

# Reference and Reinterpretation

by

Anthony Kulic

A thesis  
presented to the University of Waterloo  
in fulfilment of the  
thesis requirement for the degree of  
Master of Arts  
in  
Philosophy

Waterloo, Ontario, Canada, 2007

©Anthony Kulic, 2007

## **Author's Declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Reference is the relation held to obtain between an expression and what a speaker or thinker intends the expression to represent. Reference is a component of interpretation, the process of giving terms, sentences, and thoughts semantic content. An example of reference in a formal context involves the natural numbers, where each one can be taken to have a corresponding set-theoretic counterpart as its referent. In an informal context reference is exemplified by the relation between a name and the specific name-bearer when a speaker or thinker utters or has the name in mind. Recent debates over reference have concerned the mechanism of reference: How is it that we can refer? In informal contexts, *externalists* see the reference relation as explicable in terms of the salient causal relations involved in the naming of a thing, or a class of things, and the ensuing causal chains leading to a term's use. Opponents of this view—*internalists*—see the reference relation as being conceptually direct, and they take the external approach to rely on untenable metaphysical assumptions about the world's structure. Moreover, some internalists take the permutability—i.e. the consistent reinterpretation—of certain referential schemes to confound the externalist picture of reference. In this thesis I focus on the reference of theoretical terms in science, and I argue for an externalist treatment of natural kinds and other theoretical elements. Along the way I offer a defense of the externalist's pre-theoretic metaphysical assumptions and emphasize their central role in the interpretation of scientific languages. The externalist approach acknowledges the necessary constraints on reference-fixing that account for the schemes we employ, and this, I argue, confounds the permutation strategy.

## Acknowledgements

Countless thanks are owed to my thesis advisor, David DeVidi, who patiently endured re-write after re-write of each chapter; and whose astute direction is responsible for any substance and clarity contained herein. In addition, I'd like to thank the members of my thesis committee and other faculty members at the University of Waterloo Department of Philosophy whose admirable instruction made this thesis possible: Chris Eliasmith, Tim Kenyon, Paul Thagard, James Van Evra and Steven Weinstein. I am also respectfully indebted to the UW Department of Philosophy for their generous financial support in the form of a graduate fellowship.

Many thanks are owing to a distinguished set of interlocutors who contributed to this project in one way or another: Chloe Armstrong, Sam Cowling, Kyle Mencken and Steve Russell. The next round is on me. Thanks also to Quinn Dupont for help with  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ .

I'd also like to thank Jeffrey Foss, Scott Woodcock, James O. Young and the rest of the faculty and support staff in the Department of Philosophy at the University of Victoria for their fine instruction and assistance.

Finally, thanks to my parents, Tony and Maria, for all the help and support along the way.

*For my daughter, Ivana, ...the meaning of 'meaning'.*

# Contents

Author's Declaration .....	ii
Abstract .....	iii
Acknowledgements .....	iv
Dedication.....	v
<b>Contents</b>	<b>vi</b>
<b>Chapter 1</b>	<b>1</b>
1.1 Introduction: Reference in Language and Science .....	1
1.2 Realism and Anti-realism: A Primer.....	3
1.3 Russell's Structuralism and Newman's Objection .....	7
1.4 Ramsey, Carnap and the Received View of Theories...	10
1.4.1 <i>Ramsey</i> .....	11
1.4.2 <i>Carnap</i> .....	12
1.5 Models, Theory Modeling, and the Semantic View....	14
1.6 Realism, Truth and the Semantic View of Theories.....	19
<b>Chapter 2</b>	<b>24</b>
2.1 Putnam's Model-Theoretic Argument .....	24
2.2 Putnam's Paradox.....	29
2.3 Putnam and the Permutationists .....	31
2.3.1 <i>Quine</i> .....	31
2.3.2 <i>Field</i> .....	34
2.4 Causal Contexts .....	36
2.5 The "Just More Theory" Gambit.....	38

