofar as I am unable to pick it out from a number of others of its kind, is a notion of this type. The second sort of case Leibniz has in mind involves having infelicitous or vague definitions. In his view Aristotle’s notion of a cause (insofar as it is whatever is common to each of Aristotle’s four types) is just such a case.

Clear notions may be either distinct or confused. Either way clear notions are sufficient to allow us to recognize the individuals or properties they represent. Such notions are however confused when we cannot give explicit criteria sufficient for distinguishing their objects. The examples Leibniz has in mind here include items like our notion of green insofar as we are able to recognize green things but cannot say what it

32 My discussion of Leibniz’s epistemological framework is largely pulled from his 1684 Acta publication: “Meditations on Knowledge Truth and Ideas.” It is an essay to which Leibniz continually refers back to with approval throughout his later writings. He does so for good reason. It is the most complete and detailed account of his views on this subject—views that he did not relinquish. My aim here is primarily that of summarizing his position. See; Leibniz G. W., (1684), “Meditations on Knowledge Truth and Ideas,” reprinted in Philosophical Essays. trans. & edited by Roger Ariew & Daniel Garber, (Indianapolis: Hacket), 1989, pp. 23-28.
is in virtue of which they are green. In general, all of our particular notions from sensation, including tastes, smells, tactile sensations, and so on, will be confused insofar as we lack any definition of them. It is worth noting, at this point, Leibniz’s distinction between nominal and real definitions. The example he most frequently chooses to illustrate the concept of a nominal definition is that of the assayers notion of gold as a heavy, golden substance soluble in aqua fortis. Yet, more generally he understands by such a definition one that defines a complex notion in terms of its component notions. Real definitions on the other hand establish the possibility of thing being defined, which is to say in a rough and ready way, suitable for our purposes, that the notion does not involve any contradictions.

Whereas confused notions lack even nominal definitions, distinct notions are either themselves primitive notions or else nominally defined. This, however, is not the end of the line for Leibniz since even clear and distinct notions may turn out to be false. This follows from the fact that not every nominal definition is provided in terms of the notions primitive constituent notions. Some definitions, like that employed by the assayer, are sufficient for distinguishing their objects even though they involve notions that are as well complex. Distinct notions are thus divided into those that are adequate and those that are not. Inadequate notions are like that of the assayers notion of gold, whereas with adequate notions the nominal definitions include all and only those primitive notions that comprise the notion being defined.

A final subsidiary distinction is required in order to fully appreciate Leibniz’s basic epistemic framework. It is, however, an important one. Leibniz maintains that we

---

33 I.e. Nitric Acid.
often omit the ideas in thinking and proceed with mere signs of them. This, in Leibniz’s view, for example, characterizes most of our arithmetical and geometrical thinking, especially when the objects of those thoughts are the products of elaborate chains of reasoning. The result of such reasoning Leibniz dubs symbolic or suppositive knowledge. He contrasted this with the notion of intuitive knowledge. This latter in his view is perfect as it consists in reducing notions to their primitive constituents and as such does not involve any contradiction—at least, any latent contradiction. More importantly, it is at this level that one considers the primitive notions or ideas themselves. As Leibniz remarks in the Meditations, “We don’t perceive ideas of even those things we know distinctly, unless we make use of intuitive thinking.” That is, we gain something more exacting by the intuitions—namely, the idea.

Were it not the case that intuitions afford us some deeper epistemic privilege, there would be no non-normative reason\(^\text{34}\) that would exclude manifest contradictions from our most perfect knowledge. In Leibniz’s ideal epistemic situation we would clearly (and distinctly) be able to know that P, even if P involved manifestly contradictory constituent notions.\(^\text{35}\) Charity alone, suggest that we ought to make an epistemic distinction between ideas, and mere “notions” or “understanding” in general, on Leibniz’s behalf. Fortunately, the claim, that ‘ideas’ possess an epistemic significance over and above ‘notions’ for Leibniz, does not rest solely on charity. Leibniz explicitly marks a distinction between ‘understanding’ an expression and ‘possessing the idea’

---

\(^{34}\) Of course here I have in mind the fact that Leibniz does not appeal to standards of rationality but defines knowledge (scientia) rather in terms of the character of the notions and ideas themselves.

\(^{35}\) What would of course be explicitly ruled out, in virtue of the fact that in the ideal Leibnizian epistemic situation we consider only primitive notions, is the possibility of any latent contradiction in our knowledge.
therein putatively expressed. As he says in the *Meditations*, “at first glance we might seem to have the idea of a fastest motion, for certainly we understand what we say; but yet we certainly have no idea of impossible things.”\footnote{Leibniz (1684), p. 25.}

For Leibniz then we *perceive an idea P* iff the notions comprising P are understood intuitively and they are mutually consistent, whereas *we understand what we say* provided the notions we have are mutually consistent and understood (i.e. minimally, the notions are clear.) The distinction between perceiving the idea and understanding what we say is an important one for Leibniz’s epistemology. We only have perfect knowledge when we perceive the ideas. Yet we have all possible ideas regardless of our perception of them. To claim that an idea is false is at once to assert that we do not possess the idea but not to say that we do not understand what we say or that we do not have the constituent ideas in a manner that is clear. The principle distinctions (for my purposes) are captured in figure 1.2 below.

With this framework in place we can more clearly distinguish the prongs of Leibniz’s dilemma. Occult qualities are of two sorts for Leibniz. In the context of the above they are either merely the sensible qualities of bodies for which we can give no explanation (i.e. render distinct), or they are qualities offered in explanation of sensible qualities though they are either merely obscure or else they designate false ideas (i.e. involve contradictory notions.) The former are confused, whereas the later are again either obscure or clear and distinct though contradictory. Leibniz provides illustrative examples of each. In fact, he maintains that all particular sensible qualities of things are

\footnote{Leibniz (1684), p. 25.}
occult because they are confused. But such occult qualities are not necessarily false. We could in principle remove their occult character if we were able to render them distinct. This is just to say that their true causes are unknown while, nonetheless we have the notion of them and they are therefore possible. It is however not this sense of occult that is most pertinent to our discussion. It is useful only insofar as it sets up a contrast with non-sensible occult qualities.

More important for our purposes are those qualities that are not sensible as the intellect or pure understanding apprehends them, though perhaps with aid of our senses. It is in virtue of these notions that we attempt to explain and thereby understand the sensible qualities of bodies. In general, for Leibniz, such explanations will be causal explanations. Occult qualities in this class of notions are those that are contradictory. Examples include ‘the fastest motion’, ‘atom’, ‘void’, and notably ‘action-at-a-distance’. To illustrate the concept of an atom is in Leibniz’s view occult because it signifies a false idea as its notion is contradictory. Both the notion of extension and that of indivisibility are part of the concept of an atom. It is important to note that it is not that we fail to understand such notions. Indeed, it is because we do understand these notions that we are able to determine that they lead to contradictions and thereby signify false ideas.

---

It ought to be clear from this discussion that there is then a sense in which Laudan is right to maintain that Leibniz found ‘attraction’ to be comprehensible (understandable). If ‘attraction’ is simply identical with the notion of action-at-a-distance then it is clear enough though it involves a contradiction. The idea of action, which Leibniz believes we have, is understood straightforwardly enough in terms of mechanical forces (specifically, contact action) and is incompatible with having distant effects. Action is a principle of
mechanics. The notion of action-at-a-distance can be understood but it is impossible and hence false. Leibniz does not pursue the issue, for instance, with Clarke because the latter philosopher concedes as much. But this is only one horn of the tri-lemma that Leibniz poses for ‘attraction’ which we can now state more clearly.

Recall that for Leibniz every assertion is true, false, or unintelligible. The claim that ‘attraction’ is false follows from considering it to be simply defined as action-at-a-distance which is a notion that is, as we have seen, contradictory. I am claiming that this is something different from the contention that ‘attraction’ is unintelligible. Ultimately, the tri-lemma posed for attraction is that it is either a miracle or it is occult, or it is simply unintelligible. The first horn is I believe, expressed rather clearly in Leibniz’s assertion to Clarke in paragraph 122 of his fifth paper-- *If the means, which causes an attraction properly so called, be constant, and at the same time inexplicable by the powers of creatures, and yet be true; it must be a perpetual miracle: and if it is not a miraculous.*

To consider action in this light is not to consider it as a mechanical principle but as a principle by which God may act on bodies. Such a notion is not incompatible with anything in the notion of God since, for Leibniz, God does in fact perform miracles. It is therefore not a false notion. Yet such an activity exceeds the powers of our understanding. We know its possibility only through our inadequate understanding of God.

What remains then is to distinguish the claim that ‘attraction’ is unintelligible from the charge that it is occult. The claim that Leibniz is making when he asserts attraction to be unintelligible I maintain, *pace* Laudan, is precisely the claim that we have

---

39 Quoted above.
no understanding of it whatsoever. It is at best an insignificant term that in Leibniz’s view makes merely a vague allusion to some common cause of gravitational phenomena. As he sees it, if attraction is allowed to be explanatory then there is no reason not to believe that scholastic ‘faculties’ or ‘intentional species’ explain the phenomena. The problem with these notions is not that they fail to be sensible—they aren’t supposed to be—but that they can’t even be clearly distinguished by the intellect and it is for this reason that they are unintelligible. The distinction between what is sensible and what is intelligible is for Leibniz a simple matter.

Merely sensible notions are those particular notions we receive from the senses and that we entertain only in a confused manner. It is when we render them distinct that we make them intelligible.\(^{40}\) In other words, every distinct notion is intelligible. It follows from this, together with the fact that there are occult concepts that are clear and distinct, that even occult concepts may be intelligible. There are worse things in the conceptual realm than being occult. Whereas occult notions fail to signify or denote any real, or even possible, quality of a substance and thereby signify occult qualities, unintelligible notions fail to exist. They reduce to mere utterances that we have no understanding of. Thus, in criticizing Locke for his acceptance of the notion of attraction, Leibniz writes,

He is captious about the operations of souls when it is merely a matter of admitting what is not sensible, and here he is granting to bodies what is not even intelligible, granting them power and activities which in my

\(^{40}\) Though, of course not every distinct notion is sensible. It is the imagination or intellect through which come to have ideas and at least begin to understand for Leibniz.
opinion transcend anything that a created mind could do or understand, for he grants attraction to them.  

Leibniz is not, of course, arguing that everything that we do not in fact understand is unintelligible in principle. There are many things, that we do not understand in fact, but that we might in principle intelligibly explain. We ought not deny the former he claims, but “we are entitled to deny (within the natural order at least) whatever is absolutely unintelligible and inexplicable.”  

We can here begin to see the role that naturalism and mechanism played in Leibniz’s critique of attraction.

It is, according to Leibniz, a necessary condition of a notions being natural that it may, at least in principle be distinctly understood. With respect to bodies, those that are natural are precisely those that that can be explained in terms of shape and motion. These latter notions are the right ones for him because, as he sees it, they are the notions necessary to any distinct knowledge of bodies—i.e. they are included, Leibniz thinks, in our notion of a body. The parenthetical qualification in the above passage is intended to highlight the fact that miracles are by definition supernatural occurrences. They are of course mysterious from our point view in that we have no distinct knowledge of them and as such they are inexplicable by mechanical notions. But not every inexplicable hypothesized notion is miraculous. Insofar as attraction is supposed a natural quality of bodies, not simply identified with ‘action-at-a-distance’—this latter again being admittedly contradictory and therefore occult—it is inconceivable or inexplicable as it is

---

41 Leibniz, (1765), p. 61.
42 Ibid, p. 65.
43 See, Leibniz, (1765), p. 66: “…whereas what is natural must be such as could become distinctly conceivable by anyone admitted into the secrets of things.”
not even a candidate for being distinct. Leibniz summarizes this point nicely in the *New Essays*:

This distinction between what is natural and explicable and what is miraculous and inexplicable removes all the difficulties. To reject it would be to uphold something worse than occult qualities, and thereby to renounce philosophy and reason, giving refuge to ignorance and laziness by means of an irrational system which maintains not only that there are qualities which we do not understand—of which there are only too many—but further that there are some which could not be comprehended by the greatest intellect if God gave it every possible opportunity, i.e. [qualities] which are either miraculous or without rhyme or reason.\(^\text{44}\)

1.3.0. Conclusion

In sum, Laudan is right to claim that ‘action-at-a-distance’ is occult, but he wrongly identifies this with being unintelligible. To the extent that ‘attraction’ is an occult concept it is synonymous with the clear and distinct, though inadequate, notion of a force exerted at a location distant from the body exerting the force. “Action-at-a-distance” as a notion signifies a false idea. It is not incomprehensible or miraculous or unintelligible. It is not miraculous because it is impossible. It is not unintelligible because it is distinct. Rather it is simply contradictory. On the other hand Leibniz did not rest his critique of attraction merely on the possibility that his opponents might construe it as “Action-at-a-distance.” In fact he acknowledges that Clarke may have some other view of attraction, indeed he

\(^{44}\) Leibniz, (1765), p. 66.
must have given his concession that action-at-a-distance would have been impossible. Leibniz thus proceeded to question what alternative understanding that we might have of the notion. The alternatives, as Leibniz understood them were two-fold: either attractions had to be understood as the product of the divine intervention of God in the natural order of things, or else the hypothesized quality was simply inconceivable.

I have attempted to motivate the claim that this analysis of the structure of Leibniz’s criticism fits more comfortably with his various remarks concerning ‘attraction’. I have also tried to show how it conforms to Leibniz’s epistemology as well as his conceptual ontology. Moreover, I have suggested a number of ways in which views like Laudan’s, views that treat the horns of the tri-lemma as equivalent, are problematic in light of Leibniz’s views on the respective concepts.
Chapter II

2.0.0. Introduction

As we have seen, Leibniz’s claim that “attraction” is unintelligible stemmed from a concern for conceptual clarity. If “attraction” was to be construed neither as a miracle nor as simply synonymous with action-at-a-distance, then we have no cognitively significant notion corresponding to the term. From Leibniz’s point of view, Newton and his exponents had failed to specify any suitable idea with which the term could be associated.

As far as it goes this analysis remains incomplete. We would like to know what it is for a term to be intelligible. What we have stated so far merely pushes the question back a step to the point where we now must ask what is involved in a term being associated with a suitable idea. Given Newton’s belief that he had provided an adequate explanation of gravitational phenomena, and Leibniz’s admission that the explanation was predictive, it is reasonable to ask Leibniz for a more precise accounting of the putative failure. On what grounds could Leibniz maintain that Newton had not secured an idea? And under what conditions is an idea deemed to be a suitable one? Hadn’t Newton with his putative attraction succeeded, after all, in explaining not only the motions of celestial bodies but heavy bodies generally? And why should Newton care if his theory failed to meet some esoteric standard that Leibniz was attempting to impose on him?

These questions will be addressed in the present chapter. To that end I have three principle objectives. I argue that Leibniz was not appealing to an esoteric standard but rather that his appeal to a conception of intelligibility in terms of the association between words and ideas was broadly accepted during the period and constituted what we might
call a common minimal standard or core conception of intelligibility. Of course, this so-called core-conception was but a skeletal structure to be fleshed out by an account of the origins of ideas. Thus after demonstrating the widespread acceptance of this core conception I will argue that the claim that Newton’s “attractive force” was unintelligible stemmed from a tendency to view “attractive force” as a referential term that went beyond the bare gravitational phenomena, and that viewed in this light the term failed to be intelligible regardless of how the framework of the core-conception was filled out. Finally, and most importantly, in the closing sections I will draw out some “morals” for the development of a contemporary standard of intelligibility.

2.1.0. Sine Mente Soni

For seventeenth and eighteenth century theorists ‘ideas’ were the objects on which the understanding operated. Where philosophers did differ was on the issue of the origins of our ideas. Whereas individuals such as Descartes and Leibniz held that many of our ideas were purely intellectual and consequently independent of the senses, others like Locke maintained that all of our ideas are at root derived from experience or sensation.

While these were not insignificant differences it is important to recognize that there was a significant consensus concerning the core nature of intelligibility. It’s true that the core criterion will differ extensionally in the hands of different theorists. Theorist A may well find some term, say, “mental substratum” intelligible while theorist B views it to be no more than an empty shell of a word and hence an unintelligible locution. Either way their dispute comes down to the question of whether or not the term signifies an idea. Whereas theorist A purports to have some idea, theorist B disputes this claim. This
is precisely the sentiment Locke expresses to Stillingfleet when he claims that, “the new way of ideas, and the old way of speaking intelligibly, was always, and ever will be the same… that a man use no words but such as he makes the signs of certain determined objects of his mind.”

Locke’s statement is perhaps as explicit a formulation as any of the sort of minimal standard or core conception (CC) of intelligibility noted above. Thus we might formulate the account in the following manner:

**CC**: A term is intelligible for s iff it is the sign of a certain determined object in the mind of s.

It is clear that Locke at least thought CC reflected a common sentiment of his day. And, moreover, that the desired mental objects were ideas. Certainly, the general attitude toward the relation between words and ideas is well-noted. Hobbes (1588-1679) had written in the *Leviathan* that, “The general use of speech is to transfer our mental discourse, into verbal; or the train of our thoughts, into a train of words.” And it is according to Hobbes an abuse of language, “when men register their thoughts wrong, by the inconstancy of the signification of their words; by which they register for their conceptions, that which they never conceive.” Arnauld maintained that, “we express nothing by our words when we understand what we are saying unless, by the same token,

---

45 Locke, John, (16??). Add citation.
it were certain that we had in us the idea of the thing we were signifying by our words.”

All of these remarks reflect the commitment to something on the order of CC.

A word of caution is however in order. It way well be true, as some philosophers have maintained, that Locke construed something on the order of CC to constitute a theory of meaning. But that is in itself a matter separate from the question of whether or not it was widely accepted as an account of intelligibility or perhaps even a test for meaningfulness as opposed to a general theory of meaning. Whatever the case may be for Locke, it is not true that CC codifies a generally accepted theory of meaning. Arnauld, for instance, is clear in the passage quoted above and elsewhere that words signify the things themselves and not our ideas of them. Hobbes, on the other hand, while he, like Locke, talks about words signifying or being the signs of ideas, explains the notion of a

---


49 See for instance Loeb, Louis E., (2001). “Hume’s Explanations of Meaningless Beliefs,” *The Philosophical Quarterly*, Vol. 51, No. 203: April, pp. 145-164. Loeb claims that Lockes theory of meaning is constituted by two theses “first, the meaning of a word, term or expression in its primary or immediate signification is the idea for which it stands; and second, all ideas are ultimately derived from simple ideas, those received passively in experience.”

50 The distinction between “meaningfulness” and “meaning” that I have in mind is admittedly not obvious. For the most part I intend by it to separate, what I am inclined to think of as, external tests for the presence of the property of meaning from accounts of the property itself. Thus, to use an analogy, one might have an adequate test for determining whether a sewer is backed up independent of whether they have an account of the nature of the stoppage. Still, it is an open question whether accounts of meaningfulness are satisfactory accounts of meaning. While I believe that there is a coherent distinction to be made here, I do not expect the distinction to be entirely transparent here.

Equally, it is worth noting at this point that a number of terms that are employed throughout the present chapter will be refined and modified as I develop my account in the next two chapters. ‘Significance’, for instance, in the context of the present chapter is, from the point of view of my model, a vague descriptive property—variously referring to the object designated or the associated idea. This ambiguity is, I believe, partly responsible for misunderstandings of the early modern concern with language and will be cleared up in the chapters III & IV.
sign on analogy with the manner in which various artifacts may be said to signify the presence of some object. Thus tombstones in a cemetery generally signify the presence of a corpse and so on. Ian Hacking has argued that this is evidence that Hobbes did not possess a theory of meaning, that he was instead merely highlighting the customary connection between words and ideas.51

While Hacking may be correct in asserting that Hobbes did not possess a theory of meaning, the customary association of words and ideas is too weak to capture Hobbes’ view concerning the relationship between language and thought. For Hobbes, at the very least, words are meaningful only insofar as they truly signify the presence of some conception.52 In my view the relevant sense of sign in CC is closer to that of the role of the term in the language being functionally identical to the role of the concept or idea in thought. Both the word and the idea may equally refer to the same external object and hence it is not necessary to view the idea as providing all or part of the meaning. Nonetheless, the notion that terms might designate objects that could not be represented in thought was viewed as an absurdity. The ability to render a term intelligible, to translate it into the language of thought was meant as a check against uttering nonsense. It is this sentiment that I take to be expressed in the statements above and codified by CC. That said, it is clear that CC stands in need of some revision. Whatever the notion of ‘being the sign of’ might come to, and regardless of the relationship or connections between this expression and the “meaning” of a term, CC is too specific as a “minimal”

standard. How, we might ask, are we to determine which ideas are the relevant ones? A potential answer suggested by Arnauld’s comments is that the right ideas are those that are representative of the things signified by our words. But this answer is not as straightforward as it may at first seem. What are we to do with talk of “odors” or “tastes” generally, or to take more familiar contemporary examples, what are the requisite ideas involved in the use of terms like ‘water’ or ‘temperature’. It is not inconceivable that someone might maintain that it would have been impossible for Arnauld and his contemporaries to have had any idea of the actual signification of these latter terms. Must we then judge in hindsight that by their standards these terms were unintelligible at the time and that it was only later when we came into possession of the relevant signification of these terms that they became intelligible?53

This sort of response is certainly on the table at the outset and, as we shall see further on, some recent theorists have been tempted by it. Nonetheless, I find the response implausible on its face for a variety of reasons. First, the occurrences of this sort of linguistic phenomena obviously won’t be restricted to a small number of isolated cases. Wherever and whenever the supposed true signification of our words has to do with a subject matter presently inaccessible to the mind, anyone committed to CC in the sense under question must regard these locutions as unintelligible. Thus, we necessarily commit ourselves to the view that a rather large segment of language which seventeenth and eighteenth century philosophers widely believed to be intelligible was at base nonsense viewed from this perspective.

53 Naturally I have something on the order of Hilary Putnam’s externalism in mind here.
A more severe consequence, in my view, is that if the suggestion taken from Arnauld captures the relevant sense of what it is to have the appropriate or right idea then much of what we now might say with respect to, or in an effort to explain, various phenomena may well turn out in future hindsight to be deemed unintelligible when judged from the perspective of CC. Of course one might reasonably rejoin that this is all well and fine; all this shows is that CC does not capture the relevant relata. Instead, the rejoinder continues, we could avoid the problem if we looked externally. An accurate characterization of intelligibility has not to do with associations between words and ideas or any other mental objects but with an association between words and the physical and/or metaphysical structure of the universe most closely associated with our production of the words.

The temptation to externalize intelligibility in this way has severe consequences. Insofar as the micro-structure of the universe remains at least partly closed to us we ought to be skeptical of the idea that we ever know what we mean by a given utterance. Instead, insofar as we know anything, we will be perpetually in a situation of speaking (to borrow a phrase from Locke) of “things of which we know not what.” At this extreme historical criticisms concerning intelligibility are reduced to vacant incomprehensible utterances. Leibniz’s claim that Newton’s “attractive force” was unintelligible in truth comes to nothing more than the claim that Newton had somehow failed to succeed in referring to any coherent and distinguishable feature of the world. Yet, how could Leibniz, being in the epistemic position suggested by the rejoinder ever be in a position to level such a criticism. That is, how could Leibniz ever be in a position to know that Newton had failed
to use his terms in a manner such that they stood in some coherent relation to the hidden physical/metaphysical structure of the universe.

The problem is straightforward. If there is anything to be gleaned from CC it is that we should not take it to codify a theory of meaning and hence conflate a theory of meaning with an account of intelligibility. Arnauld’s comments are consistent with the externalization of meaning but not with the externalization of intelligibility. For individuals committed to the theory of ideas to the degree that early modern philosophers were, intelligibility was inextricably linked with mental contents and could not, in contrast with meaning, be externalized. It does not follow that ‘meaning’ and ‘intelligibility’ did not in some sense come to the same thing for the early modern theorists. But meaning in our contemporary sense was not the central concern. If some philosophers then believed that the meaning of a term was given by its associated idea this was because the terms themselves were merely placeholders for ideas. These latter entities being the proper objects on which the understanding operates.

Thus, Hacking is correct in his claim that the theory of ideas did not constitute a theory of meaning but rather stood on its own ground as a response to questions that are not explicitly within the target domain of the latter theory. Along the lines of Hacking’s suggestion I maintain that intelligibility had to do primarily with the nature of thought and the role of language in thought rather than with a theory of the meaning of words or of how words are used to communicate. One need not have in mind any of the principle concerns of the latter theories in asking what the relationship of language to thought is.
Again, Arnauld’s comments are entirely consistent with externalizing meanings but not similarly consistent with externalizing intelligibility.54

This isn’t to say that some early moderns didn’t run the two together or even that the two are not in fact related. It is, for instance, not as clear that Locke did not think that the meaning of a term was given by the idea that rendered it intelligible. Still we have shown that there is no necessary connection between intelligibility and meaning and that the principle concern of a theory of intelligibility was with the connection between language and thought.

But the response suggested by Arnauld’s remark won’t do as a response to our initial question. Understood in this manner CC falls short. It falls short because many of the terms that were found at the time to be clearly intelligible could not be associated with an idea representative of the thing signified. For example, given that the relation under question is that of “being representative of the thing signified by”, the expression ‘the cause of ρ’ will be unintelligible just in case we fail to cognitively apprehend ρ’s cause. Consequently, standing in such a relation, doesn’t capture the relationship relevant for the determination of the expressions intelligibility.

Claims that are straightforwardly observational in nature do not pose any special problem. If we simply want to discuss the phenomena then the relevant ideas can be ostensively provided. But if, on the other hand, we want to discuss or talk intelligibly about the causal mechanisms underlying the phenomena, ostensive appeals will not

54 Of course it is equally consistent with internalist accounts of meaning. I take this to be, at the very least, implicit in the discussion here the principle idea being that an account of intelligibility is compatible with both internalist and externalist accounts of meaning though not compatible with externalizing intelligibility itself.
always be possible. What is needed is some account of the abstract and unobservable designatum. This problem with abstract terms is however not insurmountable.

Leibniz takes up the problem in the *Theodicy* in the context of discussing the conformity of faith with reason. One of the central worries there concerns how, if at all, we understand mysteries such as the trinity and to that end whether or not we must be able to fully comprehend them in order to have any understanding of them. His remarks on this topic are instructive. He maintains that rather than being understood in virtue of our having the idea of their external signification, “the Mysteries receive a necessary verbal explanation, to the end that the terms employed be not *sine mente soni*, words signifying nothing.”55 It is not necessary that we fully comprehend the mystery itself. To fully comprehend a phenomenon, even a putative one such as the trinity, “it is not enough that one have some ideas thereof; one must have all of the ideas of everything that goes to make it up, and all of these ideas must be clear, distinct, adequate.”56 But Leibniz claims that, “It will be possible for a truth to be incomprehensible, but never so far as to justify the statement that one comprehends nothing at all therein.”57 Were we to truly lack any conception of say the trinity, even that is a limited and incomplete one, then the expression itself would be nothing more than a word “devoid of meaning” for Leibniz.

While it is true that Leibniz is here primarily concerned with our faith, and consequently our ability to believe, in religious doctrines like the trinity, the goodness of God, and God’s justice and benevolence, it is equally clear that he viewed this to be a

57 Ibid
general feature of the understanding even when applied to the truths of physics or natural truths. “Incomprehensibility,” he maintains,

Does not prevent us from believing even in natural truths. For instance…we do not comprehend the nature of odours and savours, and yet we are persuaded, by a kind of faith which we owe to the evidence of our senses, that these perceptible qualities are founded upon the nature of things and that they are not illusions.\(^{58}\)

Moreover,

one must not always demand what I call ‘adequate notions’, involving nothing that is not explained, since even perceptible qualities, like heat, light, sweetness, cannot give us such notions. Thus we agreed the mysteries should receive an explanation, but this explanation is imperfect. It suffices for us to have some analogical understanding of a Mystery such as the trinity and the incarnation, to the end that in accepting them we pronounce not words altogether devoid of meaning.\(^{59}\)

It is not, I think, difficult to see that in Leibniz’s view when we talk about mysteries or phenomena that we do not fully comprehend (that is phenomena whose mechanisms are not fully understood or, to spin it in more Leibnizian language, the notion of which we possess only inadequately) it is sufficient that we possess some ideas that we can associate with the terms of the discussion. Otherwise, not only is it the case that we do not comprehend the phenomena in question (i.e. know the “how” of it) but also we do not

\(^{58}\) Ibid, p. 97.

\(^{59}\) Ibid, p. 103.
understand what it is that we are talking about; our terms become as it were *sine mente soni*.

The account of intelligibility offered here constitutes a significant advance over the account suggested in the previous discussion of Arnauld’s remarks. It avoids the problem that arises for anyone who would make the intelligibility of terms depend upon their being wedded to ideas that are representative of their true significations. *A fortiori* it becomes even clearer in the context of Leibniz’s remarks that an internalist account of intelligibility (the only game in town) is consistent with an externalization of meaning but certainly does not demand it. A terms meaning, that is, may perhaps be connected with its
cognitive significance but that by no means need be taken to suggest that our ideas exhaust the terms significance.\(^{60}\) For this latter to be the case, on the Leibnizian model we would have to possess an adequate notion of the terms signification. Yet, unlike God’s understanding, this is generally not the state of our understanding.\(^ {61}\)

\(^{60}\) See footnote 40. Again the idea is that ‘significance’ is a potentially problematic notion that will come to have a more specific meaning in the model I will be proposing in subsequent chapters.

\(^{61}\) Leibniz, (1684).
On the picture that I am advancing a term is intelligible just in case it can be associated with a cognitive representation of the object (or, more broadly the object, event, or process) that it designates. This picture is captured in figure 2.1.

The account here is fairly straightforward. ‘Water’ is intelligible when we have a distinct idea of it—e.g. a colorless, odorless, heavy, solvent…. That is, in Leibniz’s terminology, the term is intelligible when we have a nominal definition of it. Not all of the ideas contained in our idea of water will themselves be distinct but they must be collectively sufficient to distinguish water from other substances. In this manner we have an inadequate though perhaps incomplete yet distinct notion of water sufficient to render our talk of water intelligible. The answer to our question concerning which ideas are the appropriate ones seems to be that the appropriate ones are those sufficient to render our the notion distinct.62

Two points have emerged in this discussion and suggest a provisional interpretation of CC. First, “sign” as it appears in CC should not be taken to mean that the idea gives the meaning of the term. Rather to say that \( \Psi \) is the idea for which “\( \tau \)” is the sign is to say that \( \Psi \) functions as a representation on which cognitive operations may be performed. For the most part such representations will stand on the sidelines, or in the cognitive background so to speak, since the signs themselves suffice as sufficient marks or space holders for the relevant ideas in thought. This last bit is important. For early modern thinkers, when serious questions arise, it is through a consideration of the ideas involved that we decide the issue and not through a consideration of the structure or form

---

62 cf. Peacocke
or truth conditions of the *sentences* that are used regularly to stand in for those ideas. Language’s primary function was to lend economy to thought. Once again Leibniz illustrates this point nicely.

> we don’t usually grasp the entire nature of a thing all at once, especially in a more lengthy analysis, but in place of the things themselves we make use of signs, whose explicit explanation we usually omit for the sake of brevity, knowing or believing that we have the ability to produce them at will… in my mind I use these words (whose sense appears only obscurely and imperfectly to the mind) in place of the ideas I have of these things… I usually call such thinking… blind or symbolic.\(^{63}\)

The second point that has come out thus far is that the ideas need not be complete. It is in fact sufficient that some idea be associated with the term. Given that terms are just stand-ins for ideas the real issue is whether or not there is any idea whatsoever associated with the term in question. In order to determine whether or not a given utterance is true one needs to carry out an analysis of the associated ideas. In the absence of *any* associated ideas it would be impossible to determine the truth or falsity of what was being claimed. Even a false idea will suffice since what is required is simply something that can be the object of analysis. Thus, we find Leibniz claiming that we can understand expressions like ‘the fastest speed’ or ‘a physical atom’ even though their ideas are contradictory and hence false. The idea we have associated with ‘fastest’ he sees as containing the idea of a limit whereas the notion of ‘speed’ has as one of its compound

---

\(^{63}\) Leibniz, (1684). p. 25.
constituents the idea of being limitless. Hence, while there is an idea associated with the expression it contains contradictory notions which render the larger idea false.