<b>Chapter 3</b>	42
3.1 Externalism and Internalism.....	42
3.2 Natural Kinds and the Causal Theory of Reference....	44
3.3 Naturalism and Constraints on Reference .....	48
3.4 Internal and External Questions .....	50
3.5 The Ontological Status of Causality .....	52
3.6 Natural Kinds and Realism Revisited .....	57
3.7 General Conclusion .....	60
Bibliography .....	62

# Chapter 1

"Reality is that which, when you stop believing in it, doesn't go away."— Philip K. Dick

## 1.1 Introduction: Reference in Language and Science

The first clear case of my daughter's referring in the English language came with her utterance, "dog". By 'referring', I mean that after a lengthy span of random babble, "dog" was the first bit of language she used repeatedly and correctly in the presence of an appropriate stimulus. Her repeated utterance of "dog" in suitable situations is what makes it reasonable to conclude that she is using the term in reference to the proximate dog.<sup>1</sup> However, her ability to do so is only the result of agreeable reinforcement from onlookers upon instances of her successfully aping their behaviour.

One way to conceive of the situation is that of an everyday causal mode of linguistic acquisition, where a term is passed from some previous user(s) to a new one; the new acquisition being one link in an historical chain of similar causal links from user to user. Some have thought to conceive of the term 'dog' as originating in a dubbing, or baptismal, ceremony, where its name was assigned as part of the name-related linguistic practices of our society. It is helpful to conceive of the term's meaning in this causal-historical way, according to proponents of this view, because it provides, in the form of a direct causal chain of events, an account of the nature in which the name of some thing can be said to refer to that object.

---

<sup>1</sup>Here, for commonsense reasons, I am setting aside Quinean worries about the inscrutability of reference—that she may have meant, for example, 'undetached dog parts' or 'temporal dog slices'—since she is only a toddler who presumably lacks the conceptual wherewithal to construct those alternatives.



However, with advancements in biological science the term "dog" (specifically, its scientific equivalent "*canidae*") has become a theoretical term used to denote a biological *kind*, examples of which are taken to satisfy certain conditions for their designation as such. Clearly my daughter does not use the term in the preceding sense. Nevertheless, over time elements of the scientific description of "dog" will permeate her understanding and, via a 'division of linguistic labour' (Putnam, 1979b), she will refer to dogs in a sense quite different from the one reflecting her present stage of development.

Another way to causally conceive of the scenario is as follows. When my daughter points and utters "dog" in the presence of one, information is deflected/emitted from the proximate dog and the specific information engenders structured neural activity that prompts her utterance from an associated memory of the affirmed correlation between the term and its associated stimulus. The preceding explanation is couched in terms of more fundamental causal processes and mechanisms—ones that are assumed by the natural sciences to obtain. Explanations of this sort usually involve more 'basic', non-ostensive elements of the natural world in such a manner as to make them available for reference. Contrast this with the former referential account of "dog", where the term's reference is conceived as being 'borrowed' from other users via a series of historical causal relations linking back to the term's baptism.

In learning, many of the terms we employ seem to be fixed in the manner of the preceding 'dog' example—by rote conditioning during our formative years. Names are attached to specific ostensive things, or classes of things, and this enables for a certain degree of successful communication between members of a linguistic community. In contrast, other terms we use are invariably theory-laden: using them correctly depends on some knowledge of the meanings of the terms used to define them, and this usually involves some understanding of a term's role in a greater network of term-meanings, as well as grasping its meaning relative to appropriate contexts for its use. Fortunately, over time and in varying ways we learn how to use such terms and apply them to aspects of the world.

Cases like my daughter's acquisition of the concept "dog" and her subsequent use of the term in reference are, from a certain *external* perspective, straightforwardly causal: they can be explained in virtue of the term's causal history and/or the causal processes and mechanisms upon which those histories are taken to supervene. In recent years the 'causal', or 'direct', approach to providing an account of reference for *natural kind* terms in science has become something of a philosophical movement. A question I engage in the latter portion of this essay is whether, and to what extent, causality can work to successfully anchor the theoretical terms

of scientific theories to their supposed referents.

Science is incontrovertibly our best epistemic engine. Anyone wishing to seriously challenge this view must explain its overwhelming success. How is it that helicopters work the way we intend them to, that antibiotics do what we expect and design them to do, and that we can accurately predict—down to the minute—each arrival of Halley's Comet? One popular view motivated by the success of science, yet sensitive to its history, takes mature scientific theories—the ones used to underwrite our scientific projects and experiments—as *approximately* true: where those theories err they are relatively negligible errors of detail. Were this not the case—so the argument goes—the overall success of science would be miraculous (Putnam, 1975: 173). The preceding general argument is taken by some to provide the justificatory basis for *scientific realism*.

Another question I engage—specifically in this introductory chapter—is how, if at all, a scientific realism born from an abductive inference from the success of science relates to orthodox metaphysical realism? My aim in examining the preceding question is to provide a backdrop for another question central to this essay: the legitimacy of Hilary Putnam's model-theoretic argument against metaphysical realism. In order to examine the nature of realism in a scientific context I consider several strategies for construing and expressing what our scientific theories say. Each of these strategies attempts to make clear the relations between scientific language and reality, and between truth, knowledge and belief. I show that, in light of problems illuminated by these theoretical reconstructions, we can (cautiously) accept the gist of a weaker realist stance—a view I call *proto-realism*—yet reject the orthodox form of metaphysical realism.

## 1.2 Realism and Anti-realism: A Primer

Realism concerns the truth and meaning of entire sentences, propositions, and beliefs (henceforth, 'sentences+'), and it can be expressed in two basic theses. First, sentences+ about particular entities, certain classes of entities, or some subject matter have (on suitable occasions) *truth conditions* which are mind/language-independent. Equivalently, one might say that there are *objective* conditions of the world that must obtain in order for sentences+ to be true. Second is the thesis that some of the sentences+ about those entities or subject matter *actually are* true, and they are true independent of whether human beings, or any other knowers, take them to be so. Such realists apply these theses to the world conceived as an object itself; they assume that the universe exists and that it does so independent of how

we take it to be. Moreover, and more importantly, metaphysical realists append this view with the following proviso: unless the world consists of some *fixed totality* of metaphysical 'world-stuff' (objects, properties, entities, events, relations, etc.) to which we can determinately refer, sentences+ about that stuff cannot be objectively true. Since knowing about the world involves some measure of equivalence between the content of our sentences+ about the world, and the way the world really is—i.e. a *correspondence* relation—without the independent existence of 'world-stuff' to fix truth conditions for those sentences+ we can have no *objective* true beliefs about the world.

Beliefs in this context should be construed as phenomenological evaluations of situations in the world. That is to say, our sensory modalities provide most of the raw material about those situations, and our beliefs constitute a cognitive appraisal of that raw material. In other words, our beliefs are judgments about certain *representations* of those situations, and those judgments, or attitudes, are themselves a form of valuative representation whose intentional object is the more basic, situational kind of representation. Scientific realists hold that science is the best mechanism by which to calibrate our beliefs about the world. This is so not only because science has proven overwhelmingly successful, but also because of the methodology underlying scientific investigation.

The 'hard sciences'—physics, chemistry and certain areas in biology—provide the most exhaustive inventory of metaphysical world-stuff. They do so by indexing the world's measurable properties, tracking the relations those properties fall into, and inferring theorems about how those properties change under varying conditions. Scientists use the data collected from experimental observation to determine the kinds of things there are, some of which are not directly observable. Unobservable *theoretical* posits are usually inferred from a combination of the available data and laws held to govern the behaviour of other well-understood posits.

Some scientific realists (e.g. Boyd, 1983; Psillos, 1999) see the aim of science as uncovering which sentences+ about our domain of naturalistic discourse are true. These realists take our best scientific theories as true, or approximately true, *linguistic descriptions* of their intended domains. Proponents of this view take theories to be linguistic entities and, as such, they take the theoretical terms of mature theories determinately refer. On this view, if a scientific theory is true, it is also true that its posits genuinely inhabit the world.