When Leibniz wrote in the Correspondence that, “every intelligible enunciation must be either true or false”, 64 he was merely expressing what, in the context of the theory of ideas, was a truism. Any idea whatsoever will be either true or false, but words in the absence of ideas hold the place of no object suitable for the understanding to analyze. Accordingly, CC should be interpreted so as to leave the association illustrated in figure 2.1 fairly loose. I offer the following “rephrasing” of CC in an effort to capture this:

\[ \text{CCR: A term } \tau \text{ is intelligible iff } \tau \text{ is associated with a representation } \Psi \text{ on which the understanding may operate.} \]

\[ \text{CCR} \text{ does not constitute a fully fleshed out criterion of intelligibility, but it does capture the core view with regard to intelligibility shared by most, if not all, early modern philosophers. The differences that remain will have more to do with the nature of } \Psi \text{ than with disagreements over the connection between words and ideas. All theorists of the time believed that words referred only in virtue of their associated ideas. The failure to meet the condition specified in } \text{CCR} \text{ captures this view.} \]

2.2.0. Elaborating the Theory of Ideas

In the case where a term fails to meet the condition specified in CCR, I will say that the term lacks “cognitive content”. The question presently before us is, “In what sense, according to Leibniz, did ‘attraction’ fail to possess cognitive content?” In order to more

---

fully appreciate Leibniz’s concern it may be helpful to impose some further structure on CCR.

Suppose I have some clear idea $\Psi$. As we saw in the previous chapter, for Leibniz to more fully appreciate $\Psi$ the understanding seeks to render $\Psi$ distinct. That is we attempt to understand our ideas by decomposing them into their constituent notions. Suppose then that a, b, and c are all constituents or elements of $\Psi$. Of course any of these constituent ideas may be susceptible of further analysis. Thus “a” might be roughly distinguished by $<a_1, a_2, a_3,\ldots>$ and so on. Along these lines if $\Psi$ is our idea of gold then it might be distinguished by {heavy, gold colored, solubility in aqua fortis, etc.}, and each of these may be further distinguished. To say that an idea is false, on this model, is to say that it has the form {$…a, \neg a,\ldots$}. Obviously the determination that some idea is actually false constitutes an epistemic achievement that can only come about if CCR is satisfied. Thus, when Leibniz claims that the concept of an atom constitutes a false notion he does so because he believes that when we come to understand the notion distinctly we will find that it contains both the notion of an indivisible as well as that of extension or divisibility. That is, the idea of an atom is identified with {...divisible, indivisible...}. Clearly, such a cognitive construct constituted an absurdity but the term ‘atom’ was not thereby rendered unintelligible.

Ideas so construed raise some obvious and interesting questions. On the one hand, on this account conceivability does not imply possibility—we can conceive the impossible. Accordingly, Leibniz holds that, “an idea is true when its notion is possible and false when it includes a contradiction.”\(^{65}\) As this quote illustrates, the very possibility

of having false ideas depends on their impossibility (i.e. the impossibility of the circumstances represented therein). On the other hand, our vulgar or common sense notion of what is impossible often seems to involve the idea that impossible situations—and specifically those which are logically impossible—describe “unthinkable” situations. And there is often a temptation to think of a criterion of intelligibility as being somehow concerned with characterizing what it is that makes some situation or thing “unthinkable.” The Leibnizian analysis of the concept of atom provides a clear counterbalance to this false tendency. For early modern philosophers ideas, say simple ideas, were like homogeneously shaped pieces of a cognitive puzzle. Any of them could be combined with any number of others resulting in a composite idea. And the existence of any such idea was sufficient to guarantee the intelligibility of its associated term—assuming there was some such term.

Still one can’t help but feel that there is something odd about all of this. How, for example, is it possible that I can think of that which is impossible? How can I think, say, of an extended object that is simultaneously located in two discrete spaces—simultaneously both here and there? How can I think of something that is at once both tall and short (given the same relative parameter)? I think that the basic worry here is well-noted and I have just two things to say with regard to it.

First, it is a natural consequence of most versions of the theory of ideas. My present objective isn’t to argue that the theory of ideas as I have been laying it out is something that we all ought to accept, but rather to clarify what I take to be the appropriate context within we can acquire a more accurate understanding of Leibniz’s
criticism that Newton’s ‘attraction’ was unintelligible. To that end the “theory of ideas” is what it was. Ideas, so understood were the primary objects on which the understanding operated, and the understanding was the faculty constituting our epistemic agency. Leibniz was not unique in maintaining that the structure of ideas was compositional in the above sense.

In the first Enquiry, Hume maintains that,

Nothing is more free than the imagination of man: and though it cannot exceed that original stock of ideas furnished by the internal and external senses, it has unlimited power of mixing, compounding, separating, and dividing these ideas, in all of the varieties of fiction and vision.66

The idea here is fairly straightforward. The understanding has an almost limitless power to geometrically manipulate the material with which it is furnished, and thereby it is endowed with the power to construct all manners of fiction—as well as notions which are materially or even logically false. The problem seems to occur when we reflect on the possibility of “mixing” ideas. How, so to speak, can we mix the idea of a divisible with that of an indivisible? And what would the resulting idea “look” like? I take this to be approximately the sort of worry that underwrites worries about thinking of impossible situations or objects. These are perhaps the sort worries Hume had in mind when we wrote in the Treatise.

Ideas may be compared to the extension and solidity of matter…Ideas never admit of a total union, but are endowed with a kind of

---

impenetrability, by which they exclude each other, and are capable of forming a compound by their conjunction, not by their mixture.\textsuperscript{67}

This claim, that the Understanding does not have the power to \textit{mix} ideas in the manner suggested earlier, resolves the pertinent worries. When we think about say \textit{indivisible divisibles} we do not have in mind an undifferentiated union of the two notions. Rather we consider only the simple conjunction of two ideas as opposed to their “total union”. In this sense the degree to which the understanding is free to mix ideas is limited.\textsuperscript{68} On the other hand the account given in the \textit{Treatise} seems, on the face of it, to conflict with that offered in the \textit{Enquiry}. But both accounts support the view that it was possible to have a contradictory idea either as a simple conjunction of two contrary notions or by their “total union.”

The worry (concerning what is conceivable) stems, perhaps, from a tendency to push the question of intelligibility back a step and onto ideas themselves. That is, we attempt to ask, “What conditions must be met for an idea to be intelligible?” There may well be a significant worry here; A worry about the status of a certain class of ideas that we might presently label unthinkable. Early modern philosophers would have simply called these “false ideas.” If the claim is that a certain sub-set of false ideas exhibits certain cognitive traits—e.g. containing contradictory notions—and that other false ideas do not, well and fine. Such discriminations are possible—e.g. on some construal the idea


\textsuperscript{68} That said, I think that there is an issue concerning the consonance of the accounts of the powers of the Understanding provided in the first \textit{Enquiry} and the \textit{Treatise}. It is not clear at present that the two can be given consistent interpretations.
of a unicorn is false but not contradictory. That is, however, not to grant that intelligibility has anything whatsoever to do with such distinctions. When it comes to questions of intelligibility, the pertinent worry is about terms and specifically whether the terms in question are associated with any idea.

2.3.0. ‘Attraction’ vs. Action-at-a-distance

With modern commentators there is a tendency to construe Leibniz’s criticism concerning the intelligibility of ‘attraction’ as a concern with the intelligibility of ‘action-at-a-distance’, and particularly with a concern with the contradictory character of its “notion.” Gerd Buchdahl, for instance, writes, “it was that gravity, regarded as ‘action-at-a-distance,’ seemed unintelligible, impossible to conceive, indeed self-contradictory.”

This quote captures what we might label the received view of the criticism that ‘attraction’ was unintelligible. If I am right, however, it would have been in fact oxymoronic for Leibniz to have claimed that ‘attraction’ was unintelligible and that its idea was self-contradictory. If there was any associated idea the term was not unintelligible and if there was not an associated idea, then it would hardly be sensible to talk of anything that was self-contradictory. It makes no sense on this scheme to talk of terms simpliciter as self-contradictory. And the worry about conceivability confuses what is comprehensible with its real possibility. The mere fact that a notion has no real, or material, possibility does not entail that its idea is inconceivable.

---

69 One has to be careful here differentiate those notions which are merely fictitious from those that are false. For instance, the notion of a unicorn is fictitious, by any account, though false, only if the notion of existence is conjoined with it.

---


71 As noted previously this view is held by others including Lauden, (1970), and Henk De regt and Denis Dieks—the authors of the view I will be considering in chapter VI.
The real culprit here is, I think, the tendency to view the issue as simply one concerning action-at-a-distance. In the remainder of this section I argue that that identification is mistaken and that in fact action-at-a-distance, at the time, was not viewed to be unintelligible. At the very least, it definitely was not viewed to be unintelligible by Leibniz.

At the outset it is worth noting that action-at-a-distance simply wasn’t the issue for Leibniz. Nor was it the issue for Newton and Clarke. It is not insignificant that we do not once find Leibniz claiming that ‘action-at-a-distance’ is unintelligible. The reason for this is simple; Leibniz did not find action-at-a-distance unintelligible. Consequently, he never claimed that it was. On the contrary, Leibniz continually claims that this notion is occult, contradictory and false. When, however, Leibniz was engaged in claiming that ‘attraction’ was unintelligible he was decidedly not equating it with the occult and false notion of action-at-a-distance.

Nor, in fact, did Newton ever construe the issue over the intelligibility of ‘attraction’ to concern either the real possibility or the conceivability of action-at-a-distance. It was, from his point of view, given that action-at-a-distance was absurd and hence did not constitute a real possibility. Consequently whatever ‘attraction’ might be taken to signify, Newton maintained that it should not be conceived of as a quality or power in bodies in virtue of which they act at a distance. Indeed, he warned against making this identification. In a frequently noted letter to Bentley written in 1692 Newton cautions,

That gravity should be innate, inherent and essential to matter, so that one body may act upon another at a distance, without the mediation of
anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking, can ever fall into it.72

Newton is not here confessing that he finds ‘attraction’ unintelligible. Rather he is insisting that Bentley not identify it with action-at-a-distance. And interestingly his reasons for not making this identification are precisely those given by Leibniz; ‘attraction’ is not to be associated with action-at-a-distance because this latter notion is contradictory. On my articulation of the issue the dialectic is easily explicable. As Newton says anyone with “a competent faculty of thinking” would see the absurdity (i.e. the impossibility) contained in such a notion. Consequently, action-at-a-distance is not the appropriate cognitive content with which to associate the term ‘attraction’.

Clarke too had taken pains in the Correspondence to disabuse Princess Caroline and thereby Leibniz of the idea that Newton’s ‘attraction’ might be treated as simply synonymous with action-at-a-distance. As he says toward the end of his fourth reply, “That one body should attract another without any intermediate means, is indeed not a miracle, but a contradiction: for ‘tis supposing something to act where it is not.”73 What this discussion illustrates is that each of these philosophers found ‘action-at-a-distance’ to be associated with a contradictory idea. The distinct idea of an action would undoubtedly involve the idea of a cause and its effect, but this latter meant that the idea of action

73 Alexander, (1956); p. 53.
would involve the notion of contiguity in time and space, Consequently, the notion of an action-at-a-distance = {…spatial-temporal contiguity, spatial-temporal discontiguity,…}. Leibniz’s response was not to insist that ‘attraction’ must be associated with action-at-a-distance despite Clarke and Newton’s claim to the contrary. Instead, as we saw in the chapter I, Leibniz rightly took his opponents to concede the soundness of that horn of his tri-lemma. Upon which he proceeded to question whether or not attraction corresponded to any idea. “Very well!” says Leibniz, “But then what does he [Clarke/Newton] mean” by ‘attraction.’ Leibniz proceeded to raise the possibility that Clarke or even Newton might believe that the hand of God or some other immaterial agent affects gravitational phenomena. I can see no other reasonable interpretation of Leibniz’s dialectic than that he is here motioning toward the other horns of his tri-lemma. That is, since the notion of action-at-a-distance was admitted by both Clarke and Newton to be inconsistent with the mechanistic view of nature (i.e. with the distinct conception of a natural action) and therefore false, ‘attraction’ had to be given some other interpretation.

From where Leibniz stood the only remaining possibility was that ‘attraction’ designated some immaterial cause. And if not that, then it was merely an unintelligible term. In the latter case we simply have no idea corresponding to ‘attraction’. To put it bluntly, ‘attraction’ had no translation in the mental language of ideas. In the former case, on the other hand, we are left with a non-naturalistic explanation of the phenomena. As the hand of God ‘attraction’ was indeed intelligible as a non-natural quality of bodies—at least in principle. We have, so says Leibniz, the clear idea of God; an idea which includes

74 Ibid, p. 94.
his ability to intervene in the course of nature. This latter interpretation of ‘attraction’ was viewed as less than desirable by both sides.

Leibniz for his part found the “attraction as immaterial cause” interpretation objectionable primarily because the notion committed one to the occasionalism he had already rejected. Newton, for his part, was somewhat more ambivalent concerning this horn of the tri-lemma. As we shall see more fully later while Newton did not close off the possibility that God might ultimately be the proximal cause of gravitational phenomena, he clearly intended the claims he had made for attraction to be empirically justified. Clarke too, was reluctant to ascribe a non-natural interpretation to “attraction.” Instead he sought to open up the notion of what might count as natural to include non-mechanical qualities writing in the correspondence,

the means by which two bodies attract each other, may be invisible and intangible, and of a different nature from mechanism; and yet, acting regularly and constantly, may well be called natural; being much less wonderful than animal-motion, which yet is never called a miracle.\(^\text{75}\)

Clarke’s response is interesting for a variety of reasons not the least of which is the fact that it highlights Leibniz’s view (discussed in the previous chapter) that there is a hard and fast distinction between the kingdom of efficient causes or what is proper to natural philosophy, and the doctrine of the monads or what constitutes the ultimate \textit{metaphysical} truth of the matter. Leibniz had maintained that it was the mechanical view that is constitutive of the former and therein necessary to any naturalistic explanation of phenomena. As Clarke makes clear here, he is not inclined to accept any such hard and

\(^{75}\) Ibid, p. 53.
fast account of what is natural. Rather he argues that the miracle interpretation does not exhaust the truth horn of Leibniz’s tri-lemma. As he goes on to say,

If the word natural forces, means here mechanical; then all animals, and even men, are as mere machines as a clock. But if the word does not mean, mechanical forces; then gravitation may be effected by regular and natural powers, though they be not mechanical.\(^{76}\)

It’s not surprising that this issue frustrated Leibniz. On the one hand Clarke sought to claim that attraction was a \textit{natural} force. Given this, our idea of it insofar as it contains the idea of cause and effect must contain the idea of spatial-temporal contiguity. At the same time Clarke maintains that it is a \textit{non-mechanical} power to produce motions analogous, in some sense, with “animal motions.” But then what is this non-mechanical motion other than the sort of action-at-a-distance that both Clarke and Newton had conceded to be false. Clarke’s response is to appeal to the \textit{regularity} of the phenomena to be explained by the supposed force. But then what other notion do we have but that of the phenomena that were to be explained? It seemed that Clarke was maintaining that we have some notion of $\Psi$ associated with attraction and that $\Psi \neq \{\ldots$ mechanical activity,\ldots$\}$, and $\Psi \neq \{\ldots$,supernatural activity\ldots$\}$, and $\Psi \neq \{\ldots$action-at-a-distance$\}$. How else could ‘attraction’ be understood save as $\Psi = \emptyset$?

The main point I want to focus on presently is that Clarke is here decidedly not interested in even hinting at a defense of action-at-a-distance. He has already explicitly indicated that such an interpretation is not even on the table. Rather, Clarke’s strategy is to argue that ‘attraction’ nonetheless designates a natural non-mechanical property of

\(^{76}\) Ibid.
bodies and is consequently a true quality of bodies; hence an intelligible one. Clarke’s response equally illustrates that intelligibility was not equivalent to mechanical comprehensibility. The fact that action-at-a-distance was inconsistent with the mechanical philosophy of the time is not what accounted for the view that ‘attraction’ was unintelligible. Indeed, one way around the claim that ‘attraction’ was unintelligible would have been to insist, as Clarke seems almost tempted to, that spatial contiguity is not a necessary constituent of our notion of cause or motion. This however begins to look too much like action-at-a-distance—so much so that neither Newton nor Clarke pursued this possibility.

There are at least two important implications of this discussion. First, Newton and his early followers did not see themselves to be engaged in arguing that action-at-a-distance cohered with the mechanical philosophy, nor did they view the issue to revolve around whether it was intelligible. Newton and Clarke, as well as Leibniz, viewed it to be intelligible but false. Thus, what was required to meet the charge that ‘attraction’ was unintelligible was some other account of the notion signified by it.

The second implication is that simply coming to accept attraction qua ‘action-at-a-distance’ as part of ones conceptual repertoire would not, contrary to what some contemporary philosopher’s have maintained, have constituted a shift in the standards of intelligibility for the simple reason that neither Leibniz nor his interlocutors found the notion of an action-at-a-distance to be unintelligible.\textsuperscript{77} The denial that this notion was false on the other hand would indeed have constituted a genuine response to Leibniz’s tri-lemma. Yet, it would have been a response that appealed directly to the then prevailing

\textsuperscript{77} Again the claim that standards of intelligibility shifted during this period will be taken up in more detail in chapter VI.
standard of intelligibility. And such a denial would require an argument. Specifically, it would require an argument effectively demonstrating that the notion of action-at-a-distance was indeed non-contradictory.\textsuperscript{78}

2.4.0. Buchdahl and the Role of Intelligibility in Theory Assessment

Few contemporary commentators have laid much stress on the role that the intelligibility of terms has played in scientific disputes. Gerd Buchdahl is a notable exception.

\begin{table}
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Dimension} & \textbf{Content} \\
\hline
Constitutive & Systematic integration of laws \\
& Data/Empirical Adequacy \\
\hline
Explicative & Meaning \\
\hline
Regulative & Maxims for Theoretical adequacy (e.g. simplicity) \\
& Metaphysical Biases \\
\hline
\end{tabular}
\caption{Figure 2.2}
\end{table}

Buchdahl argues that in order to understand historical controversies like the debate over Newton’s mechanics, it is not sufficient to merely examine the systemization of the theory (e.g. the formulation and integration of a system of laws), or the maxims presupposed in the development of the theory (e.g. the principle of sufficient causation). Instead, he distinguishes three dimensions or components considered in the assessment and development of theories. These dimensions are summarized in figure 2.2 and discussed below.

Along one dimension, that which he labels the constitutive component, theories are evaluated and understood in terms of how the laws are formulated and “integrated into a system”. This involves not only the empirical data used in the generation of the theoretical laws but the integration of the laws into a systematic whole as well as the predictions offered by the theory. As we saw at the outset Leibniz had conceded that Newton’s theory was predictive. To that extent he was allowing that Newton’s theory was acceptable on constitutive grounds. Moreover, he was not for instance claiming that Newton’s mathematical generalizations were themselves somehow incoherent or that taken together they failed to preserve the pertinent phenomena. Rather, if Leibniz’s criticisms can be situated in Buchdahl’s scheme then it will be along one of the other two dimensions.

The second dimension noted by Buchdahl consists of the regulative maxims appealed to in support of the theory. Principles like uniformity, simplicity, and the like all fall under this heading as well as preferred explanation schemes on the order of the causal-mechanical model appealed to by many of Newton’s critics. Additionally, this component contains those metaphysical biases or “laws” that might serve to ground the
theory; principles like Leibniz’s principle of sufficient reason, or maxims such as, ‘nothing comes from nothing’, etc. Buchdahl labels this the *architectonic* or, better I think, the *regulative* component of theory evaluation.

Of course Buchdahl does not think that these components are independent of one another. We can already see for example that the two dimensions articulated thus far overlap. What some philosopher’s have suggested, that unification is an explanatory goal, may or may not be generally accurate as a description of what have historically passed for good explanations. It is however plausible as an account, say, of the relation of Newton’s dynamics to Kepler’s laws of motion, and to that extent while the value expressed falls under the regulative component, the unification occurs as a product of the constitutive dimension. Moreover, Buchdahl recognizes that regulative maxims may come to form part of the theory.⁷⁹ That Buchdahl sees these two as connected is readily apparent. Elsewhere, he subsumes aspects of the constitutive dimension and regulative maxims under the general heading of a theory’s *rationality* distinguishing them from further features of the constitutive component which, for their part, contribute to the theories *probative strength*—that is, its inductive support and predictive value.⁸⁰

Insofar as the rationality of a theory is concerned with the coherence of the theory as well as the regulative standards which constrain the rational acceptability of a theory, Leibniz’s criticisms clearly related to assessing the rationality of Newton’s theory. In particular given that the mechanical view of nature functioned at the regulative level and

---


had built into it a certain conception of matter in terms of shape and motion, it functioned to constrain what could be rationally accepted as an explanation of natural phenomena. Indeed, it was the mechanistic philosophy that prevented any rational commitment to action-at-a-distance precisely because within this context such a notion was clearly contradictory. Still, given that the participants all agreed that this was not on the table, we could hardly expect to gain any comprehensive understanding of the controversy if it is simply understood to involve a worry about the theories rationality.

Instead Buchdahl maintains that historically the evaluation of theories often involves as well what he terms the *explicative* component.\(^{81}\) He associates this component with a concern over the intelligibility and meaning of the theoretical terms employed by the theory. He argues that the received view of the factors determining the acceptability of a theory has typically overlooked the fact that concerns over the comprehensibility of key terms have historically played an important role in the debate over a theory’s acceptability. He specifically focuses on illustrating the role that the concern with the intelligibility of ‘attraction’ played in the assessment of Newton’s mechanics.

The “explicative” component is, as the label might suggest to the careful eye of the astute reader, concerned with explicating the meaning of the theoretical terms or hypothetical concepts instanced by the theory. This he explicitly associates with questions concerning the theories intelligibility:

we shall also say that while the architectonic [or *regulative*] component determines the “rationale” of the theory, or of a theoretical concept,

\(^{81}\) Buchdahl, (1970b).
explication is concerned with “intelligibility”. Of course, the meanings of many terms entering into many theories antedate their formulation; still, and especially in a developing science, such meanings often undergo considerable change.82

Examples of the sort of terms Buchdahl has in mind are not difficult to come by. The “elements” of the peripatetics, the spagyrical chemists’ ‘principles’, the scholastics’ ‘faculties’, the seventeenth century notion of a ‘body’ or ‘matter’ are all good examples of the kinds of terms for which an explication might be called for. Indeed, the need for an explication of the first two of these is, as we shall see in chapter IV, part of what is at issue in Boyle’s criticisms of the attempts made by some of his contemporaries to explain various chemical phenomena. More important for our present purposes, it is the explicative dimension on which Leibniz was focusing with criticism that ‘attraction’ was unintelligible.

Buchdahl sees the issue of intelligibility, in his terminology, as simply synonymous with his explicative component. At least, this is so in the Criteria. Oddly, however, “explication” does not occur at all in Gravity. Rather, there he associates intelligibility most closely with comprehensibility; a notion he views to be “split up into (i)…‘possibility’ and (ii)…‘intelligibility’.”83 The distinction he has in mind here, between possibility and intelligibility, is, however, not obvious. It is complicated by the fact that as he proceeds he does so in a manner that suggests that he views comprehensibility, intelligibility, and conceivability to be all at least roughly synonymous. At the outset, however, we should note in fairness to Buchdahl, that it is

part of his immediate goal to trace the development of these various notions and especially that of “comprehensibility.”

Buchdahl begins by drawing a distinction between primary and secondary explanations. A primary explanation is one that provides the causal mechanism responsible for the production of a given phenomenon. A secondary explanation, on the other hand, explains the phenomenon via appeal to the activities of auxiliary agents. What he aims to show is that while Newton continually treated attraction in the *Principia* as a causally efficacious property of bodies—speaking of it as, “a ‘manifest property,’ as an ‘active principle,’ that ‘really exists,’”—Newton avoided specifying the nature of this “attractive” quality of bodies. Indeed in the *Principia* he makes no attempt to treat of the mechanics of gravitation. Only later in response to the criticisms of individuals like Huygens and Leibniz do we find Newton responding.

In the *Optics* Newton attempts to respond to the charge that “attraction” is unintelligible. The response here is to suggest that gravitational phenomena can be accounted for in standard mechanical terms involving only contact actions or impulse forces. To this end Newton makes a renewed appeal to the æther model of gravitational phenomena that he had proposed some years earlier in a letter to Boyle:

> I shall set down one conjecture more... It is about the cause of gravity. For this end I will suppose an æther to consist of parts differing from one another in subtilty by indefinite degrees: that in the pores of bodies there is less of the grosser æther, in proportion to the finer, than in open spaces; and consequently, that in the great body of the earth there is much less of

---

84 ibid, p. 81, n. 10.
85 ibid, p. 82.
the grosser æther, in proportion to the finer, than in the regions of the air:
and that the grosser æther in the air effects the upper regions of the earth.
and that the finer æther in the earth the lower regions of the air, in such a
manner, that from the top of the air to the surface of the earth, and again
from the surface of the earth to the center thereof, the æther is insensibly
finer and finer. Imagine now any body suspended in the air, or lying on the
earth; and the æther being by the hypothesis grosser in the pores, which
are in the upper parts of the body than in those which are in its lower parts
and that grosser æther being less apt to be lodged in those pores than the
finer æther below, it will endeavour to get out and give way to the finer
æther below, which cannot be without the bodies descending to make
room for it to go out into.86

The æther model indeed provided a mechanical explanation of gravitational phenomena
but it was to say the least not one that offered an explanation of the phenomena in terms
of ‘attraction’ as he had done in the *Principia*. It was rather an attempt to provide a
secondary explanation. And as Buchdahl notes,

> By constantly seeking escape from the problem of intelligibility by
> postulating a secondary explanation in terms of impulse action, Newton
did not seem to face the issue of intelligibility of gravity [attraction] as
> such, that is, of the causal implications of his language, as already hinted
> before. He sought, as we have already seen, to avoid this issue, by

attempting to “explain away” gravity through reference to the action of other physical agents.⁸⁷

It is interesting that Buchdahl describes the situation in the manner that he does. What after all, for Newton’s part, might constitute a more square response to the charge of unintelligibility than to illustrate that attraction could at least in principle be explicated mechanistically? Wasn’t Newton’s response clearly in line with the charge? Buchdahl thinks not—and I am inclined to agree with him. According to Buchdahl, in offering the æther model, Newton effectively abandoned the primary explanation of gravitational phenomena in terms of ‘attraction,’ turning instead to a secondary mechanical explanation. In doing so he conflated the “intelligibility” of ‘attraction’ with its “rationality.” Newton hadn’t rendered ‘attraction’ intelligible. Rather, he had provided an alternative mechanical explanation of the phenomena, entirely independent of the explanation offered in the *Principia*; entirely independent of any appeal to a quality of attraction in bodies.

Newton evidently didn’t appreciate this feature of his response. He viewed the æther model as a speculative account of attraction along mechanical lines. But, as a speculative model, his commitment to it was rather tenuous. As he says in the 1693 letter to Bentley already quoted from above, “Gravity must be caused by an agent acting constantly according to certain laws; but whether this Agent be material or immaterial, I have left to the consideration of my readers.”⁸⁸ Newton for his part seems to have viewed the appeal to ‘attraction’ as licensed because he could demonstrate its mechanical possibility. In Buchdahl’s view Newton had made the mistake of taking the question of

⁸⁷ Buchdahl, (1970a), p. 82.
intelligibility to be simply one concerning its possible mechanical rationale instead of a question concerning the comprehensibility of ‘attraction’ itself. Beyond his appeal to the sorts of secondary explanations implicated by the æther model and the occasionalism inculcated in the previous quote, Newton made little attempt to apologize for his use of ‘attraction’ believing that the law of universal gravitation was itself inductively established and therefore that the appeal to attraction was valid. Given that Leibniz’s concern was with the explicative dimension of Newton’s theory it is not surprising that he found this response unsatisfactory. To appeal to the inductive validity of the law was an appeal to the constitutive grounds of empirical adequacy.

The question raised by Newton’s critics was not whether or not it was possible to provide a mechanical model for gravitational phenomena. Indeed, Leibniz had explicitly recognized the possibility of such a model in a letter to Huygens claiming that he did not understand Newton’s meaning though he viewed Huygens to have, “explained it [the phenomenon of gravity] very plausibly by the laws of mechanics.” To reiterate, the Leibnizian tri-lemma was a challenge to the contention that we have some notion of a quality of attraction, not to the mechanical explicable of gravitational phenomena. Either it is true that bodies possess an attractive quality, in which case that quality must be placed into them and maintained by God since it cannot arise as an “explicable modification” of the shape and motion of bodies; or it is simply conceived of as a power in bodies by which they act at a distance, in which case it is false since it is contrary to the nature of a body; or, lastly, the quality of attraction is inconceivable in which case it is meaningless to inquire into how it might be explained in terms of the nature of bodies.

89 Leibniz, (1690), p. 309.
To be clear; I am not saying that it would not have been acceptable to reduce ‘attraction’ to a mechanical account. The feasibility of that project is a separate issue. Carrying through the project, however, would have been to give ‘attraction’ a sense other than that with which Leibniz was concerned. The sense of ‘attraction’ with which Leibniz was concerned was the sense in which he putatively designated an inconceivable quality of bodies.

2.4.1 Evaluating Buchdahl

Buchdahl’s three-dimensional model provides a useful way of situating the role played by CCR in assessing theories. It illustrates in a relatively clear manner the distinct issues involved when worries concerning the intelligibility of key theoretical terms are raised. Indeed I will be appealing to this model later to demonstrate the manner in which some recently proposed criteria of intelligibility have confused what is at issue. To that end in the next chapter I propose a structure for “cognitive significance” that parallels Buchdahl’s model of theory assessment.

That said, while I agree with Buchdahl’s claim that the explicative dimension of his model has been largely neglected by philosophers of science—to say nothing of philosophers generally—I believe that Buchdahl is guilty of a few missteps himself.

I take it that the most interesting feature of Buchdahl’s model lies in its incorporation of the explicative dimension. In my view he presents a compelling argument for the importance of including such a dimension if we want to understand the various ways in which scientific theories have been assessed historically. Obviously the case for this model rests to a large extent on its ability to explain historical controversies like the debate over ‘attraction’. Given this Buchdahl’s case would be undercut to some
extent should it turn out that there really wasn’t any broad acceptance of CCR, or for that matter if it turned out that the importance of the explicative component was not widely appreciated. Of course I don’t believe that either of these conditions obtained. I have already argued that CCR was broadly accepted. I argue, pace Buchdahl, that the import of his explicative component was appreciated at least by Leibniz and that his claim to the contrary results from a failure to appreciate the manner in which the theory of ideas and hence CCR underpinned critiques along the explicative dimension.

Buchdahl does not as a matter of fact view the independence explicative dimension of his model to have been widely appreciated. Of Newton and Leibniz he maintains that,

Newton holds that gravity can be made “intelligible” or “comprehensible” only by making it “rational,” or “consilient.” The two are absorbed into one another. Leibniz sometimes shared this conflation. Thus he asserts in the New Essays that matter cannot “naturally possess…attraction because it is impossible to conceive how this takes place, i.e. to explain it mechanically.” Again “conceivability” and “explanation” are treated synonymously.  

Buchdahl’s main point in this passage is that issues concerning the intelligibility or comprehensibility or conceivability of a notion are objectively distinct from issues concerning the theories rationality; i.e. its coherence and consilience with accepted regulative maxims. Buchdahl maintains that by conflating the various issues Newton, perhaps unwittingly, failed to address the former. By offering secondary explanations in

---

90 Buchdahl (1970a), p. 82 (Buchdahl’s italics).
terms of impulse Newton was seemingly attempting to make cognitive sense of ‘attraction’ by demonstrating the rationality of his theory; i.e. by making it consilient with the mechanistic philosophy of the time. Newton’s misstep here is not surprising as the mechanism of the day to a large extent constrained or “regulated” the rationality of an explanation. But Buchdahl’s claim that Newton, and moreover Leibniz, conflated the explicability of a theory with its rationality raises questions.

To assert that Leibniz conflated the intelligibility or conceivability of a theory with the distinct notion of its rationality at best passes over a certain complexity in the issue. As I noted earlier at the most basic level, for Leibniz, whatever is conceivable is indeed intelligible. That is to say that if a term or expression is intelligible then ipso facto it has some cognitive description or content and is hence conceivable—conceivability implies intelligibility and visa versa. This however is not to say that whatever cognitive content a term or expression might have will be consilient with the prevailing (regulative) explanatory norms. After all, we can conceive even of logically contradictory situations and surely logical norms or laws will function as regulative maxims constraining the sorts of explanations that are rationally acceptable. At the same time prior conceptions of pertinent terms like ‘body’ may, together with other maxims, function to “regulate” what notions can be consistently conjoined with them.