It is possible, however, to agree with the metaphysical realist that the world exists independently of how we believe it to be, and agree that science is our best epistemic engine, yet deny the first realist thesis—that the truth conditions of sen-

tences about certain entities and subject matter are mind/language-independent. Semantic antirealists offer an alternative account of what it takes for sentences+ to be true that is motivated, in part, by how mysterious the realist's, so-called, correspondence relation between sentences and the world actually is. The precise meaning of 'correspondence' has been exceedingly difficult for metaphysical realists to make clear. In order to avoid opacity they must provide a satisfactory account of 'facts'; that is, precisely explain the relations between facts, terms/ words, sentences, and the world. But realists have failed to achieve broad consensus on any one such account, let alone offer one without important difficulties.<sup>2</sup>

Metaphysical realists claim that the truth conditions for sentences+ are determined independent from the minds that apprehend them, so it follows from this that even our best scientific theories—the theory of evolution by natural selection, the double-helical model of DNA, quantum field theory, etc.—might still be false. To be sure, no reasonable person will deny that even our best theories are defeasible, but semantic antirealists are skeptical about the metaphysical realist's view that there is an *objective* sense of truth that obtains independent of knowers. For example, holding a particular theory as strongly confirmed, predictively successful and coherent with relevant evidence is insufficient for the orthodox realist to reckon that theory *objectively* true. This is an important point of departure between orthodox metaphysical realists and other realists for whom the above—in addition to other *theoretical virtues* (i.e. simplicity, elegance, fecundity, scope, etc.)—constitute the best possible ground for holding a theory objectively<sup>3</sup> true.

Furthermore, it is not clear to the semantic antirealist that grasping a truth condition for a sentence+ involves anything over and above understanding some minimal evidentiary condition for its assertion (or belief). If this view is correct, then both language and knowledge are inherent to the notion of truth. After all, grasping a truth condition involves, among other things, having some notion about what constitutes evidence for a warranted assertion (stated, presumably, in some language). So, semantic antirealists view the grasping of a truth condition for an assertion as more akin to a practical skill. In highly-specialized sciences this translates as having warrant (be it via some mathematical proof, some empirical ground, or some composite of corroborative evidence) for, say, an existential claim for a specific entity. All this, of course, runs contrary to realist orthodoxy, which

---

<sup>2</sup>See Lewis' "Forget about the 'correspondence theory of truth'", *Analysis* v.61, n.272, October 2001, pp. 275-280(6) for a analysis of the many problems plaguing the correspondence theory.

<sup>3</sup>I use scare quotes here to highlight the distinction between the absolute sense of 'objective' held by the realist, and the pragmatic sense of 'objective' held by the semantic antirealist.

takes truth to be an wholly non-epistemic condition. In sum, semantic antirealists disagree that truth is a property provincial to metaphysical realists alone.

Another important strain of anti-realism is error theory. Although we often deploy a realist semantics in domains where evidentiary constraints are troublesome in principle (e.g. purely theoretical domains, where entities and their attendant properties are wholly inferred), certain error theorists claim that the *prima facie* 'true' claims we make about those domains are, strictly speaking, false. Consider Hartry Field's (1980) notable error-theoretic account of mathematics. For Field the truth of the sentence, '23 is prime', involves an implicit existential assertion about an abstract object—in this case, the number 23. Platonists are realists about numbers so, in addition to the view that the number 23 exists, they hold it bears the property of being prime independent of our (or any other knowers') ability to know so. In their criticism of Platonism, nominalists often point to an important distinction between abstract and concrete objects: the former are spatiotemporally nonlocated and causally inert whereas the latter are inversely so. For a nominalist like Field the distinction demarcates the boundary separating things that exist from those that do not. Field is therefore an error-theorist with respect to arithmetic: mathematical objects are abstract, so they do not exist. Hence, a sentence like '23 is prime' is in principle false. However, despite its literal falsehood we should continue to deploy a realist semantics in the domain of arithmetic because natural science—an *ex hypothesi* good thing—requires us to do so.