All of this suggests the possibility of conceiving of something that is not in fact materially possible or for that matter possible relative to the accepted regulative norms. As I suggested earlier Leibniz clearly had in mind some distinction between simply conceiving and conceiving in the light of given regulative constraints. That is, he distinguished between simply having some cognitive content and having a true content.
Given the received structure of ideas and the epistemic nature of the understanding, it was unsurprising that some conceptions might turn out to be false. After all, the goal of science and philosophy generally was to discern those conceptions of ours that are false from those which are true. Such determinations were to be made through an inspection of the ideas themselves. It would hardly be shocking to find someone endorsing this view claiming that truth and/or explanation was necessarily constrained by the natures of the ideas involved and that what is conceivable relative to that parameter to be more narrowly defined than that which is conceivable simpliciter. To maintain this is I think nothing more than to maintain that to simply have some or another conception is not altogether too difficult, while having a true or scientifically sound conception will involve something more; namely, the consilience of the latter conception with the received regulative norms.

If I am right it is a mistake to think that intelligibility came down to mechanical explicability. While the mechanistic philosophy may have influenced opinions concerning what was “materially” conceivable, there is no evidence to indicate that it constrained what was conceivable or intelligible simpliciter. As Leibniz saw it “conceivability” and “explicability” were epistemically and logically related. Indeed, he makes this clear in the larger passage from which Buchdahl’s quote is extracted:

Whenever we find some quality in a subject, we ought to believe that if we understood the nature of both the subject and the quality we would conceive how the quality could arise from it. So within the order of nature (miracles apart) it is not at God’s arbitrary discretion to attach this or that quality haphazardly to substances. He will never give them any which are
not natural to them, that is, which cannot arise from their nature as explicable modifications. So we may take it that matter will not naturally possess the attractive power referred to above, and that it will not of itself move in a curved path, because it is impossible to conceive how this could happen—that is, to explain it mechanically—whereas what is natural must be such as could become distinctly conceivable by anyone admitted into the secrets of things.\textsuperscript{91}

To understand a subject or a quality is, as has already been pointed out, to possess a clear idea of it. With respect to Leibniz’s epistemology there is a vast difference between what is required of an explanation of natural phenomena and what is required in order for the explanation to be intelligible or comprehensible. For the later the mere presence of ideas is sufficient, whereas for the former the explanation must be couched in terms that are not only natural but also consilient with the regulative norms that govern our ideas of nature.

Buchdahl nowhere explicitly associates the worry with its consilience with the dominant theory of ideas. He does not in any primary sense accept the \textit{Core Conception} as part of his explanation of the dialectic over attraction. Rather, in his view while the worry always was centered on the intelligibility of ‘attraction,’ most individuals, including Newton, failed to grasp what was required in order to redress this grievance. In Buchdahl’s view it is only after Kant transposes attractive forces, construing them as preconditions for the possibility of the concept of matter that we meet with any legitimate response to the original concern—a genuine explication, as it were, of ‘attraction’.

\textsuperscript{91} Leibniz, (1765), p. 66. Buchdahl’s quote is from the Langley translation (La Salle, 1949).
Buchdahl’s claim that Newton’s appeals to secondary explanations failed to meet the charge of unintelligibility is squarely in line with my treatment of the critique. However, while Buchdahl’s chief aim is to illustrate what he takes to be a historically significant dimension of theory choice—namely, the concern with the explication of, or comprehensibility of theoretical terms—his predisposition to view the notion of intelligibility in the late seventeenth century and early eighteenth century as variously unclear causes him to overlook the degree to which the core conception was shared. Though he does believe that the issue involved all along a concern over the comprehensibility of attraction, he thinks that this fact was not always well understood as evidenced by the diverse attempts to meet the charge. As he says,

Certainly, we meet with a whole spectrum of cases, from one extreme of straight logical impossibility, via Hume’s approach which proceeds from his contention that *actio in distans* is contrary to one of his criteria for causal action, spatial contiguity, via a supposed lack of “familiarity,” and “inability to frame a picture,” via the supposed clash with predominant physical theories or conceptual schemes, all the way to the other extreme of straight muddleheadedness or even prejudice.\textsuperscript{92}

2.5.0. Concluding Remarks

The principal focus of the foregoing remarks has been to clarify the central worry underlying intelligibility during the 17\textsuperscript{th} and 18\textsuperscript{th} centuries and in particular to clarify Leibniz’s criticism of ‘attraction’ on those grounds. This chapter, like the previous chapter, has been primarily concerned with setting up the issue of intelligibility and

\textsuperscript{92} Buchdahl (1970a), p. 85.
providing a historical perspective on it. The centerpiece of the picture I have been painting is CC/CCR. I take this to provide two crucial desiderata for the development of an adequate criterion of intelligibility. Given that we want to develop an historically adequate criterion it ought to be able to explain the historical appeal to CC/CCR. Correlated with this is what we might label the specificity desiderata: Any adequate criterion of intelligibility will need to specify the cognitive contents responsible for a terms intelligibility. In the next two chapters I will develop a criterion of intelligibility around CC/CCR.

Despite my differences with some of the particulars of Buchdahl’s view I claim him as an ally. His model of theory assessment is a focal point for the development of the model for cognitive significance within which I situate concerns with intelligibility. Beyond this, much of the material discussed in the present chapter will resurface throughout the remaining chapters and will make a significant contribution to chapter IV.
Chapter III

Were it fit to trouble thee with the History of this Essay, I should tell thee that five or six Friends meeting at my Chamber, and discoursing on a Subject very remote from this, found themselves quickly at a stand, by the Difficulties that rose on every side. After we had a while puzzled our selves, without coming any nearer a Resolution of those Doubts which perplexed us, it came into my Thoughts, that we took a wrong course; and that, before we set our selves Upon Enquires of that Nature, it was necessary to examine our own Abilities, and see, what Objects our Understandings were, or were not fitted to deal with.93

3.0.0. Introduction

While CCR captures what is essential to any historically adequate criterion of intelligibility, it does not establish what it is for a term to possess cognitive content. Instead it says only that in order for a term to be intelligible it must have some cognitive content. It leaves the question concerning the nature of the content of cognition open. While it is true that anyone committed to the theory of ideas was thereby committed to CCR, it is not equally true that everyone was of the same mind when it came to specifying the nature of the ideas themselves. Though everyone agreed that the content on which the understanding operated ultimately derived from the perceptions to which the understanding was privy, there was not a similar agreement concerning the kinds of perceptions might come before the mind. Some theorists like Descartes and Leibniz

endorsed views according to which some of our perceptions are of a purely intellective sort. The ideas underlying principles such as the *principle of sufficient reason* are prime examples.\textsuperscript{94} Opposed to this were empiricist conceptions of the resources available to the understanding that generally restricted these resources to perceptions involving external sensations (sensory) and internal sentiments (feelings).

The account of intelligibility I develop here and in the next chapter is firmly rooted in the empiricist tradition. I do not deny that rationalist alternatives may merit further consideration but I leave that project for another time.

I argued in the previous chapter that CC/CCR and the theory of ideas more generally should not be taken to imply a theory of meaning. In particular, I argued that it ought to not be taken to imply an internalist account of meaning by demonstrating that it is neutral with respect to whether the right account of meaning is of an internalist or externalist variety or something else altogether. Nonetheless, even if this is the right stance to take there is a residual problem.

CC/CCR, and hence any criterion of intelligibility used to flesh out CC/CCR, concerns the contents, and specifically the cognitive contents, associated with terms. Insofar as it concerns the contents expressed by a term it may be objected that it is therefore an account of a term’s cognitive significance, and consequently, it can be criticized on the grounds that it offers the wrong picture of a term’s cognitive significance. There are after all, alternative views concerning what a term’s cognitive significance consists of. According to one prominent view grasping the cognitive significance consists of. According to one prominent view grasping the cognitive

\textsuperscript{94} We do not need to look as far as metaphysical principles like the *principle of sufficient reason* to find examples. For rationalist many of our more mundane sensory notions are rendered intelligible only when we remove the obscuring sensory content from them.
significance of a term consists of grasping its “use” or “inferential role” within a language game. While CC/CCR is neutral with respect to whether it is the association or the designation relation (see figure 2.1) that provides a terms meaning it is not equally neutral with respect to this latest idea. It does indeed offer a theory of cognitive significance that is incompatible with the “inferential role” alternative if either are taken as an exhaustive account of cognitive significance. Moreover, I am inclined to accept the idea that the appreciation of inferential roles does contribute to a statement’s significance.

This poses a substantive challenge. If we want to provide an empirically plausible criterion of intelligibility (which I take to be an implicit desiderata), that at one and the same time does not violate the desiderata of historical adequacy, we face a dilemma. On the one hand it is not clear that we can explain criticisms like Leibniz’s with the inferential role account of cognitive significance given its apparent incompatibility with CC/CCR. On the other hand if the appreciation of inferential roles wholly determines cognitive significance then CC/CCR is simply false. There are of course a couple of maneuvers we might opt for here. One reasonable response is simply to accept all of this and suggest that CC/CCR while false nonetheless does capture the historical worry at its core—it’s just that the historical worry was wrongheaded. Another possible response is to deny that the inferential role account provides us with anything that contributes to terms cognitive significance. However, as I have already indicated I’m not inclined to go this route. I do think the view captures features of the cognitive significance of terms. I will not however be arguing for this as I believe that it has been adequately taken up by others—some of whom will be discussed later on.95

95 See for example, Wettstein, Howard (1988); Brandom, Robert (2000).
Instead, I will opt for a third route. In section 3.2 of this chapter I will sketch out a broader model of cognitive significance that makes room for both CC/CCR and the inferential role view. The primary merit of this model is that it affords us the ability to develop a criterion that is both empirically plausible as well as historically adequate and hence avoids the problem noted above. Essentially, I maintain that neither account exhausts a terms (or sentences) cognitive significance and that both play a role therein. The model I will develop is, as I indicated in the previous chapter motivated by Buchdahl’s model of theory assessment.

3.1.0. Humean Intelligibility

Recently, Louis Loeb\textsuperscript{96} has noted an apparent tension in Hume’s philosophy. Given that Hume is interested in grounding his theory of the understanding in an empirical psychology based on impressions and ideas, he often has no other recourse but to dismiss certain terms as unintelligible. Typically, the terms in question are those that were central to the school metaphysics of his time and, consequently, widely thought to be significant. Thus, in addition to demarcating metaphysical notions like ‘causation’ and ‘substance’ as meaningless (in their metaphysical senses), Hume feels the need to explain the metaphysicians’ mistake in a manner that accords with his emerging theory of the understanding. The presence of these two projects has led Loeb to wonder whether Hume is able to “consistently set out to explain the psychological causes of a belief that is without meaning or content in the first place.” There is, Loeb suggests, at the very least, a tension between the two projects.

While Loeb and I both share a desire to dissolve this tension, I find Loeb’s resolution of it unsatisfactory. What follows is a critical examination of the problem and Loeb’s solution. In the next sub-section I will characterize the problem more precisely as a problem that arises in the context of applying Hume’s criterion of intelligibility. I will then explain Loeb’s response to the problem, which posits what he terms quasi-contents in order to explain the sort of error involved in the metaphysicians’ mistake. Finally, I will provide reasons for rejecting Loeb’s solution to the problem. In particular I will argue that there is little textual support for the claim that Hume was implicitly appealing to ‘quasi-contents’. Moreover, I will argue that had Hume adopted such a notion as part of his explanation of the metaphysicians’ mistake, he would have been involved in an even deeper inconsistency given his criterion of intelligibility. Before getting started it may be helpful to situate the following discussion in the context of Hume’s theory of ideas.

Most philosophers are familiar with the foundational apparatus of Hume’s theory of the understanding. Hume divides perceptions, the basic units of cognition, into impressions and ideas. Both ideas and impressions may be either simple (i.e. indivisible wholes) or complex (i.e. divisible into parts.) Both lie on a continuum of degree of force and vivacity. Yet, with respect to the distinction between impressions and ideas this continuum is not well ordered as both are intermingled along the spectrum. Thus, Hume tells us, “in any violent emotions of the soul, our ideas may approach to our

97 This paper is part of a larger project in which aims at offering an alternative account of how Hume may deal with the problem being suggested here.
impressions.”98 Additionally, Hume adopts what some expositors have termed the *copy principle*. According to this principle “all our simple ideas in their first appearance are deriv’d from simple impressions, which are correspondent to them, and which they exactly represent.”99 It is upon this basic cognitive framework that Hume grounds what I have been calling his *criterion of intelligibility*.

Like Locke, Hume thought that terms *signify* ideas, and he maintained that a term is intelligible or meaningful only insofar as it signifies some idea. To this extent Hume is clearly committed to CCR. But Hume’s account goes beyond this. Whether or not some idea is signified by a term is a function of the impression(s) to which the idea corresponds or from which it is copied. While meaningful terms signify *genuine* ideas, unintelligible terms or expressions signify, at best, a “feigning” of the imagination.100 Such feigned ideas are not genuine because there are no corresponding impressions from which the feigned thought is derived. As Hume saw it these later terms were unintelligible. This framework, then, grounds Hume’s intelligibility criterion and in turn, his criticisms of the

----


99 T 1.1.1, p. 4.

100 Hume variously speaks of the Imagination feigning ideas as he does when accounting for how his opponents attempt to resolve the contradiction in attributing both “simplicity” and “diversity” to a series of impressions: “In order to reconcile which contradictions the imagination is apt to feign something unknown and invisible, which it supposes to continue the same under all these variations; and this unintelligible something it calls a *substance, or original and first matter*” (T x.x.x, p. 220).” See also: EHU, pp. 21-22…
school metaphysics of his day. Characterizing it in the *Abstract*, he tells us that he practices a method according to which,

wherever any idea is ambiguous, he has always recourse to the impression… when he suspects that any philosophical term has no idea annexed to it (as is too common) he always asks from what impression that pretended idea is derived? And if no impression can be produced, he concludes that the term is altogether insignificant and it were to be wished, that this method were more practised in all philosophical debates.\(^{101}\)

In short, impressions produce the content on which the understanding operates.\(^{102}\) Accordingly, when Hume claims a term is unintelligible he is saying that no genuine idea is being signified by its use. That is, no idea whose simple parts can be traced to their corresponding impressions is signified by the expression. It is within this context that Hume claims at T. 1.4.3 that “substance” is an “unintelligible chimera” and that at T. 1.3.14 he says of ‘power’, “We never have any impression, that contains any power or efficacy. We never therefore have any idea of power.”\(^{103}\)

According to Hume then, the only objects on which the understanding may operate are those perceptions which are suitably connected to impressions, or experience. Impressions or observations are *given* in our experience and ideas are copied from them. Of course to say that ideas are “copied” may be misleading. The vast majority of our ideas are complex in nature and consequently involve a variety of simple ideas. Hume’s

\(^{101}\) A, pp. 648-649.

\(^{102}\) This statement should be treated with some suspicion. Strictly speaking, in my view, Hume does not have a theory of semantic content. See for instance Hacking, (1975) and above. Nor is it clear that Hume has much to offer as a theory of communication. Nonetheless, following Loeb I will here speak of contents in his sense.

\(^{103}\) T 1.3.14, p. 161.
contention is that the simple ideas, of which complex ideas are composed, represent or correspond to simple impressions or basic observations. If we were to confine ourselves to just those ideas that are simple we can think of them as intentional representations. Of course, on Hume’s account, the only things that these intentional representations might be “of or about” are the observations (impressions) which gave rise to them. Moreover, impressions are not themselves intentional since there is nothing of which we can have any idea that they might be of or about, that is to say they don’t represent. Talk about ‘the thing(s) that give rise to our impression(s)’ is itself a bit of nonsense in Hume’s view.\textsuperscript{104} Consequently, if simple ideas were all there were Hume’s criterion of intelligibility would be reducible to the view that \textit{a term is intelligible if and only if it is associated with a representation (Idea) which corresponds with some feature of our experience (impression)}. Fortunately, things are not so simple.

As I noted in the previous chapter, for Hume it is the function of the understanding to permute and combine our simple ideas thus deriving compound or complex ideas which may or may not (in the case of fictions) track features of our experience. As he says,

\begin{quote}
Nothing is more free than the imagination of man: and though it cannot exceed that original stock of ideas furnished by the internal and external senses, it has unlimited power of mixing, compounding, separating, and dividing these ideas, in all of the varieties of fiction and vision.\textsuperscript{105}
\end{quote}

Obviously, those of our ideas which constitute fictions do not themselves correspond to any impression or observation. Consequently, they are not intentional objects in the same

\textsuperscript{104} Add footnote.\textsuperscript{105} EHU. 47.
sense that simple ideas can be said to be. On the other hand complex ideas clearly will have some intentional content insofar as they derive this content from the simple ideas of which they are comprised. On this reconstruction we can think of Hume’s criterion of intelligibility as coming to the claim that a term is intelligible if and only if it is associated with a representation (Idea) all of whose simple constituents correspond with some feature of our experience (impression). With this apparatus in hand we can now proceed to Loeb’s problem.

3.1.1. Articulating the Problem

From the standpoint of Hume’s theory of the understanding, our idea of, say, a peach is a complex idea “form’d by the mind of the ideas of the several distinct sensible qualities” of the peach. At the same time he observes that we have a tendency to regard the peach as one thing “continuing the SAME under very considerable alterations.” But as Hume sees it these two views stand in conflict with one another. Hume devotes 1.4.3 of the Treatise to considering “the causes which make us almost universally fall into such evident contradictions, as well as the means by which we endeavour to conceal them.”

Loeb construes Hume’s project here to be that of providing a causal explanation of the metaphysicians’ belief in ‘substance’. The metaphysician posits the existence of some “substance” in order to explain the continuity and sameness of the peach in the face of such apparent alterations and differences over time. However, it is clear that ‘substance’ does not designate any permutation or combination of sensible qualities and therefore ought to be considered unintelligible given Hume’s account. Whatever putative idea ‘substance’ might be claimed to hold the place of it will not be one whose simple

\(^{106}\) T 1.4.3, p. 219
constituents track basic features of experience since all of these will be ruled out by the claim that it is not to be identified with any combination of the sensible qualities. But if ‘substance’ is unintelligible it is difficult to see how Hume can offer any causal explanation of the metaphysician’s belief if his official position is that such a belief is insignificant. According to Loeb,

Hume might try and explain why the ancient philosophers believe that their belief is meaningful…but Hume also offers an explanation of why the ancient philosophers believe that substrata exist; if the term ‘substratum’ is ‘wholly insignificant and unintelligible’ it is not possible to identify the belief Hume seeks to explain.107

The problem, then, is that if Hume were able to provide an account of the causes that induce our belief in ‘substratum’ from the standpoint of his theory, then such a belief ought to not be meaningless. In providing such a causal explanation he would in effect be analyzing the content of the idea. And this problem, to the extent that it exists for Hume, is not confined to the discussion of substance. As Loeb has observed Hume makes similar moves in his treatment of a number of other purportedly meaningless expressions including, “immaterial substrata or souls, external existence and necessary connection.”108 Nor is Loeb the first to suggest that there might be a problem for Hume in this regard. Galen Strawson maintains that,

Hume uses Causation words in a way that is arguably ruled out as illegitimate by his theory of ideas: although he officially holds that no Causation term can manage to (descriptive-contentfully) mean anything

108 Ibid
like Causation, which is what he ought to say given his theory of ideas, yet he clearly concedes in his usage that they may manage to mean something like Causation at least in the sense of genuinely referring to it.\textsuperscript{109}

The question comes to this: How can Hume maintain at one and the same time that a term is unintelligible though significant, given the close tie he draws between intelligibility and significance? How, that is, can Hume maintain that ‘substance’, for instance, denotes at best an “unintelligible chimera” and then proceed to tell us how we come by its idea? How can Hume possibly seek to coherently individuate unintelligible terms?

3.1.2. Loeb’s Solution

Loeb’s response to the problem has it that Hume was implicitly appealing to quasi-contents. While the metaphysicians’ notion of substratum is not genuinely contentful in the sense explained earlier, “the confused conception of substratum… has content-like features insofar as it is the by-product of a determinate illusion.”\textsuperscript{110} According to Loeb quasi-contents emerge from psychological pressures. In the case at hand the pressure stems from our attempt to reconcile the two views of our peach: the first, according to which we think of it as a simple whole; and the second, according to which we view it as a complex of qualities.

As Loeb sees it, the occurrence and resolution of this pressure comes about through a four-stage process. The first two stages involve our recognition of the two means of viewing the peach. In the first stage we observe our successive and gradually changing impressions of the peach. Since each successive impression closely resembles the former we are led to ascribe identity to our impressions of the peach and thus come to

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{110} Loeb (2001), p. 151.
\end{itemize}
\end{footnotesize}
view it as a single thing persisting over time. In the second stage we compare more temporally distant impressions of the peach and are led to attribute diversity to the two impressions and consequently to the whole sequence. As Hume articulates it,

When we gradually follow an object in its successive changes, the smooth progress of the thought makes us ascribe an identity to the succession…
When we compare its situation after a considerable change the progress of the thought is broke; and consequently we are presented with the idea of diversity…. ¹¹¹

Of course, strictly speaking, the conception that arises in stage 1 is mistaken. Our stage 2 conception is a more accurate portrayal of the content of our experience. Nonetheless, Hume thinks that it is natural “for men, in their common and careless way of thinking, to imagine they perceive a connexion.”¹¹² This “common and careless way,” Hume claims is characteristic of the “vulgar” and while not fully consonant with the “true philosophy” it better approximates it than the “false” or “metaphysical” philosophy.¹¹³ For while the vulgar “are apt to fancy such a separation to be itself impossible and absurd,”¹¹⁴ the metaphysical philosopher is perceptive enough to see the vulgar mistake, though not sufficiently perceptive to see the futility of seeking the connexion in matter. The metaphysician moves into Loeb’s third stage.

While the vulgar fail to appreciate the philosophical difficulty in viewing the peach as an identity, the metaphysician feels a discomfort resulting from the natural propensity that arises in stage 1 and the reflective view of stage 2. Loeb characterizes this

¹¹¹ T 1.4.3. p. 220.
¹¹² T 1.4.3. p. 223.
¹¹³ T 1.4.3. p. 222.
¹¹⁴ T 1.4.3. p. 223.
“discomfort and the effort to relieve it” as constitutive of stage 3. The vulgar never enter this stage because they are, according to Hume, unable to appreciate the step taken at stage 2. Stage 3 culminates in a minimal retreat from the view taken in stage 1 to the supposition of “the existence of a wholly unobserved object which is unchanging as well as uninterrupted, and in which sensible qualities inhere.” This supposition constitutes, for Loeb, stage 4 of the four-stage process.

What we have at this point then is a characterization of the psychological processes that lead to the supposition of substance. It is not, however, immediately clear how this is going to resolve the apparent inconsistency between what Loeb sees as Hume’s tendency to construe such suppositions as unintelligible and his desire to, in Loeb’s words, “explain the psychological causes” of the metaphysicians’ belief. The pressing question seems to be left unanswered.

Loeb however does have a response. Since the supposition is offered as an explanation of the tension engendered in the first two stages even though it is not “strictly, contentful or meaningful” it has content-like features. Thus, while the metaphysician’s belief lacks genuine content it does have quasi-content. Loeb is not suggesting here that quasi-contents are sufficient to render a belief intelligible. Rather, Loeb is suggesting that if we construe Hume as having implicitly adopted such a notion then it will be possible to see how Hume might consistently set out to explain a belief that is otherwise unintelligible and such contents can be individuated in terms of the explanatory roles that they play.

---

116 Ibid.
Loeb’s answer to the pressing question is, then, that Hume, in offering up psychological explanations, is attributing quasi-content to such locutions. Hume can then consistently maintain that the terms in question are really meaningless or unintelligible though they are not without some sort of content; namely, quasi-content. The two views are consistent since the former claim is true just in case the questionable terms cannot be rendered meaningful from the point of view of Hume’s theory of ideas, whereas the latter depends merely on there being some way of individuating the respective contents. Loeb individuates and hence identifies quasi-contents by their psychological contexts. “Strictly meaningless terms,” he claims, “serve as place holders for ideas we do not possess.” They derive their content from the role they play in our thought and in our language. In his view,

the expression ‘material substratum’, for example, is associated with the quasi content at stage (4) in virtue of being employed in the context of observations (of a succession of uninterrupted and gradually changing sensible qualities) that activate the propensity that gives rise to the illusion at (1). In this way, different expressions, ‘material substratum’, ‘inmaterial substratum’, ‘external existence’, ‘necessary connection’, though not strictly meaningful, are associated in both outer and inner speech, with different quasi-contents. Which quasi-content is in play depends upon the context in which these expressions are placed or used.

---

119 Ibid.
Here then we have a response to the pressing question. It remains then to examine whether or not Loeb is correct in attributing the view to Hume.

3.1.3. Problems with Loeb’s Interpretation

The central problem for Loeb’s view is that not only does Hume not seem to ever take notice of the problem that Loeb and Strawson characterize, but, he never once intimates that expressions can have any significance outside that granted by his theory of ideas. One would suspect someone who adopts the view Loeb attributes to Hume, to either acknowledge the problem that gave rise to the view, or give some hint that terms may have alternate, even if counterfeit, signification outside the official theory. Of course, it is possible that Loeb is right and that Hume thought that both the problem and his implicit appeal to quasi-contents would be obvious to his reader. But the problem is deeper. Hume is explicit in both the first *Enquiry* and the *Treatise* where he discusses the significance and intelligibility of terms that no further type of content exists.

Echoing the remark quoted earlier from the *Abstract*, Hume claims in section two of the *Enquiry* that, “when we entertain, therefore, any suspicion that a philosophical term is employed without meaning or idea… we need but enquire, from what impression is that supposed idea derived.”\(^{120}\) As well he makes it explicit, “that it is impossible for us to *think* of anything, which we have not antecedently *felt*, either by our external or internal senses.”\(^{121}\) Thus, if we ask the question, ‘What are quasi-contents?’, from a point of view internal to Hume’s theory of ideas we will need an account of them that renders them reducible to specific impressions. Otherwise we commit ourselves to the view that Hume appealed implicitly to unintelligible notions in order to explain the occurrence of

---

\(^{120}\) EHU, p. 22.  
\(^{121}\) EHU, p. 62.
certain beliefs. It will however be difficult to specify in any truly meaningful way the content of ‘quasi-content’ for the simple reason that, by Loeb’s own definition, there are no ideas specifically associated with the expression.

Right or wrong, Hume thought that terms get their meaning through their customary association with ideas. On this model genuine Humean contents are easy enough to explain. We derive the concept reflectively from observations of the customary association between words and ideas. Quasi-contents, on the other hand, cannot be given a parallel explanation. There is no idea or set of impressions with which the expression is associated. At best it might be associated rather opaquely with the feeling of dissonance but this won’t be sufficient to individuate such purported contents on Hume’s official view.

This isn’t to say that Hume doesn’t have an error theory of sorts. He announces just such a theory in the section that has been the primary focus of this paper. He claims that,

it being usual, after the frequent use of terms, which are really significant and intelligible, to omit the idea, which we wou’d express by them, and to preserve only the custom, by which we recall the idea at pleasure; so it naturally happens, that after the frequent use of terms, which are wholly insignificant and unintelligible, we fancy them to be on the same footing with the precedent and to have a secret meaning, which we might discover by reflection.122

122 T x.x.x, p. 224.
Loeb cites this passage and assumes that it somehow supports the reading he has offered. It is, however, difficult to see how it does. It would seem that Hume could not be more clear, than he is here, that the relevant metaphysical locutions are without any content or signification. None, that is, that would suffice to individuate them any further than as unintelligible utterances.

Cognitively speaking there is no difference between ‘substance’ and ‘attractive force’. In either case we have only the impressions of the terms themselves together with, as I will argue shortly, the sentiment of belief. From Hume’s point of view with the exception of the differing impressions of the terms themselves there is no cognitive difference between the psychological context in which we utter the one and the psychological context in which we utter the other.

That said, there is a relatively clear and obvious way that we might individuate unintelligible terms. We might well distinguish say ‘attractive force’ from ‘substance’ by the differing linguistic contexts in which they are used. While such terms may themselves be wholly without cognitive content (internal sentiments and impressions of the terms aside), they will no doubt arise in contexts in which they exhibit roughly definable logical or inferential relations to other terms that are cognitively significant. In such cases these terms might be thought to derive a sort of “semantic” content sufficient to individuate them. Despite the fact that such “inferential content” might serve to individuate some cognitively contentless terms, it will not suffice as an account of cognitive content. In the simplest case if there were no cognitive contents in the sense I have been describing and inferential content was all that there were, then our utterances would amount to nothing more than, at best, well-formed formulas in an otherwise vacuous system of signs.
On the other hand, if it is admitted that at least some of the terms in a given semantic context must be cognitively significant in my sense, then I have no particular issue, at present, with anyone who would maintain that some otherwise cognitively contentless term may, so to speak, get its cognitive bearings from its inferential association with terms which are really cognitively significant. Here again though I leave it to proponents of this view to explain their account. The main point here is that insofar as we are interested in developing an account of cognitive content any account of inferential content that is not vacuous will be parasitic upon the account being offered here. That said, this isn’t the sort of claim that Loeb wants to make. Loeb aims to individuate unintelligible terms by the psychological contexts in which they occur. I have already argued that given Hume’s criterion of intelligibility this view is incoherent on the face of it.

If Hume allows quasi-contents into his account, his arguments to unintelligibility and insignificance lose their punch. Metaphysical discourse becomes, at least partially, vindicated. Hume would then simply be making a mistake in claiming that it would be impossible to think of ‘substance’. His view would allow that we are able to think of it as “that which plays the role of unifying the several distinct sensible qualities of x.” Thus, if we are to accept Loeb’s interpretation then we must give up the notion that Hume was seriously committed to the criterion of intelligibility that he marks so explicitly or, alternatively, come to see Hume as thinking that unintelligible terms might be used in a
meaningful way.\textsuperscript{123} The first, I suggest, is prima facie implausible, and the second would leave Hume in the grips of a deeper inconsistency than the one with which we started.

The question that we began with thus remains unanswered. And it certainly deserves an answer. In next section I want to suggest an alternative response to the problem thus posed. Briefly, I will argue that the problem arises in the context of a misrepresentation of Hume’s account of the nature of beliefs. The “Big” claim that I will be supporting is that it is possible on Hume’s official theory for an individual to believe that $\phi$ without it being the case that $\phi$ is cognitively significant. That is, it is possible for me to make an epistemic commitment to a meaningless statement. In terms of the preceding I want to argue that believing that $\phi$ doesn’t entail the intelligibility of $\phi$.

3.1.4. Hume’s Solution

“[T]he difference between fiction and belief,” Hume tells us, lies in some sentiment or feeling which is annexed to the latter, not to the former, and which depends not on the will, nor can be commanded at pleasure.\textsuperscript{124} The difference, that is, between a mere idea and one that serves as the content of a belief is that the latter is attended by what he here terms a sentiment. In the next paragraph he describes this sentiment in terms of the force and vivacity it affords ideas. Of course this is not yet an account of the mechanism but rather a description of what is to be accounted for by the postulation of any mechanism. There is no further content added to the idea or any further association made between

\textsuperscript{123} One possibility is to construe the coherent use of a term as minimally sufficient for the term to meaningful, though not for its being intelligible. Bennett, for instance, motions in this direction, yet he rightly notes that Hume doesn’t appear to have such a notion of meaning (LBH: Critical Themes).

\textsuperscript{124} EHU, p. 48
ideas. Indeed Hume seems to think that no conscious or active capacity of the imagination can account for the fact that we form beliefs,

as it is impossible that this faculty of imagination can ever, of itself, reach belief, it is evident that belief consists not in the peculiar nature or order of ideas, but in the manner of their conception, and in their feeling to the mind.125

What needs to be explained then is the extra force and vivacity attending beliefs. Wherefrom, that is, does the belief derive this extra force and vivacity? Hume’s answer is not surprising. Given that every simple idea is copied from a corresponding simple impression and that the latter are distinguished from the former by their greater force and vivacity, it is a short step to the view that what enlivens our ideas and elevates them to the status of belief is a present impression which transmits some of its force and vivacity to the former. As Hume puts it, when one of the constituents of an association is presented, “the mind is not only carried to the conception of the correlative, but reaches a steadier and stronger conception of it”.126 Thus, when I see a coin tossed into the air I immediately form with much force and vivacity the thought of it falling back down. On this view nature compels us to associate ideas frequently conjoined in our experience, and it is their customary connection in the imagination that is responsible for their continued association and in this our most basic beliefs are constituted. Thus in the Treatise Hume tells us that “An opinion, therefore or belief may be most accurately defin’d, A LIVELY IDEA RELATED TO OR ASSOCIATED WITH A PRESENT IMPRESSION.”127

125 EHU, p. 49
126 EHU, p. 51
127 T x.x.x, p. 96
Given this account, it is sometimes thought that ‘beliefs’ constitute a special subclass of our ideas.\textsuperscript{128} If we have this view of beliefs then we certainly face the problem identified by Loeb. Again, in order for an individual S to believe a proposition P, P must be cognitively significant (i.e. it must take some idea as its object). But if S believes that P, then P is intelligible. Yet in the problematic cases Loeb highlights Hume has already told us that key constituents of the proposition (e.g. ‘substance’) are unintelligible and so the propositions in which these concepts occur/are expressed do not constitute beliefs.

Loeb indeed thinks that this picture of beliefs is the one traditionally accepted. At the same time he recognizes that this can’t be Hume’s official theory of belief if he is right to attribute an acceptance of quasi-contents to Hume.

if, as tradition in Hume interpretation has it, beliefs are lively ideas, and if quasi-contents are not ideas derived from simple impressions that copy experience, then we cannot have genuine belief in these cases.\textsuperscript{129}

Loeb’s response is to deny that beliefs are “lively ideas” for Hume and to claim instead that they are “steady dispositions.”\textsuperscript{130} I agree with Loeb that, for Hume, beliefs are not a special subset of ideas; namely, \textit{lively ideas}. If we maintain this, the original problem seems incapable of being solved. At the same time if we don’t construe Hume as holding this theory of belief, then the original problem may disappear altogether and there will consequently be no special reason to attribute an implicit commitment to quasi-contents to Hume.

\textsuperscript{128} See for instance Owen, p. 155.
\textsuperscript{129} Loeb (2001), pp. 152-153.
\textsuperscript{130} Ibid, p. 153
The question then is what grounds have we for thinking that beliefs are for Hume something other than “lively ideas” and what sense are we to make out of his seeming commitment, in the passage quoted above, that they are? The answer to the first part of this question has already been alluded to above. For Hume, belief is a sentiment or feeling “excited by nature.” He articulates this view most clearly in the *Enquiry*,

the difference between fiction and belief lies in some sentiment or feeling, which is annexed to the latter, not to the former, and which depends not on the will, nor can be commanded at pleasure…. Were we to attempt a definition of this sentiment… Belief is the true and proper name of this feeling; and no one is ever at a loss to know the meaning of that term; because every man is every moment conscious of the sentiment represented by it.131

“And,” Hume tells us, “in philosophy, we can go no further than assert, that belief is something felt by the mind.”132 It is the sentiment that ordinarily enlivens the idea. When Hume talks about beliefs as “lively ideas” he is primarily interested in distinguishing those ideas that are mere occurrences in the mind from those to which we have commitment. What distinguishes them is the presence of the sentiment. This does not logically commit Hume to the view that every belief involves an idea. Being “lively ideas” is distinctive of those ideas that “command our assent,” though not every putative thought that commands assent need be an idea. In short, everything that Hume says here about belief, is compatible with his maintaining that we sometimes have beliefs in the absence of any associated genuine idea.

---

131 EHU, pp. 48-49
132 Ibid
In his discussion of testimony at T 1.3.9, for instance, Hume claims that “when we receive any matter of fact upon human testimony, our faith arises from the very same origin as our inferences from causes to effects.”\(^{133}\) That is the belief here arises from the customary connection between words and ideas even in the absence of any actual connection.\(^{134}\) It’s true that most of the cases Hume considers in this light are cases where it is in principle at least possible to have the idea but his overall point and the important one I take it is that we can have the belief in the absence of the idea and solely on the basis of the testimony. The belief in these cases is associated with the propositions expressed. Moreover, in the passage cited earlier that I claimed constituted Hume’s expression of an error theory he maintains explicitly that we frequently omit the ideas which would otherwise correspond to the words. The problem with the Metaphysician is that there is no idea connected with the terms being employed. At best, it is our implicit assumption of connection between words and ideas through their customary association and the constant repetition of the words that produces the sentiment or, in other words generates the belief.

If this is the right way to construe Hume’s account of beliefs then there is a straightforward response to the problem we began with. The response, I have been suggesting, is that there is no problem to begin with, as having a belief does not require us to have the idea or for it to even be possible to generate any corresponding idea. This account I suggest fits more easily with what Hume says than any account that requires us to attribute a class of contents to Hume that he in fact never overtly takes notice of. Finally, this sort of response has the exegetical virtue of not requiring us to think that

\(^{133}\) T 1.3.9, p. 113
\(^{134}\) See as well *Enquiry*, section X.
Hume simply failed to take note of problem that seemingly should have been obvious to him.

What I want to take away from this discussion is first and foremost Hume’s criterion of intelligibility. Again that criterion can be stated as follows:

**HC**: A term is intelligible if and only if it is associated with a representation (idea) all of whose simple constituents correspond with some feature of our experience (impression).

HC is approximately the criterion of intelligibility that I shall be defending.

The second point I want to take away from this discussion is Hume’s indication of an error theory. I don’t know of any straightforward way of arguing against someone who rejects LA and accepts that some further cognitive content exists. But it is to my mind significant that proponents of HC are able to explain why certain unintelligible terms are thought by some to be intelligible. In the face of disagreement over whether or not a given term is intelligible the proponent of HC can coherently explain the error involved in their opponents view. There is, so far as I can see, no corresponding explanation available on the other side.¹³⁵

3.2.0. HC reconsidered

---

¹³⁵ This isn’t to say that philosophers of other persuasions didn’t have similar error theories. Leibniz’s comments concerning “symbolic thinking” suggest a strikingly similar account of such errors. The present suggestion however turns on the notion that the perceptions endorsed in HC were broadly viewed to supply content for the understanding, whereas the sort of intellectual perceptions suggested by alternative rationalist account enjoyed a less broadly favorable reception. To that end it is not clear to me at least what principled account we might give of the contents endorsed by rationalists. In the absence of such an account it is equally unclear how we are to account for the errors involved in committing ourselves to unintelligible locutions.
Loeb’s problem might be cast in a different light. Suppose we are committed to HC as well as to the idea that the criticism that ‘attraction’ was unintelligible is correct along the lines of this commitment. Accordingly, our claim that ‘attraction’ is unintelligible amounts to the claim that there is not anything that cognitively individuates one’s belief in ‘attraction’ from their belief in, say, ‘material substratum’—assuming this later notion to be similarly unintelligible.\(^\text{136}\) This follows if we take HC to exhaust a term’s cognitive significance.

The problem here is that there does appear to be something that cognitively individuates commitments to claims about ‘attraction’ from claims about ‘material substratum’. The contexts in which we find it “proper” to utter statements involving ‘material substratum’ differ significantly from those in which we find it proper to make claims which appeal to ‘attraction’. Moreover, we are generally fairly adept at determining in which contexts the respective use of terms is licensed. The fact that we are able to make these discriminations evinces a means for individuating the respective beliefs. Equally, it suggests an alternative account of cognitive significance in terms of an expression’s use—rather than in terms of its association with representations of the sort we have been considering. Wittgenstein appears to endorse this alternative view of cognitive significance with his contention that we oughtn’t ask for the meaning of a term but for its use.\(^\text{137}\)

\(^{\text{136}}\) Remember that we are talking about “the quality of attraction” which putatively adheres in bodies and explains gravitational phenomena. The associated representation or idea of this quality is what is would—were it available—render ‘attraction’ intelligible. Gravitational phenomena are, on the other hand, as it were, the brute facts or impressions we are trying to explain with the appeal to ‘attraction’.

\(^{\text{137}}\) Add Citation. I am not endorsing this interpretation of Wittgenstein.
The sort of criticism, of CC/CCR and hence HC, envisioned here is at a different level than the criticisms discussed in chapter II. There I was principally concerned with the thought that CC/CCR codified a theory of meaning and the corresponding potential criticism stemming from commitments to externalist accounts of meaning. My response there was to maintain that intelligibility concerns a terms cognitive significance and hence cannot, properly speaking, be externalized. The criticism here, on the other hand, is one that meets CC/CCR head-on. Those that endorse the inferential role view are offering a straightforward alternative account of cognitive significance.

On this view a statement’s cognitive significance, and consequently the cognitive significance of the terms comprising it, is given by the inferential role(s) the statement plays in the linguistic contexts in which it arises. Generally, advocates of this view hold that there is nothing further to be said concerning the cognitive significance of terms or statements. One doesn’t need to associate ‘rock’ with any particular representation of rocks in order to play the linguistic game of “rock-paper-scissors” so long as one understands the rules that govern the course of play. I’ll call the view suggested here inferentialism.

3.2.1. Inferentialism

That there is a substantive position that we may justifiably label “inferentialism” does not simply follow from Wittgenstein’s suggestion. The inferentialist account of cognitive significance has been endorsed by more recent discussants in the debate over linguistic understanding. Robert Brandom accepts something on the order of the inferentialism described above when he claims that the “master idea” behind Articulating Reasons, “is that what distinguishes specifically discursive practices from the doings of non-concept-
using creatures is their inferential articulation.” And, he understands what it is to express something as, “putting in a form in which it can both serve as and stand in need of reasons: a form in which it can serve as both premise and conclusion in an inference.”

Understanding a term consists for Brandom in understanding the “inferential significance” of the term.

Inferentialism does not straightaway preclude the relevance of representations in an account of terms cognitive significance. But it does suggest that they play little too no role in our understanding of terms and almost certainly no role in determining the meaning of a term or statement. Rather, the inferentialist maintains that the semanticist fix “his gaze upon language as a social, institutional arrangement, and upon speakers as participants in a social practice.” Moreover, Howard Wettstein—the author of the foregoing quote—claims that this approach,

in stark contrast to the Cartesian approach, denies that pieces of language become meaningful by being associated with representations, mental or objective. It is here that the connections between Wittgenstein and contemporary anti-Fregeans emerge most clearly. Consider Putnam’s slogan, slightly adapted: Meanings ain’t in, nor is it available to, the head.

Indeed, a central lesson of Wittgenstein’s Philosophical Investigations is

---

139 Ibid, p.16.
140 I distinguish here accounts of meaning and accounts of cognitive significance. Whereas the former may involve the external designation of a term, the latter in my view necessarily requires an internal account of our understanding of the term.
that there is less available to the head than one might have supposed, and further that whatever is intellectually available is less relevant to philosopher’s questions about language (and even thought) than one might have supposed.\textsuperscript{142}

Wettstein’s remark here suggests that inferentialism is sternly incompatible with anything like the account of intelligibility I am offering. For inferentialist terms or statements are intelligible insofar as they can be inferentially \textit{articulated}.\textsuperscript{143}

I am not presently interested in arguing against inferentialism. As I said at the outset of this chapter I am inclined to believe that there is something correct in the view. Rather than confront the view head on, in the next section I will propose a model of cognitive significance within which both HC and inferentialist accounts of significance may play a role. Before proceeding there are a few points worth noting.

First, insofar as Wettstein’s appeal to Putnam implies that he sees inferentialism as providing an account of meaning, I have no quarrel with him. As I noted in chapter II an account of intelligibility is neutral on the issue of where one should go hunting for “meanings”. What an account of intelligibility like HC is not neutral on is what is involved in a terms being cognitively significant—and it is an open question whether a terms cognitive significance has anything to do with its meaning.

Second, I do disagree with Wettstein’s sentiment that, “whatever is intellectually available is less relevant to philosopher’s questions about language (and even thought) than one might have supposed.” I must confess that the competing intuition; that we do “represent” strikes me as a truism. I don’t see how one could understand a statement like,

\textsuperscript{142} Ibid, p. 430.
\textsuperscript{143} See for example, Brandom (2000), p. 18.
'Matter is composed of corpuscles of different sizes and shapes arranged in different manners’, without in some way representing corpuscles—but this may just be a peculiarity of me. Moreover, what could be more relevant to Leibniz’s questions about language than this purported truism? It is not possible to understand these historical worries if we simply think of them as worries about meaning or even cognitive significance in the inferentialist’s sense. And what good is a theory of cognitive significance if it can’t explain the actual cases in which people fail to understand bits of language?

3.2.2. Cognitive Significance

At this point, then, I want to situate a criterion of intelligibility within a broader model of cognitive significance. This model is motivated by the acceptance of two principal claims:

1. The cognitive significance of a term is partially determined by our ability to represent for ourselves the term’s reference, and
2. The cognitive significance of a term is partially determined by our grasp of the manner in which the term is inferentially articulated.

I take the key idea underlying inferentialist accounts of cognitive significance to be that a term’s significance is a product of its articulation in a language or set of linguistic practices. The notion of a “term’s articulation” is intended to capture inferentialist appeals to expressions like “use,” “inferential significance,” “inferential role,” and so on. Essentially, in accepting 2 above I am accepting what I take to be the core idea underlying inferentialist accounts. That idea is that a term’s cognitive significance is partly determined by (our grasp of) the manner in which it is integrated.
into a system of language or set of linguistic practices. Of course I am here drawing a parallel with Buchdahl’s claim that theories are evaluated in part based on the manner in which the laws are formulated and “integrated into a system”. Just as Buchdahl distinguished between the constitutive and explicative dimensions of theory assessment, I want to draw a parallel distinction with respect to the manner in which we may assess a term’s cognitive significance (See figure 3.1).

<table>
<thead>
<tr>
<th>Buchdahl’s Dimensions of theory assessment</th>
<th>Dimensions of Cognitive significance</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitutive</td>
<td>Articulation</td>
<td>Inferential Role</td>
</tr>
<tr>
<td>Explicative</td>
<td>Intelligibility</td>
<td>Representation</td>
</tr>
</tbody>
</table>

Figure 3.1

---

144 See sec 2.4 above.
Strictly speaking a term’s articulation does not contribute to a term’s intelligibility. One might appreciate a term’s role within a particular linguistic practice without having any conception of what it designates. My earlier comment concerning the game of “rock-paper-scissors” presents a case in point. Moreover, this accords with the idea that Leibniz was able to understand the manner in which ‘attraction’ was integrated within Newton’s theory though he nonetheless failed to find it intelligible. He was able to draw appropriate inferences from statements involving the expression but could not conceive of what it might designate. Accordingly I distinguish between a term’s articulation and its intelligibility.

Inferentialist accounts of cognitive significance won’t constitute plausible accounts if the contention is that understanding language is generally analogous to understanding how to make moves in a linguistic practice like “rock-paper-scissors.” There is more besides. Most importantly for our purposes, learning how to participate in a linguistic practice will generally involve knowing when those practices are appropriate. A linguistic practice is partly characterized by the way in which it constitutes “a manner of speaking in the face of” certain observations. The notion of a term’s articulation is intended to include these considerations as well so that grasping a term’s articulation will also involve grasping the empirical circumstances appropriate to the practice. We speak of ‘beliefs’ and ‘desires’ in the face of certain phenomena, of ‘atoms’ and ‘valences’ in the face of others, and of ‘touchdowns’ and ‘field goals’ in the face of still others. This notion of “a manner of speaking in the face of” phenomena affords inferentialism a means of imposing a necessary empirical constraint on our utterances.
Before moving on, at this point it is worth mentioning one further feature that distinguishes a term’s articulation from its intelligibility. When we criticize a theory on constitutive grounds we are criticizing the theory as a whole; the manner in which it’s laws and statements are integrated into a systematic whole. Similarly, when we criticize the use of a term on the basis of its articulation we are doing so from the point of view of considering the manner in which it figures into a broader linguistic practice. Dramatic failures in this regard are perhaps not widespread but we can think of some schizophrenic utterances as cases in point.\(^{145}\)

Contrasted with the holistic character of articulation is what I will call the local character of intelligibility. When we inquire into a term’s intelligibility we are generally concerned with terms that putatively refer to entities, events or processes and we want to know is how we are to conceive of them. Of course “conceiving of them” does not here mean grasping the terms full conceptual character—doing that will of course entail grasping as well it’s articulation. Rather, it means understanding how to represent for ourselves the entity, event, or process putatively designated. We won’t generally say that whole theories are unintelligible and when we do it is because they incorporate terms or statements that are unintelligible. Leibniz criticism is specifically a criticism of ‘attraction’ and only derivatively a criticism of Newton’s broader theory. A criterion of intelligibility should be able to explain this. I will call this the desiderata of localism: A criterion of intelligibility must be applicable at the local level of terms.

\(^{145}\) I am thinking of something like Judge Schreber here. Perhaps equally severe instances of terms or statements failing to be cognitively significant in this respect may be found, for example, in the writings of certain continental writers.
The principle case for this model of cognitive significance is that it affords us more sophisticated means of understanding concerns (including perhaps most importantly historical concerns) regarding our cognitive apprehension of language. It is not possible to understand Leibniz criticism that ‘attraction’ is unintelligible as a concern with how the term was articulated. Consequently, even if we accept that inferentialist accounts of significance are partly correct we still require a more complex appreciation of cognitive significance in order to understand these kinds of criticisms. HC is offered as part of this broader model of cognitive significance. However, as I stated earlier HC is only approximately the criterion of intelligibility that I want to defend. In the next chapter I will be proposing a modified version of HC that deals with some remaining issues.
Chapter IV

4.0.0 Introduction

In section 3.1.4 I introduced Hume’s criterion (HC) as approximately the criterion of intelligibility to be defended here. The focus of the present chapter is on the further development of this criterion. I will begin by noting some of the virtues of HC before considering some potential problems. HC meets head on the specificity desiderata discussed in the conclusion of chapter II. It elaborates CC by restricting the “intelligibilizing” contents to those that track features of experience. Put another way HC specifies the cognitive mechanism of intelligibility. The mechanism connects terms or statements with observations by associating them with representations whose basic constituents correspond to or are copied from observations. HC as well satisfies the desiderata of historical adequacy discussed in chapter I. Insofar as it encapsulates the core conception (CC) it allows us to comprehend failures of linguistic understanding such as Leibniz’s in the appropriate manner. And finally, it is readily applicable at the level of individual terms and thus satisfies the desiderata of localism. However, before we can accept anything like HC we must face two remaining issues.

The first, which I shall call the constraint problem, was alluded to in chapter II. The basic worry here concerns which representations are the right representations to associate with a term. We can understand this as worry over which representation are the relevant representations. Leibniz would not have been satisfied to have just any representation whatsoever associated with ‘attraction’. It needed to be in some sense the
right representation. If someone, for instance, associated the idea of Adam with the term ‘atom’ we would not judge him or her to have appreciated the term’s intelligibility. There must be some conception but it also must in some sense be the right conception. Nothing that I have said up to now directly addresses this point.

The second issue concerns the problem of intentionality. The basic worry here involves the sense in which a representation of a putative entity, event, or process is a representation of that putative entity, event, or process. When I associate some cognitive image of a table with the term ‘table’, as in “this table here has a broken leg,” the associated representation can be more or less directly copied from observations of the table. The representation has the intentional content that it does because it is copied from or corresponds with the observations I make. When, however, we represent a unicorn the representation does not correspond with observations of unicorns. Rather, it’s intentional content, if this is thought of as deriving from the observations to which the basic components of the representation correspond, will be of horns and horses and the like. When we represent say Boyleian corpuscles the representations we make of them do not correspond or get copied from observations of corpuscles. What then is it that makes it the case that our representations are of or about Boyleian corpuscles rather than the geometrical figures to which they correspond?

In the present chapter I respond to these problems by appealing to features of the account of cognitive significance provided in the previous chapter. Additionally, I argue for a modified version of HC that appeals to cognitive models rather than representations. Consequently, in this chapter I will be gradually switching the emphasis off of Humean representations and onto a notion of cognitive models. It should be noted at the outset that
“representation” and “model” are not synonymous terms in this account. I begin with a discussion of the account of cognitive significance to which I appealed previously and turn later to the issue of models.

4.1.0. Semantic Agnosticism, Observation & Cognitive Significance

There are a few features and implications of HC that bear further explanation before proceeding to address the two issues described above. The first of these, to be addressed presently, concerns the relationship between a terms intelligibility and its “meaning.” Other aspects, which I shall address subsequently, concern the nature and role of observations in the criterion and the general relationship between a terms intelligibility and its articulation.

It is hardly deniable that there is a tendency on the part of some early modern thinkers to use ‘intelligibility’ synonymously with ‘meaning’. Given this, contemporary critics have tended to view early modern concerns about the intelligibility of terms simply as concerns about their “meaning” in one or another modern sense of the term. I argued in chapter II that this is a mistake and it is in part this connection that I have been attempting to dismantle. If we treat a criterion of intelligibility as just another theory of meaning then we will fail to grasp the substance of the early modern concern with intelligibility. As I argued in chapter II, the substantive concern here is not with the terms themselves, or even with their putative referents, but with our cognitive appreciation of them and the implications that these cognitive constructs have for our broader understanding. The general attitude I adopt here is a semantic agnosticism: If there is an interesting notion of meaning to be had it need not have anything to do with a

---

146 Locke and Hume are both philosophers who exercise this tendency. I have offered an explanation for this in chapter 2.
terms cognitive significance, much less its intelligibility. It is worth reiterating that this sort of semantic agnosticism is compatible with the contention that “meanings ain’t in the head.” I am simply not interested in the issue of where meanings are. What I am interested in is how we come to cognitively appreciate language. And it is in response to this question that the model of cognitive significance is offered. We understand (by degrees) a term or statement when we are able to find it cognitively significant and this may come about by either grasping its articulation or by associating it with a representation—i.e. by rendering it intelligible.

Moreover, an account of cognitive significance is not ipso facto an account of the conditions under which a statement (and certainly not a term) is true. It may or may not be the case that it is connected with this notion. The intelligibilizing contents associated with a term should not be taken to determine the truth of statements involving it. Instead a criterion of intelligibility tells us what sorts of cognitive contents get associated with terms and the manner of their association. This perhaps bears explaining. One might think that when we say things like, “There are no Boyleian corpuscles,” that we are claiming that there are no things in the world that correspond to my representation of the same. Someone who thinks this might consequently be tempted to think that Humean representations of the sort described in chapter III do play a role in determining the truth of statements in which the associated terms appear.

My response to the question of whether such representations play this sort of role is an emphatic “no!” I don’t deny that one might sometimes employ such a procedure in deciding whether to reject or accept a statement but it would be a very poor decision procedure to follow in general. Humean representations don’t have the sort of robust
content that would be necessary to achieve reliable results in making these kinds of decisions. My representation of a unicorn may well be that of a uni-horned horse but that doesn’t capture the robust content of my concept of a unicorn. If I want to see whether there are unicorns I don’t go looking around for uni-horned horses. And even if I found a uni-horned horse that is I wouldn’t be justified in claiming that it was a unicorn since what makes a thing a unicorn involves more than appearing as a uni-horned horse.

4.1.1. Observation

The appeal to observation here may suggest an overly positivistic picture of the view being defended. I don’t deny that I am to a degree a positivist sympathizer but I do not believe my account runs afoul of the same criticisms that are sometimes thought to have decapitated positivist accounts of cognitive significance. Typical criticisms, of positivism, that might be thought relevant to the present context involve the idea that there is no “observational vocabulary” with which we can reduce the non-observational terms of a theory. HC does not require this.

HC requires observationally based representations but it does not require a corresponding “observational vocabulary.” The picture of observation that I endorse is as follows. Representations are derived from seeing, hearing, touching and so on. We observe entities, events, or processes. This is to be distinguished from the sense of observation in which we say so-and-so observed that… or in which we less ordinarily say so and so observed it as… In the sense I am using it we do not, strictly speaking, observe refrigerators or beer cans. Rather we observe objects in the face of which we adopt a certain manner of speaking. What we observe is an empirical matter, which may be subjected to scrutiny. That we observe is an empirical fact, which cannot be rationally
scrutinized. Van Fraassen adopts essentially the same account of observation. To illustrate the distinction, he offers the following:

Suppose one of the Stone Age people recently found in the Philippines is shown a tennis ball or a car crash. From his behavior, we see that he has noticed them; for example, he picks up the ball and throws it. But he has not seen that it is a tennis ball, or that some event is a car crash, for he does not even have those concepts. He cannot get that information through perception; he would first have to learn a great deal. To say that he does not see the same things and events as we do, however, is just silly… 147

The view of observation being adopted here then is compatible with the popular and over-rehearsed idea that our language is entirely theory laden. A theory or other linguistic practice establishes a manner of speaking in the face of the observations we make. In doing so the terms of discussion are situated within a broader linguistic context. Because of this there may be no genuinely non-theoretical terms. On the view being offered here a term is theoretical just in case it is either articulated within a broader inferential structure or its associated representation figures in a broader network of associated representations. Given this even if one’s language consisted exclusively of the term ‘slab’ the term would nonetheless be theoretical provided the associated representation figured in a broader network of representations concerning the use of slabs and so on. Terms that are not theoretical in this sense would have little utility. Here again, the point is echoed by van Fraassen:

147 van Fraassen, (1980); p. 15.
All our language is thoroughly theory-infected. If we could cleanse our language of theory-laden terms, beginning with the recently introduced ones like ‘VHF receiver’, continuing through ‘mass’ and ‘impulse to ‘element’ and so on into the prehistory of language formation, we would end up with nothing useful.148

‘Chair’, ‘refrigerator’, and indeed ‘observation’ are every bit as theoretical as ‘electron’, ‘mass’, and ‘attraction’. And this does not entail that within a linguistic practice conventions may not be adopted regarding how we designate what we observe. The conventional assignment of observational values to terms is compatible with the latter’s being theoretical. The notion of representation being employed here thus depends on there being observations but it does not depend on their being a corresponding notion of an observational term.

4.1.2. Miscellany

The representations being described here are Humean representations. Consequently, such representations do not exhaust a term’s cognitive significance. The concept of unicorn involves in part the manner in which ‘unicorn’ is articulated within various linguistic practices. In particular it will involve the manner in which the term arises in mythology and the sorts of things we are inclined to say about unicorns there. Similarly the concept of a corpuscle is given by the manner in which the term is articulated in the context of Boyle’s theory of matter. In that domain ‘corpuscle’ plays a certain role regardless of how we represent it. Corpuscles are material atoms and hence indivisible. They have different shapes or figures. They can be variously arranged. They play a role

in various explanations and so on. We could engage in such a “manner of speaking” even in the absence of possessing any explicit representation of corpuscles just as we could talk about the ‘Loch Ness monster’ in the absence of having any explicit associated representation. We might find some creature and speculate that it is the Loch Ness monster only to later decide that it could not be since we found it in the Mediterranean Sea and there is no plausible explanation for how it might have come to be there. The decision in this case won’t be based on whether it fits whatever cognitive image or representation we might have associated with our use of that expression. It is rather a decision based on a further understanding of the way in which talk of the Loch Ness monster figures in a linguistic practice of referring to a particular creature which lives in a particular geographic region and so on.

It is a consequence of the view being described here that a term or statement may be cognitively significant without being intelligible. Consider a congenitally blind individual. It is difficult to see how one could argue that the congenitally blind individual could find ‘green’, for example, intelligible in the sense described above and codified in HC. However, let us suppose that this individual exercises great facility with her use of the term. She writes beautiful prose incorporating the term and the like. She even gains critical acclaim for her poetry. In such a case it would be hard to maintain that she understood nothing of the terms meaning. She presumably understands something of its use, seemingly even a great deal. At the same time she obviously lacks the ability to make the sort of observations that contribute to most individuals appreciation of ‘green’. It is these latter experiences that I, along with Hume, am claiming are required for the terms intelligibility.
This case is not merely hypothetical. The blind eighteenth century poet Thomas Blacklock indeed achieved a certain degree of fame in his own time in part for his very condition. His own assessment of his ability accords with the distinction drawn above:

> It is possible for the blind, by a retentive memory, to tell you, that the sky is an azure; that the sun, moon, and stars, are bright; that the rose is red, the lily white or yellow, and the tulip variegated. By continually hearing these substantives and adjectives joined, he may be mechanically taught to join them in the same manner; but as he never had any sensation of colour, however accurately he may speak of coloured objects, his language must be like that of a parrot,—without meaning, or without ideas.149

Generally on the view of I am defending a term’s cognitive significance is a function of both its articulation and its intelligibility and it is only in those cases where a term is neither well articulated nor associated with any representation that a term fails to be cognitively significant. If we need to imagine what a case of this sort of failure might look like we can think of a schizophrenic rambling about the ‘valency’ of his beliefs. Insofar as his utterance is not part of any coherent linguistic practice and it is at the same time not possible to construct any representation of the property of beliefs being postulated, we can not attach any cognitive significance to his talk. I take the presence of these kinds of failures with regard to our cognitive apprehension of language to warrant a more nuanced account of cognitive significance; one that is able to differentiate the kinds of failures involved in our understanding of language.

149 From http://www.electricscotland.com/history/other/blacklock_thomas.htm
4.2.0. Constraining Representations

While I want to distinguish between a term's articulation and its intelligibility I do not want to completely dissociate the two. Indeed it is through their association that we are able to respond to the constraint problem. Recall the problem here concerns which representations are the “right” representations. Representations of (very small) homogeneous gelatinous masses won’t do as representations of ‘Boyleian corpuscles’ any more than a representation of a stereo speaker would do as a representation of a ‘coffee pot’. I will say that the representation, if it is to be an appropriate one, will be in some sense constrained by the manner in which the term is articulated within a linguistic practice.

For many terms whose referents are directly observable the constraint problem has a seemingly straightforward solution. One has, we might suppose, an appropriate representation associated with ‘refrigerator’ if the representation is of a refrigerator. We might say that what makes a representation suitable in these cases is the fact that it results from observations of the phenomena in the face of which we speak of ‘refrigerators’. The more difficult cases will then be those in which the terms in question putatively refer to unobservable or hypothetical entities, events, or processes. We cannot copy our representation of Boyleian corpuscles directly from observations of the same. It will of course be possible to make observations of phenomena in the face of which we speak of ‘corpuscles’ but these won’t be suitable for constructing an appropriate representation of Boyleian corpuscles. To see this consider a different case as described by van Fraassen:

if a charged particle traverses a chamber filled with saturated vapour, some atoms in the neighborhood of its path are ionized. If this vapour is
decompressed, and hence becomes supersaturated, it condenses in droplets on the ions, thus marking the path of the ion... while the particle is detected by means of the cloud chamber, and the detection is based on observation, it is clearly not a case of the particle’s being observed.\(^{150}\)

Representations copied from, or corresponding to, observations of the vapor trail are not appropriate representations of ions despite the fact that we speak of ‘ions’ here. Such a representation will be of vapor trails and not ions in the same manner that a representation resulting from observations of a certain clear liquid will be of water and not of H\(_2\)O molecules. Similarly, observations made of gravitational phenomena will not constitute sufficient resources for constructing an appropriate representation of ‘attraction’ if the latter term is taken to designate a quality of bodies. What we want is a representation of the hypothetical entity designated and we want to be sure that the representation is suitably constrained. Tying representations to observations with a specific perceptual history, in face of which we speak in a certain manner, obviously will not do for terms like ‘corpuscle’, ‘ion’, ‘attraction’, and the like. (Indeed, I do not believe such an account will work even for terms like ‘refrigerator’.)

We do not typically find terms intelligible in a vacuum. Generally we find them intelligible in the context of linguistic practices or in light of their articulation. Arguably, one hasn’t found ‘refrigerator’ intelligible if their representation does not in some manner account for the fact that refrigerators are for keeping beer and meat cold. Humean representations do not accomplish this task. Humean representations are rather snapshots of experience so to speak. They only gain the additional content in virtue of being

\(^{150}\) Van Fraassen, (1980); p. 17.
associated with additional representations (e.g. of beer, meat, temperature, and so on.) I will call such a network of associated Humean representations a model (or, properly speaking, a “cognitive model”) Representations that are not situated within a broader network of representations, which do figure in a cognitive model, are little more than pictures with labels attached. This, however, suggests a natural way that representations may be constrained.

Labels or terms are already constrained by the theory or linguistic practice. Terms are inferentially related to other terms in virtue of the statements in which they occur. The modified Humean criterion of intelligibility has these inferential relations serving as constraints on the associations among representations. Accordingly, the constraint problem will be addressed by requiring that a term’s associated representation play a role in a cognitive model (i.e. a network of associated representations) that is itself constrained by the term’s articulation. Any interpretation of a theory or linguistic practice that is adequately constrained by the practice will serve to render the practice intelligible. Consequently, stick and ball representations of molecules (or the cognitive image of the like rather) will count as adequate representations just in case the representations are situated within a broader model for molecular discourse. The balls constitute representations of atoms (albeit crude ones) and serve to render ‘atom’ intelligible. The representation is constrained by the fact that it is associated with other representations, representations of bonds and the like. It is further constrained by the fact that it is associated with a practice of speaking of ‘atoms’ and ‘bonds’ in the context of discussing more fundamental constituents of matter.
Figure 4.1 presents an extremely simplified picture of a very crude molecular theory and a corresponding model of the theory. The theory may say for instance simply that ‘atoms are connected with one another by bonds in order to form molecules’—of course this won’t make a whole lot of sense unless it is further articulated say in the context of a theory of matter. We construct a representation of this claim via an analogy with sticks and balls, by representing atoms as balls and bonds as sticks and constructing a model in which the balls are connected to one another via sticks. The cognitive model will be observationally based in the sense that it is rooted in observations of actual stick and ball models or graphical representations thereof.
In practice the situation will obviously be far more complex than what is suggested by figure 4.1. There is a question concerning how “robust” the model must be before it suffices to render a term intelligible? That is to what extent must the model as a structure of associated representations capture or encode for the terms articulation within the theory or linguistic practice? This will in my view inevitably require a pragmatic answer. It will depend for instance on the level of complexity in the discourse. Balls won’t serve as reasonable representations of atoms if we want to grasp talk of their electro-chemical properties and the like. As I say terms are rendered intelligible in light of their articulation within specific linguistic practices and not once and for all across all linguistic practices. Revisions to a theory will inevitably require revisions of its corresponding models. Such revisions may include revision of the associations among representations as well modifications of the representations themselves. What is required here will depend on the discursive needs.