Field's approach is similar to Bas van Fraassen's treatment of scientific theories. For van Fraassen, science continues and *ought* to continue deploying a realist semantics with respect to our best theories, but we cannot be certain about the truth of existential assertions involving unobservable entities. Hence, it is always more rational to be agnostic about the truth of theories involving unobservable entities than to wholesale accept them as true.

It is apt to conclude this section with an emphasis on anti-realism since the segment immediately following introduces an argument that anticipates the strategy employed by Putnam in his influential argument against metaphysical realism—a principal topic at issue in this essay. The 'Newman objection' to Russell's structural realism is relevant because it shows that purely structuralist meta-theories all succumb to the same difficulty: they carry the unwanted and unacceptable implication that scientific theories are not empirical.

### 1.3 Russell's Structuralism and Newman's Objection

An important philosophical concern has been the problem of reconciling the prevailing scientific view with our 'folk theory' of the world that includes familiar objects of everyday experience (e.g. apples, people, mountains, etc.). Empirical science is loosely partitioned into domains of study, each of which has a vocabulary used to name and classify the things it takes to exist. But squaring elements in one domain with elements from another has been exceedingly difficult. Consider, for example, the difficulty in specifying the conditions for something like an ecosystem in the language of particle physics. Even more perplexing is the notion of specifying something like a bicycle—an object whose existence nobody denies—in purely physicalist terminology. However, the twentieth century ushered in major revolutions in both physics and logic that equipped philosophers with what an ambitious few of them believed were the formal resources with which to successfully reconcile the phenomenal and theoretical worlds. One such attempt was advanced by Bertrand Russell.

Russell's theory of theories in *The Analysis of Matter* (1954) is the historical predecessor of what is nowadays called *structural realism*, or *structuralism*. His version is rooted in a form/content distinction where the objective properties and relations of the mind-independent external world are its structural properties and relations, which are known indirectly by scientific description. The subjective, everyday properties and relations constituting the content of human experience are intrinsic, and known directly via the senses. On this view, the external world-structure consists of second-order properties and relations of those intrinsic, first-order phenomenal properties and relations. It should be noted that when Russell uses the terms, 'structure' and 'structural properties', he explicitly means sets (some of whose elements may themselves be other sets). When referring to our phenomenal experience—i.e. how things seem to us—the terms we use denote properties and relations of individuals, whereas our theoretical vocabulary denotes properties and relations of those phenomenal properties and relations; that is, they refer to sets of sets. So, for example, upon hearing a gunshot that I describe as a 'sound' that is 'loud' and 'shrill', such terms refer to first order, intrinsic properties and relations that constitute—in Russell's dated jargon—my auditory "percepts". In contrast, a measurement of the same phenomena in terms of, say, decibels, constitutes a value for a variable of a second-order, structural property. This reconstruction mirrors Russell's definition<sup>4</sup> of cardinal numbers as being constructed from sets of sets.

---

<sup>4</sup>In *Principia Mathematica*.

A short digression is relevant to discuss some features of Russell's overall strategy, and the general worries that motivated him. Reference to individual, concrete things—like persons, soccer balls and geographic regions—are particularly problematic for someone like Russell, who has an *internalist* view of representation that emphasizes percepts (sometimes called 'sense data'). Such external, worldly objects are paradigm cases of things we take to exist, but which are not, strictly speaking, denizens of the internal world; they are not the kinds of things with which we are *acquainted*. Russell's descriptivist strategy is to explain the capacity to refer to concrete individuals in terms of the capacity to refer to the properties and relations that are exemplified by those individuals. His description theory of reference is an outgrowth of Gottlob Frege's thoughts about reference, with some important philosophical differences.<sup>5</sup> The descriptivist theory construes terms/names as denoting phrases that are analyzed into existentially quantified logical constructions. Such phrases denote—or refer—to the extent that there is an object that satisfies the definite description. On the descriptivist view, the reference of a term/name is mediated by the proposition in which it features. So, on the Russellian picture, the values for existentially quantified variables in each denoting phrase are provided by our percepts.