The view that I am proposing then has the following structure: *Terms* are articulated in virtue of their playing a role in *statements* which are in turn articulated via their *inferential connections* within a *linguistic practice*. These components constitute a linguistic structure. What we grasp in grasping a particular such structure is the overall manner of articulation. What we lack in grasping such a structure is a cognitive interpretation of the structure, a means of linking it with observations or experience and hence rendering its components or even the whole artifice intelligible. This latter task is achieved on the present view by tying each of its components to components in a corresponding observationally based cognitive model. Terms are associated with *representations* related with one another to form more *complex representations* or
interpretations of statements. These “higher order” representations are associated with one another in the broader model. Figure 4.2 provides a representation of the general picture somewhat more elaborate than what is provided in figure 4.1.

Not all representations constitute models in this sense. Humean representations are not strictly speaking models. Models are generated through the association of representations as an interpretation of statements. Thus in figure 4.2 the higher order representation HR1 may serve as a model for statement S1 yet representation R1 is not in fact a model, properly speaking, of T1 but rather a representation of its putative referent.

The broader model will be constrained by the inferential relations among statements.
Accordingly associations \((A_x)\) between \(HR_i\) and \(HR_j\) will ideally parallel the inferential relations \((I_x)\) among statements in the theory.

Several features of the general picture presented in figure 4.2 are worth pointing out. First and foremost, insofar as a term, say \(T_1\), is articulated within a linguistic practice, its intelligibility for particular individuals will be a matter of degree. At a bare minimum it must be associated with a Humean representation. But one’s grasp of a term’s intelligibility will depend equally on the degree to which it is possible for them to construct a cognitive model mirroring the terms articulation within the relevant practices. It is not in my view an all or none sort of affair. However, objectively speaking terms either are, or are not, intelligible. *A non-logical term \(t\) is intelligible if and only if \(t\) can be associated with a Humean representation that figures in an observationally based model of the linguistic practice in which \(t\) is articulated.*

Secondly, what makes my representation of a stereo speaker insufficient for rendering ‘coffee pot’ intelligible is that such a representation will, provided it really is a representation of a stereo speaker, have associations with representations that coffee pots lack and lack others besides, which coffee pots possess. It is not, however, the case that representations are rendered deficient merely in virtue of having one or more associations that are not paralleled in the way the term is articulated. It is a pragmatic question concerning the degree of fit required. Some representations in the context of the models in which they figure will be better than others. Representations of penguins may not serve as adequate representations of birds if the goal is to render talk of birds generally intelligible but it will work for some.
Thirdly the structure presented in figure 4.2 is able to explain criticisms of the sort that Leibniz leveled against Newton’s claims about ‘attractive qualities’. From the present point of view, Leibniz’s criticism amounts to the claim there is no Humean representation associated with ‘attractive quality’. [In the context of figure 4.2 this is to maintain that there is nothing in the position of R1—assuming ‘attraction’ is playing the role of T1.] The model contains a “representational blind spot” so to speak. It is not a wholesale critique of the theory, the terms articulation within the theory, or even of the possibility of finding further features of the theory intelligible.

Finally, the model being proposed here is consistent with the sense in which I want to remain semantically agnostic. A model of a theory does not determine the conditions under which the theory is true. That we can construct, cognitively speaking, stick and ball representations of a molecule says nothing about whether there are molecules. Or if there is what they will look like. Nor does the model license inferential moves among statements within the theory—though it may suggest some. Rather, it is a cognitive tool that furthers our ability to render such moves cognitively significant. It does not license them in part because it may suggest inferential connections that are not explicitly warranted within the theory and fail to represent others. I will have more to say about this in the next section yet before moving on there are two further points that need to be addressed here.

The response to the problem of intentionality is as I see it a direct consequence of the apparatus erected here in response to the constraint problem. What makes a particular representation (e.g. balls in the example of CMT in fig. 4.1) a representation of the putative entity, event, or process designated by a term (here ‘atom’) is the fact that the
representation is embedded in a model that is constrained by the practice in which the term is articulated. It gets its intentionality derivatively as it were. Thus, for instance, what makes the various representations associated with ‘molecule’, in nineteenth century molecular discourse, have the same intentional content is their association with a particular linguistic practice.

4.3.0. Boyle, Models, & Theories

While Descartes had, from the point of view of seventeenth-century philosophers, set physics on a surer path by moving it away from explanations in terms of form and matter and toward mechanical explanations, chemistry remained in a somewhat less respectable state. It’s status as a science, or as part of natural philosophy, was at best questionable. The explanation of chemical phenomena had, by the middle of the seventeenth century become a potpourri of appeals to particulate theories of matter, peripatetic elements, and Paracelsian principles. Mechanical explanations in terms of the shape and motion were readily supplemented through appeals to the hypostatic principles of the Paracelsian tri-prima (Salt, Sulfur and Mercury), to spirits, and the composition of substances in terms of the Aristotelian elements (earth, air, water, and fire). Clericuzio maintains that this mélange of views, “was in fact a common pattern of early seventeenth-century atomism, namely an amalgam of atomism and different philosophical traditions.”

In Boyle’s view most of the theoretical positions comprising this mélange of views fell far short of providing adequate explanations of chemical phenomena. The main criticism leveled against these various views by Boyle was that they were unintelligible.

---

The following passage, from his 1674 *Of the Excellency and Grounds of the Corpuscular or Mechanical Philosophy*, is indicative of his attitude:

The first thing, that I shall mention to this purpose, is the intelligibleness or clearness of mechanical principles and explications. I need not tell you, that among the Peripateticks, the disputes are many and intricate about matter, privation, substantial forms, and their eduction, &c. And the chymists are sufficiently puzzled, (as I have elsewhere shown,) to give such definitions and accounts of their hypostatical principles, as are reconcilable to one another, and even to some obvious phenomena. And much more dark and intricate are their doctrines about the Archeus, Astral Beings, Gas, Blas, and other odd notions, which perhaps have in part occasioned the darkness and ambiguity of their expressions, that could not be very clear, when their conceptions were far from being so. And if the principles of the Aristotelians and Spagyrists are thus obscure, it is not to be expected, the explications, that are made by the help only of such principles should be clear."\(^{152}\)

Here Boyle raises a number of issues, not the least of which is his concern with the tendency of some “seventeenth-century chemists,” who, as Marie Boas has put it, “like medieval peripatetics, often became the slaves of semantic expressions, and looked very

little farther than the name, which came to dominate their thinking to the exclusion of what observation could and should have suggested.”\textsuperscript{153}

Additionally, Boyle specifically endorses a connection between \textit{expressions} and \textit{conceptions} according to which the expressions fail to be clear whenever the corresponding conceptions are unclear. Boyle is endorsing CC/CCR and claiming that the expressions merely give the illusion of being about anything real or tangible. This is supported by the metaphor he employs to characterize his attitude,

\begin{quote}
even in some of the more ingenious and subtle of the peripatetic discourses upon their superficial and narrow theories, methinks, the authors have better played the part of painters than philosophers, and have only had the skill, like drawers of landskips, to make men fancy they see castles and towns, and other structures, that appear solid and magnificent, and to reach to a large extent, when the whole piece is superficial, and made up of colors and art, and comprised within a frame perhaps scarce a yard long.\textsuperscript{154}
\end{quote}

According to Boyle often these chemists scarcely had more at hand than the appearance of a coherent language for discussing the phenomena.\textsuperscript{155} The terms they employed to discuss the phenomena, while they may have been clearly articulated, i.e. integrated into a systematic manner of speaking in the face of certain phenomena, they were not associated with specific conceptions of their supposed referents. Their terms were


\textsuperscript{154} “Mechanical Philosophy”, p. 190.

\textsuperscript{155} Though, as we saw in the preceding passage it is sometimes not sufficient even for these purposes as it was not easy to “reconcile” with “some obvious phenomena”.

139
sufficiently articulated to induce belief but, thought Boyle, closer inspection revealed them to be without intelligibilizing content. No attempt was made to ground the explanatory terms on clear conceptions.

In Boyle’s view under these circumstances we would be better off without any account than one that is unintelligible; better to plead ignorance than to utter vacuous sounds. Indeed, Boyle complained in opening *The Sceptical Chemist*,

> I blush not to acknowledge that I much less scruple to confess that I doubt when I do so, than to profess that I know what I do not: and I should have much stronger expectations than I dare yet entertain, to see philosophy solidly established, if men would more carefully distinguish those things that they know from those that they ignore or do not but think, and then explicate clearly the things they conceive they understand.\(^{156}\)

As the foregoing discussion suggests Boyle’s view concerning the importance of intelligibility fits naturally in my model with some key differences. The principle difference being that as far as Boyle is concerned intelligibility is a necessary component of any adequate explanation. A coherent well constituted or articulated theory simple isn’t sufficient. Alternatively, on the model that I am defending it possible for a well articulated theory to be adequate despite its containing terms that are not intelligible. Newton’s explanation of gravitational phenomena is a case in point. These differences aside, the nature of Boyle’s negative remarks concerning his contemporaries are readily explained on the model of intelligibility I have proposed, as I have just shown.

Beyond the general complaint that his contemporaries’ views were obscure or unintelligible, Boyle sought to found chemical explanations on an empirical basis. Being a natural philosopher and an empiricist Boyle was skeptical of *a priori* explanations. Whereas the peripatetics had sought *a priori* explanations based on form and matter (here the four Aristotelian elements), Boyle sought to ground chemistry in a corpuscular theory of matter evinced through chemical experiments. In his view the peripatetics mistakenly thought that it was “much more high and philosophical to discover things *a priore* than *a posteriore*.” He was convinced that this method, like that of many early seventeenth century chemists endorsing hypostatical principles, failed to furnish clear ideas. And his picture of the latter hypostatic chemists is a rather sardonic one, “for they deliver their hypotheses as darkly as their processes; and ‘tis almost as impossible for an sober man to find their meaning, as ‘tis for them to find their elixir.”  

Against this Boyle stressed his desire to see that theories “be intelligibly explicated, and duly proved.”

In Boyle’s view the appropriate starting point for explaining chemical phenomena was found in what he termed the “Catholick Principles.” At the lowest level there were the standard mechanical concepts; matter and motion. Beyond this he held that experimental reasoning lent credence to several additional higher order principles including figure, size, posture, rest, order and texture. Such principles were derived as possible concepts from the basic notions (i.e. matter and motion) of the mechanical philosophy and rendered intelligible together with “matter” and “motion” via analogy with, and extension from, grosser observations of nature. In his view the idea that, “mechanical principles may have a great stroke in the operations of bodies of a sensible

---

157 *Sceptical Chemist*, p. 20.
158 *Sceptical Chemist*, p. 10.
bulk,” but that they could not be equally efficacious in explaining “the hidden transactions that pass among the minute particles of bodies,” was mere sophistry. What Boyle seems to mean when he claims that they are evinced experimentally is that they are efficacious in explaining the experimental results. Boyle’s criticism of the Peripatetic and Spagyrical chemists of his day—as well as his own alternative view of what was required of an adequate explanation—demonstrate that Boyle had a fairly sophisticated conception of the relationship between explanation and intelligibility. On the one hand he argued for explanations in terms of principles consilient with the mechanical philosophy of his day and, on the other, he felt the need to ensure that such principles were themselves comprehensible. At the same time it is clear that he did not view it to be sufficient that an explanation be provided in intelligible terms. He maintained that one had to appeal to reasoning based on controlled experiments in order to be sure that they had chosen the appropriate subset of comprehensible terms.

Boyle’s view of explanation thus incorporates features of each of the three components of Buchdahl’s model. We can see features of the regulative component in the appeal to the mechanical philosophy as an explanatory paradigm, features of the constitutive component in the requirement of empirical adequacy, even though little explicit attention is given to systematic coherence, and finally a explicit appeal to the explicative dimension through the emphasis on intelligibility.

Neither explanatory efficacy nor intelligibility were sufficient to affect an adequate understanding. Both were viewed to be necessary properties of an explanation and the lack of each was variously offered as criticism of alternative accounts. But they

159 Mechanical Philosophy, p. 194.
were not at the same time viewed to go hand in hand. We might, from Boyle’s point of view, indeed have a coherent means for discussing the phenomena without thereby having an intelligible means for doing so and visa versa. Boyle’s concern was to see that theories that purport to explain natural phenomena be both explicated or made intelligible as well as demonstratively efficacious. Boyle’s view reflects the sentiment that mechanical notions provided the appropriate basis for explanation because they were themselves intelligible, whereas the alternatives (e.g. Spagyrical principles) were not.

Boyle’s account of intelligibility bears some similarity to the account adopted by Colin McGinn. McGinn holds that theoretical concepts are rendered significant in virtue of being analogical extensions of macro-level concepts. As McGinn sees it this “principle of homogeneity” (PH) implies, for instance, that we “arrive at the concept of a molecule by taking our perceptual representations of macroscopic objects and conceiving of smaller scale objects of the same general kind.”160 Similarly, Boyle maintains, to say, that though in natural bodies, whose bulk is manifest and their structure visible, the mechanical principles may be usefully admitted, that are not to be extended to such portions of matter, whose parts and texture are invisible; may perhaps look to some, as if a man should allow, that the laws of mechanism may take place in a town clock, but cannot in a pocket watch.161

Both theorists hold that appeals to micro-level or unobservable mechanisms ought to be such that the postulated mechanism can be conceived of on analogy with macro-level

161 Mechanical Philosophy, p. 195.
experiences and sensations. Both theorists however go further. They suggest that insofar as we are concerned with what can be conceived of—that is on their accounts, with regard to what is intelligible, or in McGinn’s language, the concepts available to us—cognitive contents are perceptually closed. McGinn for example maintains that we “must be careful to guard against a form of magical emergentism with respect to concept formation.” It is not simply that we ought not to attempt to employ concepts, the putative contents of which outstrip anything that is perceptually open to us, it is that, try as we might, we can never get beyond the perceptual closure to which our concept forming capacities are subjected. At least we cannot get beyond this in any epistemically significant manner. Such putative concepts have no explanatory value on this view. As Boyle says,

If recourse be had to an immaterial principle or agent… as is not intelligible… it will not enable us to explain the phenomena, …we cannot conceive, how it should produces changes in a body, without the help of mechanical principles.\(^{163}\)

For Boyle and McGinn the idea that a well-articulated theory could be explanatorily adequate is simply misguided. But if this is so, it is only because they take the crucial burden of an explanation to be the provision of an answer to questions concerning the underlying micro-structure. If on the other hand we take adequate explanations to involve not simply or even necessarily explaining how phenomena come about, but in furthering our understanding more loosely, then a formalistic construal of Newton’s theory or the various tables of affinity for chemical substances developed by Geoffroy and others may

\(^{162}\) Ibid.
\(^{163}\) Ibid, pp. 208-209.
pass as explanatorily adequate. This is because these are cases where coherent or systematic theories were provided which offered accurate predictions. And these predictions came about despite the fact that the theories did not provide for intelligible mechanisms. The latter tables, unlike Newton’s theory, were constructed in a positivistic manner without any attempt to describe the mechanism or mechanisms responsible for the known affinities.

While I do not endorse the implication present in these accounts—that the intelligibilizing contents exhaust the conceptual content associated with a term (i.e. I think there is more to cognitive significance than this)—the idea, adopted by both McGinn and Boyle, that the intelligibilizing contents for many abstract terms are extensions from macro-level experiences is in my view on the right track. On the account I am proposing these representations are observationally based models of the putative referents of the terms. The reasons why we should construe these contents as models, rather than as literal representations of the mechanisms or processes designated by the terms with which they are associated, are many-fold. First and foremost, they are not literal representations. They are not copied from, nor do they directly represent, features of our experience. Rather they constitute hypotheses about possible experiences. Boyle’s corpuscles did not designate any impression. Nor did they designate the events or experiences which were causally associated with his postulation of corpuscles. Rather ‘corpuscle’ refers to a type of hypothetical entity that plays a role in the mechanism responsible for various observable surface properties of substances. The intelligibilizing content which we associate with the word cannot be directly copied from experience.
Second, these contents are generally rough if even outright false pictures of the mechanisms referred to by the associated term. This is unproblematic since they are not the principle constituents of the theory. They function instead as vehicles to aid in our cognitive appreciation of the explanations offered up by the theory. This comes out more clearly in the context of some theories than it does in others. Boyle, for example took his notion of a corpuscle to provide a fairly accurate picture of the mechanism. Maxwell on the other hand was clear in his appeal to the fluid model of electricity that the fluid analogy was merely meant to aid in our understanding of the formal mathematical account.

The substance here treated of… is not even a hypothetical fluid which is introduced to explain actual phenomena. It is merely a collection of imaginary properties which may be employed for establishing certain theorems in pure mathematics in a way more intelligible to many minds and more applicable to physical problems than that in which algebraic symbols are used.¹⁶⁴

Beyond Boyle’s psychological stance regarding his conception there is little significant difference in the way early atomistic models, like Boyle’s, functioned in relation to the theory and the way in which Maxwell viewed his fluid model. In neither case do the models themselves explain anything. There is, for instance no reason to believe that atoms with sharp angular figures will cause a taste sensation of bitterness other than the fact that for some corpuscularians the theory independently suggested this. And as Maxwell’s picture suggests it isn’t the model that gets directly tested but the theory. The

atomistic model of nature wasn’t jettisoned because as it turns out the sensation of bitterness isn’t caused by the sharp angular edges of molecules. Insofar as it is the theory as well as the empirical data that the theory is designed to explain that get tested, these contents are not discharged but revised in light of new evidence.

As I have argued previously it is a fundamental error to think of accounts like Hume’s account of intelligibility as theories of meaning. If we do, the familiar reference fixing problem quickly emerges. When I speak or write of, say, Hume the only cognitive content that I have associated with ‘Hume’ will be, roughly, the idea of a man who wrote among other things the Treatise, who held certain ideas, who was an atheist living in the eighteenth century, and so on. It is clear, that this content does not pick out David Hume, eighteenth century Scottish enlightenment philosopher, for that individual had a number of other properties besides those included in my list and moreover may well lack some of those properties or qualities which entered into my representation of him. Nonetheless, I refer to him. The reference of ‘Hume’ in a given context generally extends beyond the intelligibilizing contents that I associate with it.

If this is right then the associated intelligibilizing contents will at best serve as a cognitive model of the referent. This model will likely possess certain inaccuracies. Whatever my representation of Hume might be it will not be copied from any direct experience of Hume. It may for all I know be accurate as far as it goes but the point is that it is not to the idea of Hume that I refer when inquiring into the truth or falsity of statements mentioning him. On the model I have been articulating this fact is easily accommodated. One only needs some cognitive model associated with Hume and that model may be the product of various experiences. The truth or falsity of claims about
Hume is a matter separate from the degree of fit between the model and the world. The objective of the model is not to model the world but the linguistic practice in which statements about the world may be made.

4.4.0. Pragmatic Virtues

The appeal to models is not new. Nor is the idea that such models are usually at least partially inaccurate representations of the hypothetical entities or processes underlying phenomena. Questions of intelligibility arise most pointedly in the context of talk about abstract terms, and generally with regard to those terms which are taken to designate some hidden or unobservable mechanisms or processes (as opposed to those which refer to classes of observable entities such as swans, dogs, etc.) Many nineteenth century chemists, for example, argued for Dalton’s atomic theory precisely because it gave them a sense of what there formulas were representing. Talk of valence and combining ratios seemed empty headed in the absence of any comprehensible way of representing the underlying structure. Atomism provided them with just such a model.

Norman Robert Campbell and Mary Hesse have argued the use of models is essential to the progress of science. Without models Campbell maintains any extension of a theory will be arbitrary.\(^\text{165}\) Hesse, for her part claims that there is an element of truth, “in Campbell’s assertion that without models theories cannot fulfil [sic] all the functions traditionally required of them,… in particular… they cannot be genuinely predictive.”\(^\text{166}\) But what Campbell, and more specifically Hesse, expect of models differs from the role

\(^{165}\) Campbell, Norman Robert (1920) *Physics, the Elements* (London: Cambridge University Press).

that I am proposing that they play. As they see it models constitute an essential component of the theory.

For Hesse, models are analogies. They have positive, negative, and neutral features. The positive features of the model are those which are known to be analogous to the mechanisms or processes being investigated. Thus, if we chose to use the motion of waves in a liquid medium (one of Hesse’s examples) as a model for the propagation of sound, then insofar as we know that the model and the phenomena being investigated share the properties of amplitude and frequency these features of the model are positive features. As well models have negative features. Thus, as sound is not wet (or what have you), the model is disanalogous; wetness is a negative feature of the model. More importantly are those features of the model about which we are uncertain. That is, there will in general be features of the model which, for all we know, may or may not be analogous to the mechanisms or processes under investigation. Such features are said to be neutral.

The presence of neutral features within the model facilitates prediction by suggesting hypotheses which may be further tested in an effort to render these features positive or negative. According to Campbell models are necessary because it is only through these features that theories become truly predictive. A purely formalistic theory is, in his view, at best only ever partially interpreted. This is because its interpretation is restricted to the phenomena that it quantifies over. However, cognitive models of the sort I have in mind serve a somewhat looser pragmatic function. While they do serve to suggest possible explanations and hypotheses and thus offer up research programs, they need not do so as their primary function is again to aid in the cognitive appreciation of
theories. Moreover, while they make predictions, it is not always the case that these predictions are testable. Boyle’s corpuscular model of nature rendered the transmutation of metals plausible, yet there was of course no way to carry out the necessary decompositions of metals at the corpuscular level. Thus, one virtue of Hesse’s and Campbell’s picture of models is that it suggests a manner in which models may serve to facilitate and advance theories.

Another helpful aspect of their picture is that it suggests a natural relation between models and theories. Theories serve to constrain models. It is clear that certain models will simply not be very useful. Newton-Smith provides a particularly striking example.

Once upon a time, many thought that the fact that there were seven virtues and seven orifices of the human head gave them an understanding of why there were (allegedly) only seven planets.

As he notes, “we need to distinguish between real and spurious understanding.”\textsuperscript{167} But making such a distinction in any hard and fast way is not only impossible in practice but unnecessary. On the view I am defending what makes one model preferable or better than another involves an ongoing interplay between the model, the theory and experiment.

4.4.1. Criticism, Rejoinder and Resolution

The idea that models are in some sense necessary if we are to have any genuine understanding has not gone uncriticized. Duhem argued that the use of models can impede scientific progress. “Concerning the procedures employed by the English in dealing with physics,” he wrote,
there is current a commonplace opinion, according to which the abandonment of care for logical unity which was so important with the old theories, and the substitution of models independent of one another for the rigorously linked deductions formerly in use bestow upon the physicist’s inquiries a suppleness and freedom which are eminently fruitful for discoveries.\textsuperscript{168}

Duhem held that this “commonplace opinion” contained “a very great share of illusion” and that the degree to which models aid in the progress of science is “very meager.” He points to the work of Lord Kelvin to illustrate his point. He maintains that Kelvin’s greatest achievements were arrived at by employing abstract systems and not by the use of mechanical models.\textsuperscript{169} Science, thought Duhem, would be better served if it confined itself to subsuming the phenomena under formal systems quantifying the relationships among physical properties. With regard to chemical method he recommended that,

All hypotheses about the intimate nature of material, about the structure of mixtures and chemical compounds, and particularly all atomic hypotheses, would be banned from the domain of science. It would not make use of principles drawn from such hypotheses. If an expression only has as much sense as is admitted explicitly or implicitly, by these suppositions, it is ruthlessly rejected. Alternatively, before adopting it, it is given a new definition, completely uninhibited by the doctrines to which it has been resolved not to make any further appeal. The definitions and propositions

\textsuperscript{169} Ibid, p. 97.
of chemical mechanics will concern, in the last analysis, only quantities which can be physically measured.\textsuperscript{170}

For positivists like Duhem a theories meaning is restricted to determinate observations and measurements. Any theoretical entities mentioned in the theory are either eliminated or reduced to measurable quantities. But such a reduction treats the meaning of theoretical terms as a function of there inferential role. Hempel, for instance, states this more explicitly.

While the testability criteria of meaning aimed at characterizing the cognitively significant sentences by means of certain inferential connections in which they must stand to certain observation sentences, the alternative approach under consideration would instead try to specify the vocabulary that may be used in forming significant sentences. This vocabulary, the class of significant terms, would be characterized by the condition that each of its elements is either a logical term or else a term with empirical significance… From its primitive terms and sentences an axiomatized theory can be developed by means of purely formal principles of definition and deduction, without any consideration of the empirical significance of its extralogical terms.\textsuperscript{171}

The formal theory is subsequently rendered meaningful through being provided with empirical interpretation,


A deductively developed system of this sort can constitute a scientific theory only if it has received an empirical interpretation which renders it relevant to the phenomena of our experience. Such an interpretation is given by assigning a meaning, in terms of observables to certain terms or sentences.\textsuperscript{172}

Statements involving uninterpreted non-primitive terms are meaningful in virtue of the deductive relations they stand in to statements in which all of the extralogical terms have been rendered empirically significant.

As noted previously from Campbell’s and Hesse’s points of view, formulas alone do not predict anything beyond their logical and mathematical implications. Without models formula and equations have only as much empirical significance as can be provided by the phenomena over which they quantify. The formalist seems consequently to be barred from talking about the causal structure of the universe as their equations have merely descriptive value and no causal efficacy. As Hesse sees it the formalist (of which Duhem is an example),

do not give a plausible account of the meaning of theoretical terms. On what [she] take[s] to be the consistent formalist view, the theory in this case consists only of a formal deductive system—marks on paper manipulated according to certain rules\textsuperscript{173}

It should by now be clear that I am going to come down somewhere in the middle of this debate. From the point of view of Buchdahl’s model of theory assessment Duhem places the semantic load on the constitutive component of the model. The only feature of the

\textsuperscript{172} Ibid, pp. 111-112.
\textsuperscript{173} Duhem (1902), p. 21.
theory that is significant in determining the meaning of the formulas and theoretical terms involved is the manner in which these latter are articulated (see sec. 3.2) within the theory. Duhem does not deny that models might constitute significant psychological aids or even that such devices might occasionally be of service to the theory.\textsuperscript{174} They are, however, unnecessary and do not in any case contribute to the significance of the theory. Hesse and Campbell on the other hand view models as essential to our grasping or understanding the significance of the theory and believe that without them theoretical terms are not genuinely meaningful. Consequently, they put the emphasis on Buchdahl’s explicative component. Put in terms of cognitive significance formalists treat the cognitive significance of theoretical terms wholly in terms of the manner in which the terms are articulated within the theory, while Campbell and Hesse treat a terms cognitive significance as primarily a matter of its intelligibility.

I, on the other hand, hold that both play a role in determining the significance of theoretical terms. To the extent that that a theoretical term is well-articulated within a theory the term will be cognitively significant. Moreover, the theory in which it plays a role will not be unintelligible provided some portion of it is empirically adequate. Whether or not the term itself is unintelligible will depend upon our ability to associate it with a suitable Humean representation in the sense described earlier. In the case of formal notions it may not be necessary to associate them with cognitive models. Thus, in my view, if Newton had simply articulated ‘attraction’ in terms of the laws, instead of being tempted to treat it as designating a hidden mechanism responsible for the phenomena

\textsuperscript{174} Add reference.
quantified by the laws, then the complaint concerning its intelligibility would have been less poignant.

One might claim that ‘attraction’ is on this picture a superfluous element of the theory. This may be true. But the point is irrelevant. A consistent formalist can’t holdout for any greater significance. At the same time it seems incorrect to maintain that the term requires more significance—as does Hesse. I take Hesse’s point that ideally we would like to know how the formula are instantiated by the phenomena. We are tempted to take ‘attractive forces’ to designate a mechanism and insofar as we do this the term hasn’t got cognitive content unless we have some cognitive model of the mechanism itself. But I don’t hold, with Hesse, that explanations are disqualified simply on the grounds that they are non-mechanical and merely formal explanations. Newton’s theory was successful despite the inability to provide any mechanistic interpretation of attraction.

In connection with the criterion of intelligibility that I am proposing this discussion is primarily meant to situate intelligibility in a broader framework of cognitive significance. The basic criterion that I want to defend is the Modified Humean Criterion I alluded to earlier:

**MHC:** A non-logical term \( t \) is intelligible if and only if \( t \) can be associated with a Humean representation that figures in an observationally based model of the linguistic practice in which \( t \) is articulated.

However, as previously noted, I want to clearly distinguish MHC as a criterion of intelligibility from a theory of meaning. I am happy to let meaning lie with contextual and referential features of the language. Also, it should be clear that MHC prescribes that
the associated intelligibilizing contents to be understood as models which are not necessarily directly copied from impressions or experiences of the referent.

Finally, intelligibility is not a necessary feature of theoretical terms. This is a seemingly inevitable consequence of theories like Newton’s as well as quantum mechanics and the like. However, I don’t take this to be a shortcoming of the view but a benefit. An account of intelligibility is an empirical hypothesis concerning how we understand language and in particular, language associated with hypothetical entities. Cases like that of the blind poet, as well as the history of science itself, suggest that linguistic understanding is not a simple matter. These cases suggest that we understand language both by grasping how terms are used in relation to another as well by attempting to associate are use of these terms with significant and relevant experiences. The fact that quantum mechanics is not intelligible on one view or another is not a conclusive argument against quantum mechanics anymore than the fact that aspects of Newton’s theory are unintelligible constitutes a conclusive argument against his theory. Ideally we would like for theories to be maximally cognitively significant but this is not always the case.
Chapter V

5.0.0. Introduction

For the better part of the twentieth century the only available criteria of intelligibility came in the form of positivistic principles of meaningfulness. Like their earlier positivist allies, twentieth century logical positivists sought to restrict the domain of meaningful discourse to observable quantities and qualities. In addition to providing and account of the structure of scientific reasoning, one of the central aims was the demarcation of sense from nonsense. Motivating this goal was the positivists shared conviction that metaphysical principles, unlike the claims of science, failed to yield knowledge. And the explanation offered for this epistemic failure was that the claims of metaphysics, again unlike those of science, were unintelligible or meaningless, “at best express[ing] a certain mood or spirit.”\textsuperscript{175} The challenge was to specify in a more principled way how the theoretical claims of science were on any more secure a footing than those of metaphysics.

Officially the goal of the Vienna Circle, announced in the Manifesto, was to develop and promulgate the “scientific conception of the world.” In 1929 this goal amounted to “linking” and “harmonizing” science and the “achievements” of its practitioners, which, in turn, led to the search “for a neutral system of formulae, for a symbolism freed from the slag of historical language.”\textsuperscript{176} And this meant determining meaning through “logical analysis or, more precisely, through reduction to the simplest


\textsuperscript{176} Ibid
Positivists like Carnap thought the central fallacy of metaphysics consisted in its committing logical or grammatical mistakes,

In Metaphysical theory, and even in the very form of the questions, there are two basic mistakes: too narrow a tie to the form of traditional languages, and a confusion about the logical achievement of thought. Ordinary language for instance uses the same part of speech, the substantive, for things (‘apple’) as well as for qualities (‘hardness’), relations (‘friendship’), and processes (‘sleep’); therefore it misleads one into a thing-like conception of functional concepts…”

A. J. Ayer summarizes the positivists’ charge against the metaphysician not as the claim, “that he attempts to employ the understanding in a field where it cannot profitably venture, but that he produces sentences which fail to conform to the conditions under which alone a sentence can be said to be literally significant.” Carnap’s manner of expressing this point is perhaps more perspicuous.

The researches of applied logic or the theory of knowledge, which aim at clarifying the cognitive content of scientific statements and thereby the meanings of the terms that occur in the statements, by means of logical analysis, lead to a positive and to a negative result…. In the domain of metaphysics… logical analysis yields the negative result that the alleged statements in this domain are entirely meaningless.