Russell was aware of, and motivated by, the philosophical problems that accompany claims of determinate reference for scientific terms. First, one need only undertake a cursory review of the history of science to appreciate that it reads like a chronicle of outmoded theories. Moreover, as theories are succeeded—especially during occasions of *revolutionary* change—a successor theory will often disavow an earlier theory's posits. A classic example—and one that Russell had, at the time, just witnessed—is that of luminiferous aether, a theoretical posit taken at the time to be the medium in which light propagated. Aether theory was eventually superseded following Einstein's advancement of the special theory of relativity in 1905. The point is that, clearly, any such example should diminish our confidence that the entities existentially quantified in even our best scientific theories have concrete existence. But another problem pervasive in science, and particularly disconcerting in physics, is that most classes of scientific entity seem systematically explicable in terms of sets of more basic entity classes. Hence, we face the

---

<sup>5</sup>Here it should be mentioned that, whereas Frege and Russell are often taken as sharing a similar descriptivist strategy, their respective conceptions of a proposition differ significantly. Specifically, Frege rejected the view that a physical object could be the constituent of a proposition. Russell disagreed, holding that propositions might indeed have physical objects as components, claiming that such propositions could be grasped only by someone who is acquainted with all of the constituents of the proposition, where acquaintance requires the kind of perfect and complete knowledge that we can have only in virtue of universals.

venerable problem of infinite ontological regress.

Russell's theory was meant to allay these concerns by explaining the theoretical predicates of scientific theories non-reductively,<sup>6</sup> since they are only indirectly related to the world. That is, the reference of theoretical posits can be explained only in terms of the reference of the observational properties and relations of the theory; this curtails the problem of ontological regression because all theoretical terms refer to properties and relations that constitute second-order (or higher) subclasses of properties and relations of first-order individuals. In other words, the ontological buck stops at phenomena. To reprise an earlier example, if an instrument is used to measure the sound of a gunshot, the sound might be measured and quantified in decibels. 'Decibel', given this picture, is a higher-order property whose referents are the related (first-order) percepts of the instrument-reading, gunshot-hearing observer. Theoretical terms refer to the world only indirectly because their referents are always first-order phenomenal properties. Structural realists take this view to dispel concerns over theory change in the following way. Although the conceptual posits of a particular theory may be succeeded by another that is, for example, more elegant, predictively powerful, and coherent with available data, both its observations (and their consequences) and (many of) the theory's substructural elements are preserved from one theory to the next. So, when theories change the primary ingredients in Russell's reconstructions are taken by structuralists to survive the transition.

Russell's view came under fire in a 1928 article by Cambridge mathematician, Max Newman, whose objection concerned the obvious empirical character of scientific theories. Specifically, Newman held that on any reasonable account of scientific knowledge it must be the case that true scientific claims are non-trivial. Russell's account of theories, Newman argued, fails on this commonsense requirement. Newman showed that any set of empirical objects of sufficient cardinality can be used to construct any theoretical structure, since a structure is merely a set of sets of objects. Scientific theories, on Russell's view, provide structures that are second-order properties and relations of our observations. In set-theoretic terms these properties and relations translate as sets of sets of sets of...objects. But any second-order structure which satisfies any basic class of objects of a given cardinality up to isomorphism also satisfies any other class of objects of the same cardinality. This, of course, gives rise to important questions regarding the evidential basis for choosing a particular structure, since isomorphic structures are constrained *only* by cardinality. That is to say, if scientific theories represent the

---

<sup>6</sup>I use "non-reductively" in the sense whereby reduction constitutes explanation only in more-fundamental, theoretical terms.



world only in virtue of such isomorphisms holding, then knowing that a scientific theory is true amounts to knowing only that the theory's target system—what the theory represents—has a specific set-theoretic structure. The problem underscored in Newman's critique was that Russell's theory offers no way to distinguish an empirical structure from a non-empirical, albeit isomorphic equivalent. Newman showed that any and all isomorphic structures that satisfy the phenomenal vocabulary also come out true, making a purely structural theory of the world only trivially true. Hence, in order to evade triviality an account of theories requires some way to distinguish the empirically relevant from the empirically irrelevant isomorphic structures. But, as Russell duly conceded,<sup>7</sup> the requisite distinction cannot be drawn from within a purely structuralist framework.

Newman's argument is significant for a number of reasons. Mainly, it threatens the realist component of Russell's doctrine. If structure constitutes the *objective* properties of reality that we can know, yet *any* isomorphic structure satisfying the phenomenal vocabulary of the theory comes out true, our scientific knowledge of the world—purportedly the best kind of knowledge there is—is only trivially true. Hence, the structural realist faces the problem that her formal reconstructions of scientific theories require an *interpretation* to evade triviality.

## 1.4 Ramsey, Carnap and the 'Received View' of Theories

Another respect in which the Russell-Newman exchange is significant is the role it played in influencing subsequent attempts to formally reconstruct our scientific knowledge.<sup>8</sup> Russell's logic and his ensuing formulation of scientific theories served as the launching platform for the philosophical movement known as Logical Positivism. The notion that scientific theories might be expressed as sets of truth-bearing sentences in the language of logic strongly appealed to a group of (predominantly) German and Austrian philosophers and scientists. They believed that rational reconstructions of our theories would enable them to display, explicitly and clearly, the relationships of meaning and evidence that confer a special status on the sciences. The Positivists' approach to theories was undertaken as part of a more general program aimed at discrediting what they saw as a disturb-

---

<sup>7</sup>Russell's concession came in a letter to Newman, quoted in its entirety by Demopoulos-Friedman (1985).

<sup>8</sup>Demopoulos-Friedman (1985) draws out these comparisons at greater length and in greater detail.

ing trend toward metaphysical obscurantism, popularized most notably by G.W.F. Hegel and his followers.

The 'new' logic—introduced by Frege and further developed by Russell and A.N. Whitehead—provided scientific meta-theorists with a formal vocabulary by which to express scientific inferences. However, Newman showed that the empirical character of scientific theory mandates an experientially legitimate interpretation of the structures which appear in them. To mitigate this difficulty Russell's followers assumed the notion that observable objects have observable properties, and they took this view to be unproblematic. In rational reconstructions such objects and properties were expressed in what became known as the 'observational vocabulary', which was taken by the Positivists as empirically meaningful. Moreover, any statement formed out of the logical vocabulary combined with the observational vocabulary likewise counted as empirically meaningful. Thus, the challenge for the post-Russellians (the Positivists, in particular) became the following: how might one give meaning to *theoretical* terms that are neither part of the observational vocabulary nor the logical vocabulary?

### 1.4.1 Ramsey

An influential figure sandwiched between Russell and the Positivists was Cambridge philosopher, F.P. Ramsey. Inspired by the Russell-Newman exchange, and troubled by the problem of defining theoretical terms, Ramsey developed a method for eliminating overt reference to theoretical posits in the formal reconstruction of scientific theory. Overt appeal to such entities was viewed as problematic because, after all, unobservables are not empirical—they are inferred from the continuities and changes among clusters of variable values—namely, measurements. In addition, the dubious ontological status of certain theoretical posits was illuminated by the recent Einsteinian revolution in physics, where, for example, terms such as "mass" were completely redefined.

Ramsey's strategy for answering the problem of defining the theoretical vocabulary runs as follows. Take the axioms of the formal system (for convenience, supposed to be formulated in first-order predicate logic) which expresses the scientific theory in question; substitute every theoretical constant with a distinct variable, then apply existential quantification over the propositional and property matrices thus obtained. Introduce laws, or axioms  $T(\Phi_1, \dots, \Phi_n)$  and correspondence rules (rules that explicitly link some of the theoretical terms to terms whose meanings are given directly in observation)  $C(\Phi_1, \dots, \Phi_n)$  which contain  $V_i$  terms  $\Phi_1, \dots, \Phi_n$ :