---

177 ibid, p. 306.
178 ibid, p. 307.
For positivists, like Ayer and Carnap, the significance or meaningfulness of a statement was thought to be grounded in observations. Exactly how observations were to ground various kinds of statements was a matter of some dispute and different views continued to evolve for the next two decades. Nonetheless, throughout this period, positivists remained committed to the basic view announced in the 1929 Manifesto; the reduction of lexical significance to experience via logical analysis:181

Since the meaning of every statement of science must be statable by reduction to a statement about the given, likewise the meaning of any concept, whatever branch of science it may belong to, must be statable by step-wise reduction to other concepts, down to the concepts of the lowest level which refer directly to the given.182

Moreover, “Logical investigation… leads to the result that all thought consists of nothing but a transition from statements to other statements.”183 The “scientific conception” thus analyzes statements purely with respect to their formal relations to other statements down to level of statements which can be given a purely empirical interpretation in terms of observations—that is without concern for their semantic properties. And in fact, three years later, we find Carnap elaborating just this view in Elimination of Metaphysics.

Elimination is interesting for two reasons. First, it contains Carnap’s most concerted attempt to analyze the concept of ‘meaning’ to which he had made, and would continue for the next several years to make appeal. Second, he expresses his affinity with nominalists.

181 Hahn (1929).
182 Ibid, p. 309.
183 Ibid, p. 308.
Carnap begins *Elimination* by notifying the reader that “logical analysis” as a philosophical method has rendered both a positive and negative result. The positive result, Carnap tells us is to be carried out by the empirical sciences. His purpose in *Elimination* is rather to articulate and defend the negative consequence. He sets the stage by asking, “What, now, is the meaning of a word?” In responding to this question Carnap is explicit:

> Since the meaning of a word is determined by its criterion of application (in other words: by the relations of deducibility entered into by its elementary sentence-form, by its truth conditions, by its method of verification) the stipulation of the criterion takes away one’s freedom to decide what one wishes to “mean” by the word.

Like Hempel, Carnap treated the cognitive significance of terms as a matter of the inferential relations among statements together with the empirical application of the language. In order to determine the significance of theoretical terms one had to specify a language consisting of a set of observational terms (termed the observational vocabulary or $V_0$) with which observation statements were composed. These statements were interpreted directly in terms of observations. All other extralogical terms (the theoretical vocabulary or $V_T$, as well as the statements involving them) derive their cognitive significance from the deductive relations they stand in to this observational basis. It is

---

184 This is, of course, hardly an innocuous claim as it suggests that the only legitimate philosophical enterprise consists in articulating the formal properties of so-called empirical theories—which of course leaves one to wonder about the semantic status of the formalizations.


186 Ibid, p. 63. This idea, that logical analysis will and ought to take away one’s freedom to choose what they mean by a word, is one that will become a key focus for us later on.
this basic structure which grounds the positivists’—perhaps oversimplified—principles of
significance such as the so-called *verificationist principle* according to which *the
meaning of a statement is its method of verification*. In order to verify a theoretical
statement one derives and tests its observational consequences.

This view bares a strong relation to earlier positivistic accounts like Duhem’s.
Recall Duhem’s recommendation concerning how Chemistry ought to proceed:

All hypotheses about the intimate nature of material, about the structure of
mixtures and chemical compounds, and particularly all atomic hypotheses,
would be banned from the domain of science. It would not make use of
principles drawn from such hypotheses. If an expression only has as much
sense as is admitted explicitly or implicitly, by these suppositions, it is
ruthlessly rejected. Alternatively, before adopting it, it is given a new
definition, completely uninhibited by the doctrines to which it has been
resolved not to make any further appeal. The definitions and propositions
of chemical mechanics will concern, in the last analysis, only quantities
which can be physically measured.\(^{187}\)

Both theorists fall into the category of those who see a term’s cognitive significance to be
wholly determined by what is, on the model I am defending, the term’s articulation.
Unlike Duhem, however, logical positivists had more room to maneuver and could
accommodate talk of atoms or corpuscles or valency properties. As Duhem saw it these
hypothetical entities were to be simply ruled out in favor of meaningful talk in terms of
measurable quantities. Carnap and others, as we shall see in some detail below, however

\(^{187}\) Duhem (1902), p.113.
could permit talk about corpuscles and the like provided these were, like Maxwell’s fluid, understood as collections of properties that were in principle observable. One could in principle look and see if corpuscles really come in different shapes. One could not however “look and see” whether, say, monads constitute the fundamental substratum of reality.

There is, however, a problem with this picture. At least for some theoretical terms the cognitive significance afforded them under the positivistic model fails to line up with the cognitive significance they actually have. Take a claim like ‘Hydrogen is monovalent’. Independent of a model of atomic structure we might assent to this claim because of various tests that we have performed, examining the proportions in which it combines with other substances, etc. For positivists the significance of this term is wholly reducible to the observations we make. But this underdetermines the significance of the term since we intend at least in some contexts to refer to the mechanism or property of atoms responsible for valency.

Analogously, consider a situation in which we had never seen an airplane, but in which we had on various occasions seen vapor trails. And suppose that we wanted to explain how these vapor trails occur. Driving along some day we noticed the trail of exhaust left by a poorly maintained automobile. At this point we hypothesize that someone somewhere has constructed some sort of vehicle that can travel in the air and that it is these vehicles that are responsible for the vapor trails we observe. Now we say things like ‘Look there is an airplane!’ or ‘There must be two airplanes there’ and so on. What is the significance of ‘airplane’ on the positivistic account?
On the positivistic account ‘airplane’ has no more significance than can be garnered from our observations. It does not have the significance suggested by the model. Of course the model itself may provide a means of rendering ‘airplane’ intelligible even on positivistic grounds provided that it affords us observational tests with which we could judge the truth of statements involving the term. But it does not contribute to the significance of ‘airplane’ discourse in the example. The model is not entailed by any of our observations. Rather ‘airplane’ ends up being a “manner of speaking” in the face of certain phenomena. The problem of course is that if we want to say what it is that makes ‘airplane’ intelligible, it is not its bare association with the phenomenon of vapor trails. Not only do positivists fail to recognize the contribution the model makes to the term’s overall significance, they fail to appreciate the manner in which the ability to construct models may contribute to our understanding of the phenomena. The airplane model may for instance spur us to construct physical models that can reproduce the vapor trails or attempt to shoot one of these hypothetical entities out of the sky and so on.

I’ll return to these worries latter but first I want to examine a somewhat similar line of criticism directed at the positivists and distinguish this critique from my own. Two theorists in particular have targeted the positivists’ criteria of significance. The first was Hilary Putnam in 1962 and the second was Peter Achinstein in 1968. They were thought to have collectively placed the final nail in the positivist coffin. I however will argue, focusing particularly on Putnam’s 62’ critique that the criticism misses the target. My primary motivation stems from the fact that I believe there is indeed more to the positivists’ account of significance than often meets the eye and again I will pick up this line of thought later in the essay.
5.1 Contra-Positivism: Putnum’s Critique

While Carnap held empirical data to be given as observable, he viewed the choice of an observational vocabulary \( (V_0) \) itself to be a matter of conventional assignments of observational terms to the given data. In turn, he construed the meaningfulness of statements involving theoretical terms as a function of the relation they stand in to the observational vocabulary. In doing so he relied on a distinction between ‘observation terms’ and ‘theoretical terms’—the so-called observational/theoretical distinction.

Critics of Carnap typically take him to view the observational/theoretical distinction to be factual. That is, they take him to be claiming that there is as a matter of fact an epistemically privileged class of terms that refer to, or have as their content, certain observable features of the world. In addition to this privileged class of terms there is another, somewhat less privileged class of terms which do not refer to observable features of the world but whose content is nonetheless logically related to the content of terms in the former class via reduction pairs. On this view Carnap’s observational-theoretical term distinction is assumed to be an effort at demarcating a natural division among terms that is fixed, hard and fast.

For example, in his 1962 essay, “What Theories Are Not,” Hilary Putnam reports that Carnap’s distinction is “intended as an explicative and not merely stipulative one.” He goes on to clarify, claiming that “the words ‘observational’ and ‘theoretical’ are not having arbitrary new meanings bestowed upon them; rather, pre-existing uses of these words… are presumably being sharpened and made clear.”

Peter Achinstein conveys a similar view, claiming that Carnap’s distinction “is founded upon a prior classification of

---

terms as theoretical or non-theoretical,” and is ultimately grounded in a fixed criterion of ‘observability’. Achinstein’s general criticism is that there is no suitable criterion of observability upon which to base the desired distinction. Nonetheless, here and elsewhere, Achinstein presumes that Carnap viewed ‘observation’ to be an unproblematic and natural cognitive category. As well, Achinstein accepts Putnam’s interpolation that Carnap is not using these terms in any unusual way. Achinstein and Putnam are thus of a similar mind in claiming that Carnap’s distinction purports to capture a distinction natural to language; they also share the view that no such distinction can be made.

However, Carnap’s critics weren’t always of the same mind about what is wrong with his distinction. Putnam’s criticism, unlike Achinstein’s, does not rest on the view that there is no coherent notion of ‘observation’ available, but rather that no linguistic categories can be demarcated on its basis. This is a very different criticism than Achinstein’s, which argues instead that there is no workable notion of observation upon which to base a concrete distinction. Both Putnam’s and Achinstein’s articles merit careful attention. Indeed, Frederick Suppe went as far as to claim that the papers constituted the “coup de grâce for the observational-theoretical distinction.” Here I focus exclusively on Putnam’s 1962 argument, aware that more may need to be said to address Achinstein. I maintain that Carnap’s confirmationist semantics does not presuppose that the observational-theoretical term distinction be made “hard and fast” and indeed, at many points, militates against such a view. Critics of logical positivism

---

190 See for instance, Achinstein (1965).
192 Suppe, Frederick (1977).
have often misrepresented this point. Concerning Carnap at least, the revolutionaries in Suppe’s coup took down a straw man.

Putnam forwards the strong thesis that the observational-theoretical distinction endorsed by Carnap is “in fact, completely broken backed.” He goes on to make four claims that are intended to support this thesis:

(A) If an ‘observation term’ is one that cannot apply to an unobservable, then there are no observation terms.
(B) Many terms that refer primarily to what Carnap would class as ‘unobservables’ are not theoretical terms; and at least some theoretical terms refer primarily to observables.
(C) Observational reports can and frequently do contain theoretical terms.
(D) A scientific theory, properly so-called, may refer only to observables. (Darwin’s theory of evolution, as originally put forward, is one example.)\(^{193}\)

(A)-(D) do not, however, constitute adequate grounds for rejecting Carnap’s observational-theoretical distinction. Concerning (A), Carnap did not think of observation terms as terms that only apply to observables. (B) and (C), taken together, misconstrue the point of Carnap’s actual distinction. Given that they are directed at a different issue they do not constitute a problem for Carnap. Finally, (D) is irrelevant to the issue. Even if we take Putnam’s (doubtful) assertion about Darwin’s “theory of evolution” to be correct, there is nothing in Carnap’s distinction, or inherent in positivism, that precludes its possibility. In fact, as we saw previously positivists like Duhem maintained that meaningful theories ought not to refer to unobservables at all. It is difficult to see what

Putnam had in mind in posing this as a criticism of positivists. In any case, Putnam provides no further exposition of (D) and I shall likewise have nothing further to say concerning it.

There is an explanation for why Putnam misrepresented Carnap’s distinction. Putnam draws his understanding of it from two passages in Carnap’s 1956 essay, “The Methodological Character of Theoretical Concepts”, from which Putnam concludes that Carnap, “thinks of observation terms as corresponding to qualities that can be detected without the aid of instruments.”\textsuperscript{194} The two passages are:

(1) the terms of $V_0$ [observational vocabulary] are predicates designating observable properties of events or things (e.g., ‘blue’, ‘hot’, ‘large’, etc.) or observable relations between them (e.g., ‘$x$ is warmer than $y$’, ‘$x$ is contiguous to $y$’, etc.).\textsuperscript{195}

(2) The name ‘observation language’ may be understood in a narrower or in a wider sense; the observation language in the wider sense includes disposition terms. In this article I take the observation language $L_0$ in the narrower sense… An observable property may be regarded as a simple special case of a testable disposition: for example, the operation for finding out whether a thing is blue or hissing, or cold, consists simply in looking or listening or touching the thing, respectively. Nevertheless, in the reconstruction of the language it seems convenient to take some properties for which the procedure is extremely simple

\textsuperscript{194} ibid, p. 242.

(as in the examples given) as directly observable, and use them as primitives in $L_0$.\textsuperscript{196}

Beyond these two passages Putnam claims he can find no “other clarifying remarks” from Carnap that help elucidate the observational-theoretical term distinction.\textsuperscript{197} Nonetheless, he attributes to the Carnap of “Testability and Meaning”\textsuperscript{198} the view that observation terms designate the observable qualities of objects and that they cannot by definition be used to refer to unobservable qualities. He admits that he can’t “find any explicit statement” on this latter point but still maintains that “Carnap must be neglecting the fact that all terms—including the ‘observation terms’—have at least the possibility of applying to unobservables.”\textsuperscript{199}

Putnam’s own view of the distinction is not difficult to see. His understanding relies on a dichotomy between the observable and unobservable qualities of bodies. He does not deny that such a dichotomy exists. He endorses it, claiming that he does not deny the intelligibility or use of observation reports in science. Rather, he takes issue with the idea “that the distinction between observation reports and, among other things, theoretical statements, can or should be drawn on the basis of vocabulary.”\textsuperscript{200} In other words, on Putnam’s view there are as a matter of fact observational as well as non-observational qualities of bodies. What he maintains is that no observational-theoretical

\textsuperscript{196} ibid. p. 63.

\textsuperscript{197} For anyone familiar with Carnap’s corpus and his persistent efforts at clarifying the observational-theoretical term distinction, Putnam’s comment is mystifying. Carnap devoted an entire section of “Testability and Meaning” to the notion of an observable predicate—a section that Putnam makes no reference to, despite the fact that he explicitly attributes the view he extracts to Carnap (1936).

\textsuperscript{198} Carnap, Rudolph (1936). “Testability and Meaning,” \textit{Philosophy of Science}.


\textsuperscript{200} Ibid, p. 244.
term distinction, or at least not the one on offer from Carnap, will parse language in a manner that maps onto this factual basis. (A) through (C) emphasize this point. Putnam claims that there will be elements of $V_0$ that refer to unobservable qualities as well as elements of $V_T$ (the theoretical vocabulary) that refer to observable qualities.

One consequence of Putnam’s interpretation is that Carnap must hold $V_0$ to be fixed. The assumed goal of the distinction is to capture all and only those terms that refer to the natural category of observables. That is, Putnam sees Carnap as endorsing the antecedent in (A). Given that the terms of $V_0$ form the semantic bedrock of our language (and of any sub-language therein) they must be fixed by their reference to the observable qualities of things and thus cannot be used to refer to any unobservable qualities. This interpretation comes out clearly in the “decisive argument” that Putnam offers to elaborate (A). Putnam claims,

There is not even a single term of which it is true to say that it could not (without changing or extending its meaning) be used to refer to unobservables. ‘Red’, for example was used by Newton when he postulated that red light consists of red corpuscles.201

This view of the distinction underwrites all four of Putnam’s assertions in (A)-(D).

Yet, to what extent do Putnam’s remarks pose a challenge for Carnap’s actual project? Carnap and Putnam do agree on one important point. Unlike Achinstein, both accept the notion of ‘observation’ (as distinguished from observational terms). Putnam’s criticism, again, is that no set of terms can be selected which refers exclusively to observable entities and qualities (and therefore theoretical terms cannot be explicated in

terms of an observational vocabulary.) Carnap, however, treats ‘observation’ as a basic term in the development of his semantics. All other terms of the theory are in turn explicated in terms of ‘observation’.\textsuperscript{202} Having announced this foundation, Carnap next provides an explanation (note: not a definition) of the notion of an observation term:

\textit{Explanation 1.} A predicate “P” of a language L is called observable for an organism (e.g. a person) N, if, for suitable arguments, e.g., “b,” N is able under suitable circumstances to come to a decision with the help of a few observations about a full sentence, say “P(b),” i.e., to a confirmation of either “P(b)” or “¬P(b)” of such a high degree that he will either accept or reject “P(b)”\textsuperscript{203}.

The first feature that leaps out of \textit{Explanation 1} is the fact that nothing Carnap says here suggests (or implies) that observable predicates or terms cannot be used to talk about unobservable entities or qualities. Carnap does not view his project to be that of constructing $V_0$ such that its elements refer exclusively to what is observable. Rather, what is claimed is that being confirmable by observations is a sufficient criterion for a term’s being observational. Such observations fix the meaning of P but they do not determine the range of P’s application (i.e. the arguments that it may take.) It may well be that some arguments fail to be “suitable” (e.g. those that refer to unobservable space-time points.) This point can be illustrated by looking at Putnam’s “decisive” counterexample concerning “red corpuscles.”

\textsuperscript{202} Actually, Carnap accepts two terms as basic; the other is “realizable”. This latter term, though important for Carnap’s overall project, is not directly relevant to the present issue. Thus, for simplicity we have left it out.

\textsuperscript{203} Explanation 2 concerns “realizable” predicates and is left out of the discussion here (see previous note).
The force of Putnam’s counterexample rests on the presumption that since from the standpoint of early eighteenth century natural philosophy or physics “corpuscles” designate unobservable entities, any term predicated of them will necessarily have that fact as part of its meaning and thus will not, strictly speaking, be an observational term. Consequently, Putnam reasons, ‘red’ is not an element of \( V_0 \). Putnam is wrong. The question in dispute is the semantic status of the class of terms. How are we to understand Newton’s explanation of the observed redness of a body in terms of its corpuscular micro-structure? Carnap’s answer is that we understand Newton’s claim because, among other things, we know what observations would confirm it provided we could make them. The fact that we cannot make them in this case does not entail that “red” has a meaning beyond its observational significance. To accept Putnam’s argument we would have to acknowledge that we have no idea what might confirm Newton’s claim—and this of course is not something that Newton believed.

So much for the logic of Putnam’s “decisive argument.” Beyond the fact that Carnap allowed for the possibility that observational terms might be predicated of unobservables, he also explicitly denied that the observational vocabulary was fixed. Instead, he maintains that the construction of a suitable language involves the \textit{pragmatic} selection of primitive terms.\textsuperscript{204} Moreover, Carnap explicitly rejects the notion, implicit in Putnam’s critique, that the standard of what counts as a term in \( V_0 \) is given by the ordinary, or possible, uses of the term outside the constructed language. On Carnap’s model a scientific language \( L \) is a deliberate construction wherein \( V_0 \) is conventionally fixed. Putnam’s criticisms, on the other hand, derive their force from the assumption that

\textsuperscript{204} Carnap (1936), pp. 168-175.
Carnap’s observational-theoretical term distinction is intended to parse the vague and ambiguous speech of ordinary natural languages. For how else are we to understand Putnam’s claim that ‘we also observed the creation of two electron-positron pairs.’ counts as an observational statement for the simple reason that such and such a thing was claimed to have been observed?

Carnap’s aim is rather to construct a language system $L$ (of which $V_0$ and $V_T$ are components) suitable for conducting scientific inquiry and to carry out this construction in a way that ensures that the sentences of $L$ are meaningful. Such a project is clearly different from, and even at odds with, the idea of demarcating a pre-existing distinction in a natural language. Thus, Carnap claims, “We may construct $L$ in whatever way we wish. There is no question of right or wrong, but only a practical question of convenience or inconvenience of a system form, i.e. of its suitability for certain purposes.” Among the practical questions that arise in the construction of $L$ is our selection of $V_0$ as a basis for $L$. According to Carnap, our choice will be influenced by who we wish to include in the linguistic community, our physical constitution, and the like. But whatever decision we make, it will not include or translate every statement of a natural language.

Moreover, as is clear from Carnap’s use of ‘observational predicate,’ he does not intend this expression to be synonymous with the vague and often elliptical use of

---

205 Putnam (1962), p. 244.
206 On this view, were it Carnap’s, we should be compelled to say that all sorts of statements are true because they were, so to be speak, “observed”—or, directly confirmed. Thus, if Jones reports that he “observed a barn in the meadow”, we ought to conclude that there is a barn in the meadow even though Jones bases his claim on the fact that there is a certain well constructed barn façade in the meadow. Of course this is wrong. Some things we claim to observe in our vulgar way of speaking simply aren’t observable, properly speaking.
“observation” in ordinary parlance. Strictly speaking, the claim that ‘we observed the creation of two electron-positron pairs’ is an elliptical way of saying ‘Two electron-positron pairs were created…’ What makes this statement meaningful is that it is analyzable down to the level of observational predicates. Yet “the creation of an electron-positron pair” will not be a predicate of $V_0$ for precisely the reason alluded to by Putnam. Such predicates, like “corpuscle” in eighteenth century physics, do not refer to observable bodies.

Putting the point this way may mislead. Consider another example of Putnam’s. He claims that “satellite” is a theoretical term and draws our attention to the fact that the term typically refers to observable objects. Carnap will grant this while pointing out that the fact that the thing referred to is observable (e.g. that satellite) does not make every predicate we attach to it an observational predicate. To take an even more mundane example, to ascertain whether or not an object is a clock may require us to make a range of observations, none of which individually constitutes observing it to be a clock.

Putnam (like Achinstein) incorrectly surmises that Carnap is trading in ordinary language. In fact Carnap is proposing a theory of meaningful scientific discourse that is partly descriptive and partly normative. In the next section I will examine the merits of Carnap’s account (or the positivist account more generally) and return to the worries I raised toward the end of the last section. I end this section with a brief summary of how we see Carnap replying to Putnam’s (A)-(C):

A. True enough, but Carnap’s account does not assert or imply that observation terms cannot apply to unobservable features or properties.
B. There are several problems with this claim. First, it presupposes a class of terms that might suitably be called “theoretical” for which Putnam, unlike Carnap, gives no account. Second, Putnam’s only example (“people too little to see”) is not a case in which the relevant predicates are non-observational. Rather, it is a case in which an observational predicate is supplied with “unsuitable” arguments.

C. The fact that S purports to “observe X,” where “X” involves a theoretical term Q, does not entail that “Q” is an observational term.

5.2.0. Contra-Positivism: Critique from the standpoint of MHC

Like Putnam I see the confirmationalist semantics offered by positivists like Carnap as problematic. Unlike Putnam my worry does not rest on any dispute over the status of the observational-theoretical term distinction. For even if it is possible to make this distinction the resulting picture of the cognitive significance of many terms, most importantly theoretical terms, will be misrepresented. Whereas the original Humean criterion of intelligibility, when construed as an account of cognitive significance, fails to recognize the importance of a term’s overall articulation, positivistic views determine a term’s cognitive significance exclusively in terms of its articulation within an observationally interpreted theory-structure. As a consequence, the class of cognitively significant terms on the positivists’ model of will include terms which are classical examples of terms which were historically thought to be metaphysical and hence unintelligible. This is the main criticism that I aim to level against the positivists’ account below. If I am right the positivists’ account of intelligibility is overly broad.
Additionally, as I mentioned earlier, according to the positivistic account models of the mechanisms underlying phenomena contribute to the cognitive significance of terms only if they provide observationally testable interpretations of them. Typically, many of the tests that would have to be made to verify statements involving problematic terms can be made only in principle and not in practice. This in and of itself is not a particularly damning criticism since mechanistic models are not necessary on this view. Again, the significance of terms is sufficiently captured by their articulation within a theory together with its empirical significance. This, however, raises two worries for me.

First, this picture of significance seems to suggest that models will play a fairly diminished role in determining a terms overall cognitive significance, arguably performing little more than a heuristic function. If this is right and assuming that the criterion of significance offered can be fairly taken to be an account of intelligibility, then it gets intelligibility all wrong. This criticism is closely connected with the main worry stated above and will be argued in conjunction with it below.

My second worry concerns the nature of the acceptable models. Given that the problematic cases are generally ones in which the theoretical terms under consideration designate hypothesized mechanisms and given the speculative nature of the model itself, the connection between the model and observation reports will seemingly be tenuous at best. Moreover, occasionally there is not a single model of the hypothesized mechanism, but a cluster of models which may be incommensurate with each other as well as the experimental data as a whole. In such cases the model can not be understood as being entailed by or entailing the observational data that it is hypothesized to explain. Strictly
speaking this is more of a worry than a criticism. I will nonetheless briefly illustrate the worry later on.

I am not arguing that terms like ‘atom’, ‘corpuscle’, ‘affinity’, ‘valence’, ‘luminiferous ether’ etc, in their appropriate scientific contexts, turn out to be unintelligible on the account under consideration. Rather, it is significant that they all do in fact turn out to be intelligible on this account, and that terms like ‘substratum’, ‘vital force’, ‘monad’ and the like do not. Nor am I saying that the present account misconstrues what it is about at least some of the terms on the first list that makes them intelligible. Logical positivists judged terms like ‘atom’ to be significant only because they could be associated with observations that could in principle be made. One could in principle look and see whether atoms come in different shapes, arrangements, textures, and so on. We could in principle look and see whether these properties are experimentally related to the observational data in the manner described by the theory.

On the other hand terms like ‘attractive force’ or ‘affinity’ have what significance they do only in virtue of the inferential relations they stand in to direct experimental observations. There is no model of the mechanisms these terms were variously taken to refer to. This isn’t to say that there couldn’t have been a model. As noted, Newton had in fact proposed a corpuscular model for the mechanism of attraction. But this model doesn’t contribute to the significance of ‘attraction’. ‘Attraction’ is rather understood as a quantifiable property of bodies in relation to one another. Its significance is determined by the phenomena it explains and predicts and not, as in the case of corpuscles, by an independent representation of it. As far as I can tell, Newton’s own corpuscular model of
gravitational phenomena does not, from the point of view of the positivistic conception, make any contribution to the significance of the term.

We can accordingly distinguish two ways in which a term might be judged to be significant on the positivists’ account. These will not, however, be two separate criteria but cases of the more general criterion according to which a term’s significance is determined by the inferential relations that it (or more properly speaking the statements in which figures) stands in to observations (or observational reports.) Since I want to make use of the distinction later on it will be use to restate the principle disjunctively.

(Positivist Criterion): A term is cognitively significant just in case it is either (i) inferentially associated with experimental observations, or (ii) characterizable in terms of a set of “in principle” observationally significant traits.

The latter disjunct is of course an approximation of MHC. That logical positivists accepted the latter condition is hardly debatable. It is a straightforward consequence of the more general principle since whatever observational properties serve to define the term will be deductively related to it. In any case, at least some positivists clearly endorsed the idea. Hempel for instance writes,

Theories about the elementary particles constituting the atomic nuclei of various elements, or about the molecular structure of the genes, are presented as accounts of the actual structure of the systems in question, and not just as analogical models. Like any other theory in empirical science, such microstructure theories are put forward “until further notice,” i.e., with the understanding that they may have to be modified or
completely withdrawn in light of subsequently discovered unfavorable
evidence.\footnote{208}

The acceptance of (ii), however, raises questions. If models play a role in determining the
cognitive significance of terms like ‘atom’ then why don’t models more generally,
models like those defended by Hesse and others, play a role in the determination of a
terms cognitive significance. For at least some models the answer is clear enough.
Models like Maxwell’s fluid model are in fact merely analogies and their explanatory
value resides strictly in the fact that, and to the extent that, they instantiate the laws
articulated in the theory. They are not strictly speaking representing explanatory
mechanisms.\footnote{209} But why don’t models like Newton’s corpuscular model of ‘attraction’
apparently contribute to the significance of that term? It is notable that when Hempel
discusses the view he doesn’t mention the corpuscular interpretation.\footnote{210} This is
presumably because it contributes nothing of value to the explanatory force of Newton’s
gravitational theory. While one has to be cautious about conflating the positivists’
account of explanation with that of significance it appears that the primary determination
of a term’s significance is given by its satisfaction of (i) and not (ii). And this will, I
contend be generally true since, provided (i) is satisfied, mechanistic models will be
unnecessary and only useful to the extent that they make an essential contribution to the
explanation. This picture is reinforced by Hempel’s account of what is of primary
importance in explanations.

Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. (New
\footnote{209} Ibid, pp. 433-437.
\footnote{210} Ibid, 433.
Hempel maintains that, “it is covering laws or theoretical principles that are crucial to scientific explanation.” Terms like ‘vital force’ or ‘entelechy’ fail to be explanatory because the putative explanations that make use of them, do not specify under what conditions a vital force will exert its influence and what specific forms its manifestations will take, nor, in the case of external interference with an organism, to what extent the entelechy will compensate for the resulting disturbance.

Once a term has been situated within the framework of a scientific theory complete with a set of experimental observations and explanatory laws, any preexisting models cease to contribute to the term’s overall significance. What is important is the manner in which the term designating the mechanism is logically related to the phenomena via the theory. Models of the mechanism itself are dispensable. This gets the story backwards. The worry over intelligibility was not historically a worry over whether or not abstract terms were logically related to observational consequences. Again I note that even after conceding that Newton’s theory was predictive Leibniz’s worry over the intelligibility of ‘attractive force’ persisted. We can account for this fact if we understand intelligibility to reside in the model, as I contend, rather than logical connections between the hypothesized mechanism and testable consequences. Without a physical for ‘attractive force’ the term remains unintelligible. However, it is cognitively significant for precisely the reasons suggested by the positivists, i.e. because it is a well articulated empirical concept.

---

211 Ibid, p. 433.
212 Ibid.
In contrast, for positivists, models contribute little or not at all to the cognitive significance of a term in its explanatory context. That is to say that a term may be independently intelligible because of its association with a physical model but that this does not contribute to its meaning or significance within the theory. The decreased value of models has the consequence that, when taken as an account of intelligibility, the positivists’ criterion of significance is unable to explain the key historical cases where real concerns were raised. To see this more clearly we can distinguish between the sort of observations mentioned in (i) and (ii) above. Call these type I and type II observations respectively. Type I observations accordingly consist of the empirical data as well as the empirical predictions made by a theory. Type II observations include any observable property whatsoever. Speculative models about the nature of theoretical entities often involve specifying the constitution of the entity in terms of analogies between the hypothesized entity and type II observations. Hence Newton’s “red corpuscles” or Boyle’s “Catholick Principles.” The difference here is not a difference in kind but rather a difference regarding the relations that the respective observations stand in to theoretical terms.

Unlike type I observations, type II observations are meant to give us a “picture” of the mechanism or theoretical entity designated by the term. And this project does not go hand in hand with the project of articulating terms in the context of the empirical data (type I observations). Presently, then, I will argue the latter project does not suffice to render a term intelligible.

Consider ‘affinity’. ‘Affinity’ in the eighteenth and nineteenth centuries was a term on par with ‘attraction.’ A “force of affinity” was postulated to account for the
relative tendencies of substances to combine with other substances. Like attraction the mechanism underlying the affinities of substance was unclear. As Partington notes, “The physical authors who attempted to explain the nature of affinity usually invoked obscure forces.” This however did not halt chemists’ efforts at understanding affinity. Geoffroy’s tables were an attempt to order substances in terms of their relative affinities. And this project was pursued throughout the eighteenth century. Moreover, the study of affinity was carried out with methodical experimentation. In 1783 Bergman provided the following principle for determining the relative affinities of substances:

Suppose A is a substance for which other substances a, b, c are attracted (appetunt). Suppose that A combined with c to saturation (which I denote Ac) should on addition of b tend to unite with it to the exclusion of c, then A is said to attract b more strongly than c, or to have a stronger elective attraction for it. Lastly, let Ab on the addition of a have it’s first union relaxed and b ejected, a taking its place, then it follows that a exceeds b in attractive power and we have the series a, b, c in respect of efficacy. What I here call attraction others call affinity.

‘Force of Affinity’, as illustrated by this passage and throughout Bergman’s 1783 work, was logically related to observations. The conditions “under which it would exert its influence” were specific, and the manner in which it would “manifest itself” was clear. It was equally part of a predictive theory. It predicts for instance that if portions of b and c, with the respective forces of affinity for A given in Bergman’s example are introduced

---


into a solution of A that the amount of Ac created will vary according to the proportion of A and b present in the system while the amount of Ab will depend only on the amount of A and b and not c (assuming the temperature remains constant.)²¹⁵ Of course the theory varies in different hands and this is only a crude statement of Bergman’s theory. Bergman in fact connects affinity with additional variables such as temperature and so on. Nonetheless, the case as it stands suffices for my point.

‘Affinity’ in Bergman’s use of it is a classical theoretical term which is, on any of the later positivistic criteria, cognitively significant in virtue of the deductive relations it stands in to type I observations. I contend however that its association with type I observations did not capture its full significance and that given its full significance it was viewed by some to be a characteristically metaphysical notion. Most individuals thought of ‘affinity’ as a power or force residing in substances and this fact formed part of the terms cognitive significance. What specifically this force consisted of was, as noted by Partington, generally obscure. That is to say ‘affinity’ had no clear association with type II observations and was viewed by some to merely designate a mysterious metaphysical quality of substances.

A century later, upon reviewing the history of affinity, J. W. Langley, in his address before the American Association for the Advancement of Science, announced,

The word ‘affinity’ is in bad odor. We see how enormously complicated the phenomena of chemical action have become, and we have lost faith in

²¹⁵ Of course it is also assumed that the temperature does not differ significantly from the temperature at which the ordering of the respective efficacies of the substances was made.
hypotheses which can be evolved by the mere force of metaphysical speculation.\textsuperscript{216}

The implication is clear. Regardless of its type I associations the notion of a ‘force of affinity’ remains empirically suspect. The failure of ‘affinity’ to meet the criterion specified in the positivists’ second disjunct (ii) renders it metaphysically suspect despite the fact that it satisfies the (i). This supports the view that critiques concerning the cognitive significance of terms can be pushed along different dimensions corresponding to the constitutive and explicative aspects proposed by Buchdahl and that in fact they were. That these two dimensions have historically been the subject of scientific research is further evidenced by the continuing evolution of the theory of affinity.

The middle of the nineteenth century saw the first formulations of the theory of atomicity or valency. The first statement of the new theory was given in 1852 by Frankland.\textsuperscript{217} In order to account for the proportions in which chemical “affinities are best satisfied” Frankland offered the following “law”:

Without offering any hypothesis regarding the cause of this symmetrical grouping of atoms, it is sufficiently evident,… that such a tendency or law prevails, and that, no matter what the character of the uniting atoms may

\begin{footnotesize}
\begin{itemize}
  \item \textsuperscript{217} Kekule’ is often thought to have originally developed the view and indeed claimed this for himself, a claim that was subsequently repeated by others. That Frankland had actually first pronounced the view was clarified in 1878 by Frankland himself in the \textit{American Journal of Mathematics}, (“Extract from a Letter of Dr. Frankland to Mr. Sylvester,” \textit{American Journal of Mathematics}, Vol. 1, No. 4 (1878), 345-349), and in a previous “Note” in the same journal, (“Historical Data concerning the Discovery of the Law of Valence,” \textit{American Journal of Mathematics}, Vol. 1, No. 4 (1878), 282). This is also supported in Partington’s account of the history, (Partington, (1961-1965). Vol. III, pp. 535-537.).
\end{itemize}
\end{footnotesize}
be, the combining power of the attracting element, if I may be allowed the
term, is always satisfied by the same number of these atoms.\textsuperscript{218}

While Frankland seems here to be disinclined to speculate concerning the nature of the
cause or mechanism of the phenomenon it is clear that it remains a principle concern of
the research. The language used here itself suggests that he views ‘atomicity’ to designate
an essential material property of atoms. Later in the same publication he identifies the
cause with the “electro-chemical character” of atoms. Unlike earlier views which
accounted for the chemical properties of compounds in terms of their structure the theory
of atomicity held that the “nature of a body” is, “essentially dependent upon the electro-
chemical character of its single atoms, and not merely upon the relative position of those
atoms.”\textsuperscript{219} Rather than giving up the notion of a ‘force of affinity’ the theory of atomicity
or valence is an effort at clarifying how this force is exerted, the suggestion being that the
atomic mechanism responsible for an atoms electro-chemical properties is the mechanism
instantiating ‘forces of affinity’.

The moral I am maintaining is that if type I associations did exhaust a terms
cognitive significance then there would have been little or no reason to pursue the project
of achieving a representational understanding of the mechanism responsible for affinity.
Nor should we expect to find scientists like Langley disturbed over the potential
metaphysical character of such terms. Both the pursuit and the disturbance suggest a
more sophisticated view of the significance of terms like ‘affinity’ as they occur in
scientific discourse. Scientist and ordinary people alike use terms to mention

\textsuperscript{218} Frankland, E. (1852). “On a New Series of Organic Bodies Containing Metals,”
\textsuperscript{219} Ibid, p. 441.
unobservable, hidden and even obscure or metaphysical qualities, entities and processes and this fact will inevitably bear on questions concerning the significance of these terms. A criterion of intelligibility needs to be able to distinguish terms which are significant from the point of view of their type II associations. MHC accomplishes this goal.

MHC entails that ‘affinity’ or any other term which is not susceptible of, or for which we do not possess, a model will be unintelligible. The fact that we find episodes in which the significance of terms was questioned along this dimension and corresponding to Buchdahl’s explicative dimension of theory assessment despite their being satisfactory along other dimensions supports the view that significance is not a simple function of inferential roles. This is not to give up positivism wholesale. MHC is situated in a broader framework of cognitive significance. Implicit in this framework is the idea intelligibility is not a necessary property of theories. Moreover, as I mentioned previously, there are clear parallels between (ii) and MHC and between (i) and (ii). Collectively and the broader framework of significance I have proposed. Positivists essentially conflate intelligibility and articulation and in so doing have too loose a notion of intelligibility.
Chapter VI

6.0.0. Introduction

A more recent criterion of intelligibility has been proposed by Henk De regt and further elaborated on in collaboration with Denis Dieks.\textsuperscript{220} According to De regt and Dieks whether or not a theory is intelligible is determined by an individual’s ability to grasp the theoretical consequences. They offer the following criterion as a “general Criterion for the Intelligibility of Theories”:

\textbf{CIT: A scientific theory }T\textbf{ is intelligible for scientist }S\textbf{ (in context }C\textbf{) iff }S\textbf{ can recognise qualitatively characteristic consequences of }T\textbf{ without performing exact calculations.}\textsuperscript{221}

CIT differs from MHC in several substantive respects. Unlike MHC, CIT is deliberately framed as a pragmatic criterion in the sense that theories that satisfy CIT always do so for particular individuals. As a consequence, the means by which }T\textbf{ becomes intelligible may differ significantly among individuals. Whatever aids a scientist or group of scientists in making predictions (without of course performing exact calculations) will serve as a means of rendering the theory intelligible }\textit{for that individual or group}. Whereas one individual may appeal to visualizations or geometric representations of a theory to grasp its consequences another scientist may rest their predictions on causal reasoning. Still other scientists may find the theory unintelligible altogether. They will, as a logical consequence of CIT, necessarily find a given theory unintelligible if they do not

\textsuperscript{220} De regt, Henk W., & Dieks, Denis (2005).
\textsuperscript{221} Ibid, p. 13.
appreciate its characteristic applications or are unable to assess its implications without making a concerted effort at carrying out the theoretical calculations.

The means by which a given individual may come to grasp a theory’s implications and hence find the theory “intelligible” are, for De regt and Dieks, indeed quite broad. Even a nondescript familiarity with the theory’s application may serve to satisfy CIT. As they say,

many theoretical physicists have developed a familiarity, and intuition for, the general behaviour of the solutions of the mathematical equations they use. This enables them to acquire a feeling for the qualitative behavior of the described systems without invoking picturable physical mechanisms. The claim that such physicists make, namely that they really understand the theory they are working with, is on our analysis perfectly legitimate.\(^{222}\)

MHC is on the other hand not so radically subjective. While the representational contents, which render theoretical terms intelligible, may vary from individual to individual, the constituents of any representation (i.e. representations of more basic experiential contents) must in principle be cognitively accessible to any (cognitively) normally functioning individual. Coming from the standpoint of MHC, I have a worry concerning the pragmatic character that the authors claim for CIT. One of the central criticisms of CIT that I will be offering later in the present chapter concerns what I perceive to be an overly subjective character in CIT.

Another characteristic of CIT, related to its subjective character and which equally distinguishes it from MHC, is its contextual nature. Not only is CIT subjective in

\(^{222}\) De regt, Henk W., & Dieks, Denis (2005). p. 16.
the above sense but it is also relativized to the conceptual schema available to the members of a linguistic community at a particular moment in time. Thus, as De regt and Dieks describe the situation, the problem that Leibniz, Huygens, and Newton himself, had with the intelligibility of Newton’s theory stemmed from their commitment to what was at the time a broadly accepted corpuscular-mechanistic view of nature. De regt and Dieks maintain,

The controversy about the intelligibility of Newton’s theory of gravity hinged on the acceptability of action-at-a-distance as a tool of understanding. Seventeenth-and-eighteenth century physicists who objected to Newton’s theory endorsed a corpuscularist ontology that implied contact action…This preferred metaphysics was a ‘canonization’ of the tools that previously had contributed to the achievement of scientific understanding. Thus Huygens’s rejected action-at-a-distance as unintelligible…he had learned to understand the natural world…by means of corpuscularist principles and models. The same applied to Newton himself.\textsuperscript{223}

The contextual nature of intelligibility helps to explain the subjectivity endorsed by De regt and Dieks. Whether or not an individual finds a theory intelligible depends on the conceptual resources available to the individual. Which conceptual resources are available will be influenced by factors such as the individual’s capacities and background

\textsuperscript{223} Ibid, p. 19. Of course the authors are wrong to imply as they do that Leibniz was a corpuscularist. Also, as I have already noted on a number of occasions, and argued in chapter 1, he did not find action-at-a-distance unintelligible. I would add that it is not clear that Newton found it unintelligible either. Newton’s own language concerning action-at-a-distance suggests that he understood it in a manner similar to that of Leibniz’s own view of the notion, namely, that it was an ‘absurdity’ and hence false.
beliefs, including any commitments they may have to metaphysical principles and scientific methodologies. Accordingly, anyone committed to a mechanistic ontology, will be unable to find Newton’s non-mechanistic account of gravity intelligible. This, in the view of the authors, was true of Newton. They contend that Newton did not understand his own theory: What applied to Huygens above “applied to Newton himself. The generally accepted tool of contact action canonized in corpuscularist metaphysics—did not enable him to understand his own theory.”

The sort of view of Newton’s theory taken by many of his contemporaries and perhaps even by Newton himself is not necessarily misguided in De regt’s and Dieks’ opinion. On their model, “rejecting a scientific theory as unintelligible for essentially metaphysical reasons – as Newton and Huygens were inclined to do – is not necessarily irrational: there may be strong heuristic reasons for attaching value to accepted metaphysics.” Metaphysical commitments, along with the capacities and additional background beliefs come to form what the authors label the ‘conceptual toolkit’ available to a scientist or community of scientist at a particular historical point. The tools present within the toolkit shift overtime and tools available to one community may cease to be available to subsequent community. Equally tools that were earlier unavailable may latter come in to the possession of future members of a scientific community. Specifically, resources like methodological commitments to visualizability, or causal-mechanical explanations may be given up during one historical period and reclaimed at another.

---

225 Ibid, p. 19. The authors go on to claim that “scientific understanding of it was still possible in a different, purely mathematical, way.” This, however, is inconsistent with their contention that intelligibility is both necessary and sufficient for achieving scientific understanding.
This picture of the historical significance of intelligibility leads the authors to accept what I will call the *Shifting Standards Thesis* (SST.) Essentially the idea behind SST is that which is described above: what accounts for a theories intelligibility, or lack thereof at one moment in time may not be what accounts for it at another moment in time. Moreover, a theory that is unintelligible may later come to be intelligible depending on shifts within the conceptual toolkit. Thus, during the eighteenth century Newton’s theory became intelligible as action-at-a-distance emerged as a conceptual tool. De regt and Dieks are explicit in their endorsement of SST. In fact they go so far as to maintain that, “one cannot deny that intelligibility standards actually vary,”\(^\text{227}\) and this is evinced by, “changes in the structure of scientific theories” which “make it difficult to maintain that the same intelligibility standards (if defined as particular properties theories must have) remain valid throughout the history of scientific thinking.”\(^\text{228}\)

Unlike the authors of CIT, I do not accept the shifting standards thesis. My primary reason for rejecting it is simply that it strikes me as inaccurate as a description of the actual historical worry concerning intelligibility. Moreover, it is difficult to see how anyone who accepts SST can provide a substantive means of distinguishing theories that are unintelligible merely at given point in history and those which are unintelligible without qualification. Perhaps the latter is simply an empty class. For the most we can ever say is that a given theory fails to satisfy CIT *at the present moment*. But, for all we know, it may latter come to be intelligible as new tools become available. This view seems to pull some of the teeth from the charge that a theory is unintelligible. It doesn’t give us reason to reject the theory, as it seemingly should given De regt’s and Dieks’

\(^{227}\) Ibid, p. 6.
\(^{228}\) Ibid, p. 9.
claim that intelligibility is necessary for scientific understanding.\textsuperscript{229} On my view intelligibility is a function of our, largely fixed, biological and cognitive makeup. This issue concerning SST will be picked up later in the essay along with that of CIT’s radical subjectivism.

A third feature of CIT distinguishing it from MHC is its holistic character. CIT does not enable us to evaluate terms in isolation. It is whole theories that get evaluated as either meeting or not meeting the condition specified in CIT. On the other hand, according to the account I have offered a theory is judged unintelligible only derivatively as a product of its incorporation of unintelligible terms. Historically, when theories were criticized on the grounds that they were unintelligible it was because of their incorporation of troublesome notions putatively designated by terms like ‘attraction’, ‘affinity’, ‘sulfuric quality’ and the like. It is difficult to make sense of these criticisms applying CIT.

A final distinguishing feature concerns the sort of criterion that CIT is. CIT does not specify what intelligibility consists in. Rather it gives us an extrinsic test for determining whether a given scientist does indeed find a theory intelligible. It does not tell us what it is internal to a scientist and/or the theory that makes the theory intelligible. Indeed it cannot since the mechanisms, which provide for the satisfaction of CIT, will vary among individuals. By contrast according to MHC representations constitute the mechanisms by which terms are rendered intelligible. MHC thus provides a substantive

\textsuperscript{229} Note that I am not claiming that my account doesn’t pull some of the teeth from the charge. It is rather that I am pulling different teeth. On my model a term that is unintelligible is unintelligible with without qualification. That is, if t is genuinely unintelligible, as opposed to being merely confused or misunderstood, then it can’t at any point become intelligible while retaining its original meaning. Also I want to be clear that I am not maintaining that standards of explanation do not shift over time.
hypothesis in response to the question concerning what intrinsic properties of individuals are responsible for a term’s intelligibility. As we shall see below, the only answer available to De Regt and Dieks is often no more than an un-specifiable familiarity with, or intuition for, the theories consequences.

6.1.0. Elaborating De regt’s and Dieks’ Model (DND)

De regt’s and Dieks’ stated goal is to provide an account of what makes various kinds of explanation genuinely explanatory. They note that despite extensive debate over the nature of explanation no single model yet provides us with an account of what makes different explanatory schema all explanatory. That different forms of explanations have been employed historically is not a contentious claim. As we saw previously, Newton’s explanation of gravitational phenomena was viewed as a departure from the causal-mechanical structure of explanation employed by most of his contemporaries, and in the nineteenth century Brodie’s chemical calculus was the antithesis of causal explanations.

A number of philosophical accounts of explanation have been offered which construe it variously in terms of notions such as unification, causation, visualization, familiarity, etc. W. H. Newton-Smith characterizes the current state of our understanding of the nature of explanation as one in which,

We have an embarrassment of riches. We have explanations by reference to causation, to identities, to analogies, to unification, and possibly to other factors. Philosophically we would like to find some deeper theory that explained what it was about each of these apparently diverse forms of
explanation that makes them explanatory. This we lack at the moment (130-31).^{230}

De regt and Dieks aim to fill the lacuna Newton-Smith highlights by providing us with a model of scientific understanding. As they see it what unifies different forms of explanation is that they all generate understanding. CIT then comes as part of a broader theory of scientific understanding according to which, “A phenomenon \( P \) is understood iff a theory \( T \) of \( P \) exists that is intelligible (and meets the usual logical and empirical requirements).”^{231} This Criterion, which the authors label CUP (Criterion for Understanding Phenomenon), is meant to capture a general aim of science. All else being equal, a phenomenon is understood when we have an intelligible theory of it. While De regt and Dieks see CUP as a universal aim of science it incorporates a pluralistic view of explanation. Different types of explanation will be explanatory just in case they lead to understanding. They articulate this picture by distinguishing between three levels at which scientific activities may be analysed: “the macro-level of science as a whole; the meso-level of scientific communities; and the micro-level of individual scientists.”^{232}

According to the authors,

the three-level distinction can reconcile the existence of universal aims of science with the existence of variation in the precise specification and/or application of these general aims. The macro-level characterisation of universal aims must necessarily be rather general in order to accommodate micro-level differences, but it may still provide us with information about

\[^{230}\text{Newton-Smith, W.H. (2000).}\]
\[^{231}\text{De regt and Dieks (2005). p. 12.}\]
\[^{232}\text{Ibid, p. 3.}\]
scientific practice . . . This leaves open the possibility that scientists in different historical periods or in different communities have quite different specific views about precisely how scientific understanding is to be achieved.233

Whereas understanding is a general aim of science as a whole, different scientific communities at different times may employ different explanatory strategies and conceptual tools to achieve it. At the level of the individual scientist achieving understanding will be influenced by their background beliefs and abilities.

The principle use of any explanatory strategy is however not directly to yield understanding. Rather, the primary goal is to generate an appreciation of a theory’s qualitative consequences. It is the recognition of the theoretical consequences that, in turn, enables a scientist to render a theory intelligible and as a consequence understand the phenomena under investigation. The central idea behind CIT is the notion that what different explanatory strategies, whether it be causal-mechanical reasoning, visualization, analogical thinking or another strategy altogether, provide the scientist with is a means of becoming familiar with, or of developing an intuition for, the theoretical consequences. This is a theme that is perhaps brought out most clearly in the quote concerning theoretical physicists provided earlier. But it is a theme that is continually rehearsed by the authors.

Early on in their essay De regt and Dieks contrast their view with that of more positivistically minded philosophers who, in their view, see the aim of science as the production of factual knowledge and true predictions. They maintain,

233 Ibid, p. 4.
It is beyond dispute that a mere list of true descriptions or predictions is not regarded as true science. In science we want more than just factual knowledge. Think of a hypothetical oracle whose pronouncements always prove true. Although empirical adequacy would be ensured in such a situation, we certainly would not speak of a great scientific success and not even of science tout court, because there is no insight in how these perfect predictions were brought about.234

As this quote makes clear predictiveness in and of itself is not an intelligibilizing feature of a theory. Rather, “In contrast to an oracle, a scientific theory should be intelligible: we want to be able to grasp how the predictions come about, and to develop a feeling for the consequences the theory has in a particular situation.”235

Of course, one might gain a feeling for a theories consequences via different means. Further on they describe this as an “intuitive feeling for how the theory works,”236 and maintain it as an instance of deploying an explanatory strategy to achieve understanding: “Causal reasoning is used to achieve the goal of qualitative prediction; it is an instrument that helps us to get a feeling for the situation.”237 While causal reasoning might constitute the preferred strategy in one scientific domain, visualization may be the preferred strategy in another. De regt and Dieks note that the latter strategy has played an important role in some domains: “As a simple illustration, consider the idea of field lines

234 Ibid, p. 5.
235 Ibid.
236 Ibid, p. 12.
in electrostatics... Although intuitive application of this idea is possible only in simple situations, it is quite useful to get a feeling of how electrostatic systems behave.\(^{238}\)

On the whole these remarks illustrate the general character of intelligibility implicit in CIT. Implicitly, what makes a theory intelligible on De regt’s and Dieks’ model is that we are able to develop an intuition or feeling for its application by whatever means necessary. As they say, “A central role in this criterion is played by deductions (though not of the formal kind) from a theory; namely the intuitive apprehension of characteristic consequences.”\(^{239}\)

The characterization of intelligibility in terms of ‘feelings’ and ‘intuitions’ is problematic. It appears clear enough, given the pluralism they adopt with respect to explanatory strategies, De regt and Dieks have little choice but to articulate the criterion with the aid such language. Given their view that intelligibility does not consist in a particular property of either theories or individuals, they have little recourse to any criterion other than one that merely “describes a sufficient (possibly also necessary) condition for the presence of intelligibility.”\(^{240}\) They are unable to be more specific concerning what these diverse explanatory strategies provide the scientist with. And that, as I shall argue below, is precisely what an account of intelligibility needs to provide. Before turning my attention to the criticisms that I have alluded to, in the next section I

\(^{238}\) Ibid, p. 15.
\(^{239}\) Ibid, p. 20. It is not clear to me in what sense precisely these are deductions “from the theory.” Maxwell’s geometric models, for instance, were not part of the theory at all but rather simplified interpretations or structures to which the theory might be applied. MHC appreciates this relationship between theory and model by holding that at the level of theories the models are constrained by the theory as being a conceptual instance of it (see chapter 4).
\(^{240}\) Ibid, p. 12.
examine the broader notion of understanding that De regt and Dieks appeal to. Not surprisingly this notion is problematic in its own right.

6.2.0. Understanding: Concerns and Confusions

While De regt’s and Dieks’ account of understanding isn’t my chief concern, the manner in which they are inclined to think about understanding influences the account of intelligibility that they offer. It is, therefore, worth considering whether and to what extent the notion of understanding they employ captures interesting features of scientific practice. My overarching worry here is that have conflated different senses of intelligibility and understanding, and the degree to which this is problematic for the account that they offer. At the outset, I want to stress that I am not offering this as a decisive argument against their view. Rather I believe there is reason to worry about the use of these terms in the development of their view and appreciating this may help the reader to better understand some of the criticisms that I will be elaborating in the later sections.

The claim on which De regt’s and Dieks’ analysis of scientific understanding is founded is “that one has scientifically understood a phenomena when one has an intelligible scientific theory of it.” After making this claim De Regt and Dieks provide the following motivation for this claim:

Realists typically claim that scientific theories provide understanding of the observable phenomena by telling a true story about the underlying processes. But this makes sense only if these underlying processes are intelligible; or else one has merely replaced one unintelligible phenomenon by another. This implies that the theories themselves should be intelligible…
for how could a process be intelligible when the theory that gives the true
description of it is completely ununderstandable.241

As this passage indicates De regt and Dieks are inclined to talk equally of the
intelligibility of phenomena and theories, and that they are comfortable doing so even in
the same breath. Moreover, it suggests that they view ‘understandable’ and ‘intelligible’
and cognate expressions to be largely synonymous.

Unlike De regt and Dieks I’d prefer not to use ‘intelligible’ and its cognates with
respect to phenomena. However, the tendency to do so is not unique to them. It is indeed
part of our ordinary practice. We might, for instance, hear someone remark that they
“don’t understand how electricity works”, or alternatively, that they find the “manner in
which electricity works to be unintelligible.” Nicholas Rescher spends a chapter of
Nature and Understanding in an effort to account for why, “the majestic lawful order of
nature is intelligible to us.”242,243 I have no particular objection to the practice of using
these expressions synonymously in application to phenomena. The worry arises once we
realize that they are also used synonymously in a different sense.

In this second sense we talk about whether purported concepts, expressions, terms,
or statements are intelligible or even, as with CUP, the intelligibility of theories (as
opposed to the phenomena they are about or the “mysterious order” they are allegedly
aiming to capture.) At the same time we come across claims concerning what it is to

241 De regt and Dieks (2005), pp. 6-7.
243 Others engaged in the discussion of “scientific understanding” use ‘Understanding’
and ‘intelligibility’ in a similar manner. See for instance; Cushing, James T. (1991)
“Quantum Theory and Explanatory Discourse: Endgame for Understanding,” Philosophy
understand a theory. My worry arises when the synonymy is carried over across these two domains; the domain of language versus the domain of phenomena. This is true for either term generally.

We ought not equate understanding a theory with understanding the phenomena that it purports to explain for the simple reason that we can understand theories that are false (or otherwise deficient). Whatever understanding a phenomena might consist of, it is not at all clear to me that it involves merely having (an empirically adequate) understandable theory about it. Were this the case then every “how possibly” account of a phenomenon that is understandable would yield understanding of its target. Contrary to this view of understanding phenomena, I incline toward the idea that ‘understanding’ at least in this domain is a success term. That is, is one doesn’t understand the motion of an automobile if they think that the energy producing it is derived from combustion when it is in fact produced by an electrical motor any more than they would if they believed it to be produced by monkeys under the hood chasing bananas on a treadmill. Note that each of these accounts, of how the automobile’s motion is produced, however unlikely, is arguably intelligible.

De regt and Dieks do not endorse this view. That they see understandable theories as sufficient (together with their being empirically adequate) to generate understanding of their target phenomena follows directly from CUP. It is, perhaps, possible that my worry can be accommodated by saying more concerning what they have packed into empirical adequacy; but they have not done this and I can’t see any principled way to do this simply on grounds of empirical adequacy—unless “empirical adequacy” has built into it the epistemologically problematic notion that theories are only empirically adequate
provided the story they tell concerning the underlying processes is true.\textsuperscript{244} In any case, I think there is an even stronger claim to be made here.

It is not at all clear that ‘intelligibility’ or ‘understanding’ in the sense in which they are applied to theories or language more generally make any sense when applied to phenomena. When we raise questions about the intelligibility of theories we are inquiring into their content and the degree to which this content is cognitively accessible. When we ask, for instance, about the significance of Dalton’s atomic theory and in particular about the character Dalton’s atoms we are asking what Dalton meant by ‘atom’. We are asking how they are to be modeled or thought of? And if it should turn out that they can’t be thought of or cognized, (i.e. if they are like Lockean “substance”—a term by which we “signify nothing”), then we judge them to be unintelligible.\textsuperscript{245} On the other hand, “a phenomenon that can’t be thought of or cognized” seems to be simply a contradiction in terms. Dalton’s atomic hypothesis was put forward to explain combining ratios, the phenomenon or fact that elements combine in definite proportions. That elements combine in definite proportions is a fact we record or take note of and subsequently seek to explain. I can’t see how it makes any sense to ask whether the brute fact is intelligible in this sense once we have observed it.

On the other hand, when we seek to understand a phenomenon we are asking perhaps about how it is related to other events or what possibly its significance might be

\textsuperscript{244} It is epistemologically problematic because we will, in many cases, inevitably be unable to discern the truth of the story being told. Thus, for instance, when Boyle proposed his corpuscular theory of matter there was no way to test its truth. Nonetheless, Boyle’s view was offered precisely on the grounds that it provided for a more intelligible explanation of chemical phenomena. This case is a dramatic illustration of what strikes me as an otherwise obvious principle: What we can understand (in the context of language) ought to be independent of epistemic questions concerning its truth.

\textsuperscript{245} Essay, p. 95.
for future events, or what causes its occurrence, or even how we might intervene in events of its kind. In short, it appears that what we are seeking in these instances is to develop a greater facility with the phenomena. We might even say, along with De regt and Dieks, that we aim to become more familiar with the phenomena. It may even be true that different explanatory strategies can help us to achieve this goal. But they do not do so by rendering theories intelligible. Nor, in my view, is a theory’s intelligibility sufficient to produce understanding of this sort—monkeys chasing bananas on a treadmill is an intelligible yet insufficient explanation (though perhaps not one that provides a physically plausible or even possible mechanism) if understanding the car’s motion is a theoretical goal.

The conflation of these two senses of intelligibility is, I suspect, what leads De regt and Dieks to adopt the notion of an “intuitive feeling” as their core idea. Essentially, they maintain that a phenomenon is understood when we are familiar with it. This occurs when we are able to tell some story concerning the processes underlying it. But, as they say, this makes sense only if this story is itself intelligible; a notion, which, in turn, they unpack in terms of “familiarity,” “feelings for,” and “intuitions” concerning the theories application. Again, I do not take this to decide the matter against their view but it may help clarify the sense of ‘understanding’ and ‘intelligibility’ that De regt and Dieks are operating under.

To be fair, they don’t do so on De regt’s and Dieks’ account either. Rather they serve to permit the scientist to make predictions and it is this later ability that “indicates” the theory is intelligible. The sense of indicating here is analogous to the manner in which the clicks of a Geiger counter indicate the presence of radioactive substances. It is in this sense that the account can be thought of as providing for an extrinsic mark of intelligibility.
The idea that ‘understanding’ is a general aim of science has not been well received by some. Individuals like Hempel and van Fraassen have considered ‘understanding’ to be too pragmatic to be of much use in distinguishing explanations as such. Hempel’s remarks in this direction are especially interesting as they not only clarify the objection but also help to illuminate the specific sense of “understanding” that is at issue.

He is not unreflectively critical of the idea that understanding is an aim of explanation or science in general. Rather, he accepts that understanding, construed in the proper light, is a goal shared by scientists interested in providing explanations. “But,” he maintains,

it is important to distinguish here understanding in the psychological sense of a feeling of emphatic familiarity from understanding in the theoretical, or cognitive, sense of exhibiting the phenomenon to be explained as a special case of some general regularity. The frequent insistence that explanation means the reduction of something unfamiliar to ideas or experiences already familiar to us is indeed misleading.247

It is misleading, Hempel believes, because scientists frequently seek to explain or understand familiar phenomena through appeals to more abstract and less familiar principles. Hempel points to Newton’s explanation of the free fall of bodies as a case in point. Accordingly, Hempel holds that familiarity is not a necessary condition on explanations.

---

De regt and Dieks are not without a response here. One virtue of their view is that it does afford us an objective means for determining whether a scientist is familiar with a theory. We need only observe the degree of success with which the scientist makes predictions. There is no issue of theories being more or less familiar than the phenomena they are intended to explain. Whatever facilitates the determination of qualitatively characteristic consequences ipso facto enables the scientist to better understand the theory, and consequently the phenomena. For De regt and Dieks scientific understanding always involves telling a story about the underlying processes.

It is important to appreciate that in the first instance the sort of theories that they have in mind are mathematical in nature (though they mean for their account to cover scientific understanding generally.) At this first order of abstraction the theory may indeed be unintelligible or unfamiliar and consequently unable to provide us with any genuine understanding of the phenomenon. I take it that this is the moral of their contention that Newton and his contemporaries did not understand his theory or the gravitational phenomena he sought to explain in virtue of it. Hempel, and I am inclined to agree with him, has a different assessment of the situation.

As Hempel sees it Newton did understand his own theory despite the fact that the notion of an attractive force remains as an unexplicated element. This is supported by the fact that Newton and others, like Leibniz, were able to apply the first order mathematical theory to gravitational phenomena generally. There doesn’t appear to be any principled

248 What I have in mind here is simply that a first order of abstraction consists, for De regt and Dieks, in the raw uninterpreted mathematical formula. Second order abstractions involve interpretations of first order quantities like “attractive force” and the like. This may involve visualizations of their referents or analogical or geometrical interpretations and so on. As near as I can tell the second order abstractions are necessary in order to make qualitative judgments concerning the theoretical consequences.
reason to contend that scientific understanding is not achieved here. Furthermore, De regt and Dieks concede as much when they grant that, “scientific understanding of [Newton’s theory] was still possible in a different, purely mathematical, way.”\textsuperscript{249} If this is right then Hempel’s contention that intelligibility (in De regt’s and Diek’s sense) is not a necessary property of theories is correct.

Hempel, moreover, does not view the generation of “emphatic understanding” or “familiarity” to be a sufficient condition for the adequacy of explanations either. As he sees it the problem is that,

Much of the appeal of the “method of understanding” seems to be due to the fact that it tends to present the phenomena in question as somehow “plausible” or “natural” to us; this is often done by means of persuasive metaphors.\textsuperscript{250}

Yet, historical episodes wherein the use of suggestive metaphors led to a false sense of understanding are numerous. Appeals vital spirits, hypostatic principles, and divine purposiveness constitute just a few. Newton-Smith’s particularly striking case, to which I appealed in chapter 4 is worth restating here:

While certainly an explanation must be capable of giving an appropriately tutored person a psychological sense of understanding, this is not likely to be a fruitful way forward. For there is virtually no limit to what has been taken to give understanding. Once upon a time, many thought that the fact that there were seven virtues and seven orifices of the human head gave

\textsuperscript{249} De regt and Dieks (2005), p. 19.
them an understanding of why there were (allegedly) only seven planets.

We need to distinguish between real and spurious understanding.251

Hempel doesn’t deny that we can sensibly talk about understanding as a universal aim of science. Instead, he distinguishes scientific understanding from the sort of emphatic understanding described above; the sort of understanding appealed to by adherents of what he labeled the “method of understanding.” He maintains that, “the kind of understanding thus conveyed must clearly be separated from scientific understanding.”252 Unlike emphatic understanding, scientific understanding comes only after we have demonstrated that the phenomenon follows from a general and empirically acceptable law together with empirically adequate statements describing the conditions under which the phenomenon was observed. Consequently, “A class of phenomena has been scientifically understood to the extent that they can be fitted into a testable, and adequately confirmed, theory or system of laws.”253 Or what is the same, “the explanation of a phenomenon consists in subsuming it under general laws.”254 The moral, according to Hempel is that, “the criterion of [an explanation’s] soundness is not whether it appeals to our imagination, whether it is presented in terms of suggestive analogies or is otherwise made to appear plausible.”255

If Hempel and others, like Newton-smith, are right then understanding, or intelligibility, in De regt’s and Dieks’ sense is neither a necessary nor a sufficient

---

255 Ibid.
condition on a theory's ability to produce scientific understanding. This does not, of course, entail that CIT is false. It may be the case that while CUP is false, that CIT nevertheless provides an accurate picture of a pragmatic feature of theories in the sense that theories that satisfy CIT are generally though not always preferable to those that don’t. Presently, then, I will turn to considering the adequacy of CIT itself.

6.3.0. Criticisms of CIT

My concerns with CIT correspond to the four “characteristic features” of CIT articulated above (section 6.0). For convenience these characteristics are:

i. CIT’s pragmatism; or, to adopt an admittedly more biased descriptive vocabulary, what I shall term its radical subjectivism: intelligibilizing mechanisms vary from individual to individual as their capacities and background beliefs vary.

ii. The commitment to SST: CIT is deliberately committed to the view that standards of intelligibility shift over time as new conceptual tools make their way in and out of the “conceptual toolkit”.

iii. CIT’s holistic character: It is theories, as opposed to terms or expressions which either possess or fail to possess the property of intelligibility.

iv. The extrinsic character of CIT: CIT provides an external test for the presence of intelligibility within a given context.

6.3.1. CIT’s Radical Subjectivism

What makes CIT subjective is that it is relativized to individual scientists. That is, a theory is never intelligible in and of itself but always with reference to someone who understands it. What makes CIT radically subjective is that finding a theory intelligible
consists for the individual in having an “intuitive feeling for” or “familiarity with” its application. Subjectivity in either sense has deleterious consequences.

CIT’s subjectivity entails that theories cannot be objectively intelligible. For a given theory, if no one understands it (i.e. no one is able to qualitatively—without carrying out the mathematical calculations—determine its consequences) then we are forced to say that the theory is unintelligible. This, of course, is only a relative fact about the theory. Someone S may come along who is able to qualitatively determine the characteristic consequences of the theory at which point we should say that theory is intelligible for S, though not for us.

The obvious point to make here is that this picture simply doesn’t capture what was at the heart of the historical criticism. Individuals like Boyle and Leibniz saw intelligibility as an objective property of theories. While it is true that it was viewed to be a property that theories possessed in relation to persons, it is persons or humanity as a biological or psychological kind. Moreover, as has been noted at various points throughout this dissertation, the issue did not concern the possibility of grasping the theories consequences. If that were the issue, arguably, we should be surprised to find Boyle arguing that the spagyrical chemists were speaking unintelligibly. For what could be clearer than the idea that adding more of the sulfuric principle to a substance would, among other things, make it more combustible? The problem wasn’t that. The problem was, and is, that we have no idea what it is that we would be adding. Similarly, Leibniz didn’t claim that he couldn’t understand the consequences of Newton’s theory. He understood them all to well.
There are a few possible responses available to De regt and Dieks. Perhaps the most obvious response available to them is the “so-much-the-worse-for-them” response. According to this response Leibniz and Boyle, didn’t really understand their own criticisms. Or at the very least, and perhaps more charitably, they didn’t really appreciate what it is about our psychological makeup that is responsible for our finding theories intelligible. What kept them from understanding these theories, De regt and Dieks maintain, was their commitment to the mechanical philosophy.

I don’t deny that views were criticized for their failure to conform to the mechanical philosophy but these criticisms are better situated within the regulative component of Buchdahl’s model of theory assessment. The demand for intelligibility was a demand for an explication of the questionable terms. It is notable that mechanistic commitments, while widespread, did not affect everyone equally. Clarke notes in his fourth reply,

That one body should attract another without any intermediate means, is indeed not a miracle, but a contradiction; for tis supposing something to act where it is not. But the means by which two bodies attract each other, mat be invisible and intangible, and of a different nature from mechanism; …if [natural forces] does not mean mechanical forces; then gravitation may be effected by regular and natural powers, though they not be mechanical.256

What are we to say of this passage from Clarke? Since he did not suffer from the mechanistic myopia that ailed most of his contemporaries including Newton, are we to say that Clarke better understood Newton’s theory than the author himself? Perhaps. But

---

256 Alexander, H. G. The Leibniz-Clarke Correspondence. (New York: Manchester University Press), 1956; p. 53.
the problem is more severe. We would like to know what it is that Clarke possessed that his afflicted colleagues lacked. CIT does not provide a plausible response to this query. If CIT is right and, Clarke did indeed find Newton’s gravitational theory intelligible in ways that Newton and others did not, then Clarke was familiar with the consequences of the theory in manner that others were not. The problem is that there is no evidence that Clarke had a better appreciation of the “qualitatively characteristic consequences” of Newton’s theory than either Newton or Leibniz. And if he did not then there is no reason to believe that the commitment to the mechanistic philosophy was what prevented Leibniz and Newton from appreciating those consequences. Consequently, De regt and Dieks lack a plausible explanation of why they found the theory unintelligible given that on their account lifting of the mechanistic constraints on theories should have alleviated this problem, as it eventually did.

Moreover the “so-much-the-worse-for-them” response doesn’t address the second worry. It simply isn’t true that Newton and Leibniz did not appreciate the consequences of Newton’s theory. Nowhere is Leibniz’s charge that ‘attraction’ is unintelligible tied to possibility of determining its characteristic consequences. More importantly, Leibniz concedes that he can appreciate the consequences of Newton’s theory and even that it “explains” certain celestial phenomena maintaining that, “although Newton is satisfactory when one considers only a single planet or satellite, nevertheless, he cannot account for why all of the planets of the same system move in the same direction, using only impetuosity together with gravity.”257 Leibniz here suggests not only that he understands

---

the positive consequences of Newton’s theory but that he equally understands what is left unexplained by the theory and you don’t get the latter by carrying out precise calculations. If CIT is correct and Leibniz really did appreciate the qualitatively characteristic consequences of Newton’s gravitational theory then, despite his claims to the contrary, Leibniz really did find the theory intelligible. In sum the “so-much-the-worse-for-them” response doesn’t account for the historical facts.

As an alternative we might modify CIT’s modality so that it reads, “A scientific theory T is intelligible for it is possible for some scientist S (in context C) iff S can recognise qualitatively characteristic consequences of T without performing exact calculations.” But Leibniz’s familiarity with the consequences of Newton’s theory suggests that the problem of intelligibility should not have arisen. Leibniz’s familiarity is a realization of this possibility and despite this, the problem of intelligibility persisted for him. Of course one might rejoin that Leibniz simply failed to appreciate this but this is merely to return to the failed “so-much-the-worse-for-them” response. Neither response seems capable of explaining the historical reality that De regt and Dieks are committed to explaining. Of course, this doesn’t prove that intelligibility is an objective property of theories but it does show that the sort of subjective familiarity with a theory’s consequences, implied by CIT, does not capture the historical worry about intelligibility during this period. Moreover, the subjectivity thesis (if I may call it that) is inconsistent with the core conception that I argued for in chapter 2. I will pick up on this theme in the next section.

258 May he now rest in peace!
Turning then to the idea that intelligibility is radically subjective we are presented with what strikes me as the most troublesome aspect of CIT. We might think: “well, if it’s not plausible that intelligibility is subjective in De regt’s and Diek’s sense, then why even worry about the thesis of radical subjectivity.” And the reason is that the two are not mere matters of degree. It may well be that it isn’t subjective in the manner suggested above (i.e. that it is not a function of an individual’s abilities and background beliefs), but that it is nonetheless a matter of possessing an “intuitive feeling” that we are all, as a psychological kind, capable of developing. The problem here is that the notion of an intuitive feeling is overly vague.

In section 6.1.0 I provided a number of citations illustrating their commitment to the idea that intelligibility consists in having an intuitive “feeling for” or “familiarity with” a theory’s (qualitative) consequences. Given this it strikes me that a reasonable rephrasing of De regt’s and Diek’s requirement might run as follows:

\[
\text{CIT}^R: \text{A scientific theory } T \text{ is intelligible for scientist } S \text{ (in context } C) \text{ iff } S \\
\text{has an intuitive feeling for the qualitatively characteristic consequences of } T \text{ without performing exact calculations.}
\]

Admittedly, the rephrasing in CIT\(^R\) is rhetorically motivated. I don’t believe that anyone would find CIT\(^R\) compelling. Simply put, it is just not very informative. We all have intuitive feelings about a host of things: how to swing a golf club, why American beer is better than French wine, what would happen if the monkeys fell off the tread mill, and so on. The idea that such intuitions generate understanding is on the one hand unsatisfactory because it doesn’t tell us anything about what this understanding consists in. This may be forgivable given the authors’ pluralistic commitments. But the criterion is also false for,
among others, the reasons cited by Hempel and Newton-Smith above. Having an intuitive feeling for how to swing a golf club does not entail that one understands any theory concerning the mechanics involved. On the other hand someone might have a good understanding of the mechanical theory involved and yet lack the intuitive feeling for how to swing a golf club that a seasoned professional such as myself possesses. “Intuitive feelings” appear to be neither necessary nor sufficient for understanding. Perhaps this consequence can be avoided by providing a more precise explanation of what these intuitive feelings are.

6.3.2. The Commitment to SST

The primary motivation for De regt’s and Dieks’ adoption of a pluralistic view of intelligibility, and consequently what leads them to unpack intelligibility in terms of familiarity, is their commitment to the shifting standards thesis (SST). As they see it there is a dilemma involved in filling the lacuna noted by Newton-Smith. We are faced with potentially mutually exclusive goals:

The dilemma in answering the question of what scientific understanding is, appears to be: Should we take the views of practising scientists seriously, with the danger of being led into a relativistic view on which understanding is merely a matter of arbitrary fashion? Or should we look for a philosophical, generally valid conception of scientific understanding?259

When we look at the practices of scientists throughout history we observe preferences for different explanatory strategies at different periods and even within the same period.

---

259 De regt and Dieks (2005), p. 2.
among different groups of scientists. If we take this variation within actual scientific practice seriously we are likely to be led to the conclusion that it is not feasible to give a unified account of explanation. But that is, as Newton-Smith suggested, precisely what we are after.

De regt and Dieks believe that their view, “bypasses this dilemma by proposing an approach that offers a general characterisation of scientific understanding that can encompass the historical variation of specific intelligibility standards employed in scientific practice.”260 The macro-meso-micro-level structure discussed earlier (captured in figure 6.1) is the means by which they hope to achieve this. Briefly, the differences we observe in the practice of science are construed as differences in preferences for how best to achieve understanding, and are accounted for at the micro and meso levels. At the macro level, understanding remains as a universal aim of science. The differences at the former levels are essentially, though not exclusively, differences concerning which explanatory strategies constitute the preferred means for achieving the broader goal of understanding. Thus, it is their commitment to a pluralistic account of explanation together with the idea that explanatory strategies are the means by which we render theories intelligible, that accounts for their adopting a criterion of intelligibility that is at its core pluralistic.

Thinking through this picture raises a rather dizzying number of concerns, some of which are more naturally pertinent to the issue of intelligibility than others. I will mention the less pertinent ones first. I discuss these concerns first in the hope that doing so may help to clear out some of the conceptual fog that has settled here.

260 Ibid.
While my principle concern isn’t with the nature of explanation, it is unclear what the authors’ view explanations to be. On the one hand they appear to have a rather broad notion of explanation encompassing visualizations, causal mechanical reasoning, D-N accounts, analogical and geometric models, and so on. It is not clear to me that visualizations, for instance, are explanations. Nor do the authors give us any reason to why we should consider them to be explanations other than to claim that they aid in our understanding of theories. But this surely isn’t sufficient since a wide range of things may aid us in understanding a theory that we would nonetheless be disinclined to label “explanations.” The periodic table of elements is not an explanation or a theory, though it is certainly a tool or record that organizes information in a systematic way that may help to facilitate our understanding like a tax-to-income table or Geoffroy’s tables of affinity. The later were not theories or explanations. The theory was that substances have definite affinities for one another. A table was a record of the varying affinities among substances (see figure 6.3). The explanation, when it is forthcoming, concerns what is responsible for the observed affinities—e.g. the electro-chemical properties of the respective substances. Again, I don’t have any stake in whether these sorts of things (tables and other representational devices) are considered to be explanations. My point is simply that whether or not they count as explanations can’t be a merely matter of whether they aid in our understanding a theory. They surely do that, but so does a thermometer.

To carry this worry a bit further, we can connect it with a further worry concerning what exactly the relation between theory and explanation is for De regt and Dieks. They view the central idea behind CIT to be that intelligibility is achieved by means of, “deductions (though not of the formal kind) from a theory; namely the intuitive
Yet, as I note previously, it is unclear in what sense these “informal deductions” are “deductions from the theory”. Consider geometrical representations. A geometrical representation of a limit or sum is not part of the theory of Analysis (i.e. the mathematical theory which explains the theorems of calculus). Maxwell’s geometric models were not part of the thermodynamic theory he was proposing. Nonetheless, for De regt and Dieks it is through the use of such devices that we draw out theoretical implications. It is evident that they are operating under some implicit distinction between theories and the explanatory strategies that we employ in order to understand them. But it is not clear what that implicit distinction is. That is, what is it that we’re supposed to recognize the “qualitatively characteristic consequences” of? The model? The theory? Dalton’s atomic model is indeed a theory about the microstructure of matter, which would, if true, at least partly explain combining ratios. But what theory are we trying to find intelligible here? Dalton’s theory or the theory of definite proportions? And what’s doing the intelligibilizing work here? Is it Dalton’s atomic model, or is it the fact that Dalton’s model was visualizable? I don’t know the answers to these questions and I am not convinced they can be given.

One problem is that the use of devices like visualization, analogical reasoning, geometrical representation, and the like, while they may aid us in understanding theories, often carry a risk. Often such devices incorporate, to borrow a phrase from Hesse, negative components. This, as you will recall from chapter 4, is precisely why I maintained that as intelligibilizing mechanisms models are non-semantic devices—they don’t determine the truth of statements involving the terms or expressions with which

\[261\] Ibid, p. 20.
they are associated. The occurrence of such negative components explains, in part, the worry raised by both Hempel and Newton-Smith. While such devices may serve to generate understanding of a sort, they may also mislead us into a false sense of understanding.

At this point I want to mention one final worry before proceeding to my main criticisms of the authors commitment to SST. Given their pluralistic view of intelligibility and the radical subjectivity that it involves, I can see no way that this model can give any substance to the historical fact that scientists themselves disagreed over which sorts of views were genuinely explanatory. During the nineteenth century, positivistically minded thinkers like Duhem and Brodie held that Dalton’s atomic model was not explanatory. The only principled response that we can find forthcoming from the point of view of CIT is that they were simply wrong: insofar as Dalton’s atomic model had recognizable, or facilitated recognition of, qualitatively characteristic consequences of some theory then it was explanatory. This sort of response harkens back to the so-much-the-worse-for-them response discussed above.

Insofar as we have a philosophical interest in the history of any intellectual worry, we want in part to be able to provide a substantive explanation of the character of these historical concerns. CIT does not achieve this. Any theory that meets the criteria specified in CIT is intelligible. Any historical figure that disagrees with this assessment is *ipso facto* wrong. To be sure their mistake is explicable. It stems from their micro and meso level commitments to disparate explanatory strategies. The criticism here isn’t that there is no explanation, but that the explanation trivializes the intellectual worries involved. It doesn’t achieve the professed goal of providing an account of the historical
concern with intelligibility. Rather it assumes that all theories on all sides can be made intelligible despite historical claims to the contrary. There is, in my view, nothing historical in the development of CIT.

I don’t want to rest much on the foregoing criticisms. I suspect that clarification of their model might help to relieve some of the worries mentioned above. I leave that project to them. I do however have two central problems with their commitment to SST. The first problem is that historically the criticism that certain views were unintelligible was generally a homogeneous one—call this the homogeneity criticism. That is, a misguided understanding of the history of the worry motivates the commitment to SST. The second worry that I have is what we might term a structural worry. The structural worry concerns what exactly is going on when shifts in the conceptual toolkit come about, and equally questions about how the newly incorporated concepts avoid concerns about intelligibility in their own right?

The homogeneity criticism stems from considerations taken up in chapter II. Specifically, I argued there for the widespread endorsement of the what I labelled the core conception (CC) of intelligibility. According to CC, intelligibility is a property that terms possess in virtue of our ability to associate them with definite cognitive constructs. For seventeenth and eighteenth century philosophers the constructs in question were comprised of ideas. What CC explains is the fact that criticisms to the effect that various terms or expressions, or even theories, were unintelligible were readily enough understood. That is, when we look at the historical literature we don’t come across disagreements of the following sort:
A: As I see it all bodies are endowed with a quality of affinity by which they are variously attracted to other bodies.

B: Well put, my good fellow, but I do not see how this “quality of affinity”, as you say, is intelligible.

A: How could you fail to grasp my meaning? This affinity is a natural cause of the varying attractions that we observe.

B: Yes, but what idea is it that we are to conceive of in contemplating this quality?

A: Ideas, Schmideas! I don’t understand why you and your lot keep insisting on having ideas associated with terms? I have a perfectly respectable causal account of our various attractions. What more do you want? I don’t understand this preference for “intelligibility” you seem afflicted with.

If CIT were true we ought to expect to find dialectical episodes like that caricatured in the above dialogue within the historical literature. Simply put, we don’t! Clarke’s response (provided above) comes as close as any, yet the issue there is not over how to understand the claim that ‘attraction’ is unintelligible, but whether the only natural forces are mechanical forces. The reason that we don’t find widespread confusion concerning the nature of “intelligibility” is, I contend, because of the general acceptance of CC as a constraint on language. Given that, De regt’s and Dieks’ assertion that standards of intelligibility vary is false.

Of course one might simply contend that while standards did not shift during this specific period they did at others. Well and fine. However, De regt and Dieks not only
contend that standards shifted during this period but that the shift in question is a paradigm example of shifting standards.

Historically-minded readers will observe that the preference for physical mechanisms can plausibly be related to the success of this concept within nineteenth-century science. But even in pre-twentieth-century physics causal-mechanical explanation was not always the norm. It is true that Newton’s theory of gravitation was criticised because it failed to conform to the Cartesian intelligibility ideal of contact action; its implication of actio in distans was unacceptable to most seventeenth-century physicists. But between 1700 and 1850 action-at-a-distance rather than contact action and causal chains dominated the scientific scene … It was only after 1850, as a result of the success of ether theories, that contact action and causal processes à la Salmon became an acceptable intelligibility standard again.262

As this passage makes clear, not only do De regt and Dieks believe that standards shifted they believe that the shift was from a causal-mechanical standard of intelligibility to a standard that permitted explanations in terms of action-at-a-distance. The central problem here is that the causal-mechanistic view of explanation was never a “standard of intelligibility” properly so-called.

The causal-mechanical view of nature was a standard of explanation for natural phenomena. It is true, as I suggested in my discussion of Boyle, that it was preferred in part because causal-mechanical explanations were readily intelligible. But this does not

entail that every non-mechanical explanation or concept will be unintelligible. The idea that it was a standard of intelligibility ignores two important facts.

First, even mechanistic philosophers like Leibniz did not believe that the standard applied to metaphysical explanations. It was not the case that every non-mechanistic notion was viewed to be unacceptable.

Second, as was pointed out in the discussion of Leibniz in chapter I, Leibniz did not view ‘action-at-distance’ to be unintelligible. On the contrary it was precisely because it was intelligible that he was able to determine that it was false. Moreover, I take it that one of the main points that Leibniz is making is that if we are inclined to think of it as unintelligible then we are committing a mistake. We can only find contradictions in terms that are intelligible. And Leibniz, Clarke and Newton endorse the view that this expression designates a contradictory notion. What this shows is that the causal-mechanical view of nature, while it was a standard of explanation within the domain of natural philosophy (as opposed to metaphysics), was not the “standard of intelligibility”. It follows from this, together with the fact that ‘action-at-distance’ was not and is not unintelligible, that the shift that De regt and Dieks have in mind here was not a shift in the “standards of intelligibility.”

I think it is fair to say that there was a shift nonetheless, but this was a shift in what was deemed acceptable as an explanation and it was indeed a shift from the mechanism that had up to then constrained explanations. Naturally, it is because they view explanations to be the mechanisms by which we are able to render theories intelligible that De regt and Dieks are inclined to see intelligibility standards as shifting.
If there was a shift in the core conception of intelligibility this has not been demonstrated by De regt and Dieks. And as I argued in chapter II no such shift is in evidence.

The final concern I have with SST is the structural concern. For the sake of argument let us assume (contrary to the facts) that ‘action-at-a-distance’ was thought by Leibniz and his contemporaries to be unintelligible—as De regt and Dieks contend. Let us also accept the idea that after a period of time the expression became intelligible. The question I want to raise concerns what accounts for this apparent shift. From the standpoint of CC there is clear answer regarding what would have to happen: The expression could only become intelligible by being associated with an idea. But what account does CIT offer?

According to De regt and Dieks it simply gets incorporated into our “conceptual toolkit.” But a criterion of intelligibility is a constraint on which terms may be incorporated into our “conceptual toolkit” based on their cognitive content. Thus, it is not sufficient to account for the putative shift by pointing out that the expression was incorporated into our conceptual toolkit. That is just to make note of the fact that the shift occurred. What we want to know is what facilitated the shift. How did the expression come to be cognitively contentful? Moreover, in order for there to have been a genuine shift there must have been a change in the conception of what it is for a term to be cognitively contentful—otherwise it seems clear that the term will have taken on a new meaning, or else it was genuinely intelligible all along. I don’t see any response to this worry can be provided from the standpoint of CIT.

There may be a temptation to say that what happened is that we gained an intuitive feeling for the expression. Yet if the standard of intelligibility—what makes it
the case that a term is sufficiently cognitively contentful—is whether we have an intuitive feeling for it, then no shift in intelligibility has came about. Rather, we simply find some term intelligible that we formerly did not based on precisely the same criterion. Of course, this would be an overly vague account of intelligibility, and in any case I am not sure that this is the sort of thing that either De regt or Dieks would want to say. I conclude that there is no evidence for SST. Nor do I think that standards can shift in the relatively short time periods the authors suggest. One implication of MHC is that intelligibility is a relatively fixed product of our cognitive constitution.

6.3.3. CIT’s Holistic Character

A notable feature of CIT is its holistic character. It is not individual terms or expressions that have the property of be intelligible but whole theories. Here again the problem is one of fit with historical reality. Generally it wasn’t a whole theory that was deemed unintelligible but specific terms that seemed to have referents that could not be conceived. It is true that Boyle rejected the whole spagyricist program but this was because it made rampant use of notion like the ‘sulphuric quality’ of bodies and the like. It was the presence of these terms that made the theory unintelligible.

In general, though, it is terms and not whole theories that are the subject concerns about intelligibility—terms like ‘attractive force’, ‘affinity’ and latter ‘atomicity’. In principle it should be possible to ask about the intelligibility of terms independent of the theory. We should be able to ask whether terms like ‘liger’ or ‘chimera’ or ‘attraction’ are intelligible. We should be able to do so without rejecting whole theories. When some chemists in the nineteenth century worried over the intelligibility of ‘atomicity’ they did so in light of the contention of others (structuralists) who construed it to designate a
quality of substances. When it was considered as a merely quantitative notion the worry dissipates without any substantive change in the theory. There doesn’t seem to be any reason to preclude the possibility of considering a single term’s intelligibility. CIT does not capture this possibility and given that it does not fit the historical reality.

6.3.4. CIT’s Extrinsic Character

The final criticism I offer of CIT is that it is the wrong sort of criterion. CIT doesn’t provide us with an account of the cognitive mechanisms responsible for a theory’s intelligibility. Rather, it provides us with a test for judging whether a given individual finds a theory intelligible. The fact that a theory passes the test stated in CIT doesn’t tell us what the intrinsic property is that renders the theory intelligible.

Given De regt’s and Dieks’ use of the “oracle analogy” noted above I find this consequence ironic. The moral they draw from the “oracle analogy” is that a theory needs to be more than a list of true descriptions or predictions. It must tell us something about the underlying processes that are responsible for these the events which it predicts. They conclude from this that the theory needs to be intelligible. But the theory of intelligibility they provide doesn’t meet their own criterion in any substantive way. The most that we can say about the underlying process is that it involves an intuitive feeling. How is this anything more specific than what we might expect of an oracle? Surely, at a minimum the oracle has an “intuitive feeling.” Nonetheless, I don’t want to rest my complaint against the extrinsic character of CIT here.

I think that there is a deeper worry with the sort of extrinsic criteria offered here. To see this we can borrow an example from Wittgenstein. Suppose we wanted to come up with a criterion for when two people are playing chess. We might be tempted to say
something analogous to what De regt and Dieks say with respect to when someone understands a theory. Let us suppose then that we offer the following *Criterion for Playing Chess*.

**CPC**: A game of chess $C$ is being played by two individuals $Q$ and $S$ (in context $C$) iff $S$ and $Q$ have an intuitive feeling for the legal moves in $C$ without consulting a rulebook.

Now consider the following case: Suppose that two individuals are watching you and I play chess and further that these individuals have never seen or heard of the game. They may in fact not even be sure what exactly it is that we are doing beyond moving a few pieces around on a checked board. Suppose sometime after we finish we come back and observe them moving the pieces around in turn, and in the manner we did earlier. At this point we want ask two questions. First, are they playing chess? Second, does their activity satisfy CPC? If the answer to the first question is “No” and the answer to the latter question is potentially “Yes”, then CPC is false.

On the assumption that they are just mimicking our earlier behavior, we may conclude that the answer to the first question is indeed negative. What about the second? Part of the problem involved in answering the second question revolves around the murky question of whether or not they have an “intuitive feeling for the legal moves in $C$.” I don’t have any principled proposal for how we ought to clear up this question other than to appeal to De regt and Dieks. When it comes to CIT the only way that we have for determining whether or not it is satisfied for a given scientist is to “look and see” whether they actually make the right qualitative determinations. On that basis the individuals in our example appear to satisfy CPC. If this is right then CPC is false.
Of course we can’t conclude from this analogy that CIT is false. That’s not the point. The point is that CIT at best provides us with a symptom that goes along with finding theories intelligible. For all we know an individual might exhibit this symptom for very different reasons. Until we know more about the cause we run the risk of misdiagnosis.
Chapter VII

7.0.0. Conclusion

The criterion of intelligibility defended here situates worries respecting the *intelligibility* of terms within a broader framework of concerns over the *cognitive significance* of theoretical terms. The broader concern can be thought of, more generally, as a question regarding the cognitive mechanisms involved in our understanding of the theoretical language of science. A term’s intelligibility, I have argued, is to be understood as a function of but one of two complimentary mechanisms that facilitate our understanding of theoretical terms. On this account the sort of understanding involved in our apprehension of the theoretical language of science will ideally include both our grasp of the inferential role, or use, of the term within the theory as well as our ability to represent the referents of the relevant terms. Both mechanisms—the term’s “articulation” within the theory as well as its representational content—contribute to our understanding of the term. The model of cognitive significance appealed to here can thus be viewed as providing a pluralistic model of linguistic understanding.

A pluralistic model of cognitive significance yields immediate and decisive explanatory benefits. The principal benefit is that such a model affords a better explanation of the historical data. Simply put: unlike the present account neither of the proposals discussed in the last two chapters is able to fully explain the historical controversies wherein intelligibility or, more broadly, the understanding of theoretical terms was raised as an issue. Both views construe the cognitive significance of the theoretical language of science exclusively in terms of the manner in which the theory is
articulated. The deep problem for these views resides in the fact that they fail to capture the sort of concerns that motivated individuals like Boyle, Leibniz, Hume, and others to maintain that portions of the explanatory language being employed by their contemporaries were unintelligible. As demonstrated in the early chapters, these early modern theorists simply weren’t principally concerned with the way in which these terms were being used (in the relevant sense) or with the empirical adequacy of the theories in which they played a role. Rather, they were expressing worries respecting our general capacity to conceive of, or represent, the items being referred to by the terms in question. Equally, the views canvassed in the last two chapters are unable to account for the transitions within chemical theory that occurred in the 18th and 19th centuries and which were prompted in part by the competing views concerning the cognitive requirements of theory. These views arose in the critical debate between positivists and structuralists. In the context of this exchange not only did positivistic prescriptions concerning cognitive significance fail to be persuasive but it is arguable that had they been successful in this regard, science, at least chemical science, would look much different today.

My central argument for the criterion of intelligibility offered here is thus an explanatory-historical argument and I am happy to let its defense rest largely on its laurels in this domain. Still, I think there are prima facie intuitive grounds for adopting the proposed pluralistic attitude concerning the nature of a term’s cognitive significance. Indeed it seems intuitively right that there is something in the early modern discussion of “material substrata” and the like that distinguishes the discussion there from, say, Boyle’s effort at explaining chemical phenomena in terms of what he called “catholick principles”. It seems, that is, that we are indeed to understand the latter language in a
manner that we are unable to understand the former. If this is right, then on the assumption that there is a coherent framework in which to situate talk about ‘material substrata’, then there is here prima facie evidence in favor of the view that the comprehension of theoretical terms cannot be understood simply as the product of a homogenous cognitive process involving the grasp of a term’s usage. At the same time, it seems incorrect to hold out that we understand nothing of the talk of a ‘material substratum’. We appear to understand distinctly the difference between this language or practice from that in which ‘attractive forces’ figure. These observations taken together suggest that a more complex picture of cognitive significance is in the offing and it is that picture which I have attempted to develop and defend in the foregoing chapters.

Of course, our intuitions may fail us. None of the arguments I have offered can license the conclusion that there isn’t a third view waiting in the wings; one capable of reducing one of the proposed components of a term’s cognitive significance to the other in a substantive and historically adequate way. Rather, the arguments provided here and above are offered in the naturalistic spirit of raising the empirical plausibility of my proposal over and above that of the alternatives.

Having said this there is work to be done. Little has been said here concerning the nature of how we come to grasp a term’s articulation. Instead, the focus has been on developing an account of how we come to find terms intelligible. But the issue over the appropriate account of intelligibility cannot be detached from issues concerning conceivability, explanatory adequacy, and the like. The present account opens up terrain to be canvassed. On the question of what is conceivable, on the present view, we ought to frame the question in terms of which mechanisms our requisite to satisfy the claim that a
proposition P is conceivable. A natural consequence of the present view is that a term’s being cognitively significant may well fail to be sufficient to render it conceivable. This, for the simple reason that a term’s being well articulated may well suffice to render it cognitively significant though not intelligible. In other words, it would be quite natural to ask where in the nexus of cognitive significance “conceivability” might lie. With regard to explanatory adequacy it would be natural to ask at this point whether intelligibility is a necessary property of explanatorily adequate theories. I have suggested that it is not but the contrary view is not without its advocates.

Setting aside those issues, there are also questions of development. The account of intelligibility provided here is best viewed as a working hypothesis. There is still work to be done in furthering the account. What, for instance, is the nature of the association between words and their representational correlates? Some hand waving was done in the direction of suggesting means by which the articulation of a term might constrain the sorts of representations that suitable to rendering a term intelligible. Ultimately, it will be necessary to specify the connections between the components in more detail. Moreover the concept of “association” which admittedly does some heavy lifting will need to be fleshed out. In addition to these related projects the view defended here raises the issue concerning the role of models in understanding. It does so in two respects. First, by making cognitive models central to the account of intelligibility it raises the issue of their nature as well as manner in which they come to fulfill the task of rendering terms intelligible.

Second, it raises the issue of how models in science generally contribute to our understanding of the theories they play a role in. To take just one example, we might well
ask what the periodic table contributes to our understanding of chemical discourse. The answer, I strongly suspect, cannot be “nothing that we couldn’t have gotten otherwise.” While it may be true that whatever information the table captures could have been otherwise encapsulated, and while it may be true from a logical point of view that there is nothing special about its organization, it isn’t plausible that from an empirical view that it contributes little to our understanding of the chemical world or the theories offered in explanation of that world. If this is right then there is a significant question concerning the manner in which such models make their contribution. These are all empirical questions that in my view have ultimately to do with human psychology. They are significant questions which aren’t going to be answered by rational analysis but through empirical reflection with an eye toward developing reasonable empirical hypotheses.

That is what I have aimed here. I have attempted to provide an empirical hypothesis adequate to the data. There are gaps. There are holes. There is work to be done. But with what empirical hypothesis in science is that not the case. If you want more, turn your efforts toward speculative metaphysics.


----- (1710). Reprinted; *Theodicy; essays on the goodness of God, the freedom of man, and the origin of evil*. Edited with an introd. by Austin Farrer; translated by E.M. Huggard (New Haven, Yale University Press), 1952.