$$(\exists\Phi_{*1}) \dots (\exists\Phi_{*n}) [T^*(\Phi_{*1}, \dots, \Phi_{*n}) \wedge C^*(\Phi_{*1}, \dots, \Phi_{*n})] \quad (1.1)$$

where the  $V_i$  predicate terms  $\Phi_1, \dots, \Phi_n$  in  $T$  and  $C$  are replaced by the distinct predicate variables  $\Phi_{*1}, \dots, \Phi_{*n}$  thereby obtaining  $T^*$  and  $C^*$ .<sup>9</sup> Ramsey called the theory's observational vocabulary its 'primary system' and its theoretical vocabulary its 'secondary system'. Later Rudolph Carnap successfully proved that a sentence stated in the primary system follows from the theory if and only if it follows also from the Ramsey sentence of the theory.

The significance of Ramsey's contribution is this. Although theoretical terms and predicates are pervasive in scientific theories, the really important feature, as Newman stressed, is the theory's empirical content. In the legitimate reconstruction of a given theory, we need not treat its theoretical terms and predicates as "names", or entities. Rather, by treating them as genuine variables bound by existential quantifiers we can construct a truth-valued sentence instead of an open formula, and the sentence can thereby be used to express a cognitive judgement. Moreover, the Ramsey sentence implies that not all correspondence statements are false, which entails that the theory is realized by the world. Hence, implicit in the Ramsey sentence is the view that there are Russellian structures that realize the Ramsey sentence, but the Ramsey sentence itself does not commit the theory to the existence of any specific set of structures. Hence, a 'Ramseyfied' theory is understood as cognitively meaningful via the conjunction of its empirical content along with its implicit claim of realization.

### 1.4.2 Carnap

Strongly influenced by the work of Russell and Ramsey, Carnap developed a theory of theories as partially interpreted formal systems, and his mature formulation of this system became the paradigm by which Positivists construed scientific theories. Carnap (1967: 5) even adorned his *Aufbau* with an introductory motto from Russell: "The supreme maxim in scientific philosophizing is this: Whenever possible, logical constructions are to be substituted for inferred entities." Russell's form-content distinction was retained, as was the notion of a logical structure, but Carnap ambitiously held that *all* concepts of science could be explicitly defined within a single constructional system. In other words, he aimed for complete formalization where an entire theory might be expressed as a set of deductively

---

<sup>9</sup>Suppe (2000), 390.

closed sentences in the language of logic. Carnap's mature system was so influential among his peers and followers that it became known as the 'received view of theories'.

Recall that Russell formulated our objective knowledge by using 'experiential' properties to form higher-order theoretical structures. Carnap, on the other hand, follows Ramsey by eliminating higher-order structures and recasting them as variables of first-order individuals. Focusing instead on separating the factual from the non-factual content of a physical theory, Carnap took the factual to be exhausted by the theory's Ramsey sentence.

Carnap frequently emphasized that virtually every unreconstructed sentence of the language of science and everyday life has both factual and non-factual aspects, so that the sharp separation of our language into factual and non-factual sentences is meaningful only relative to a reconstruction of that language." (Demopoulos, 2003: 384)

Carnap then isolates the non-factual content by way of the 'Carnap sentence',

$$C(\theta) = (R(\theta) \implies \theta) \tag{1.2}$$

where C is the Carnap sentence, R is a Ramsey sentence, and  $\theta$  is the theory in question. The factual content of  $\theta$  is associated with  $R(\theta)$  because the observational consequences of  $R(\theta)$  and  $\theta$  are equivalent. Furthermore, the Carnap sentence—like the Ramsey sentence—expresses the claim that it must be the referents of the terms of the unreconstructed portion of the theory that satisfy the matrix of  $R(\theta)$ .

In reconstructing our theories, Carnap—like Russell and Ramsey before him—attempts to show that we can give an authentic representation of the theoretical knowledge offered by our best scientific theories. It is done by carving a clear distinction, within the theory's domain, between the observational and the theoretical. If we segregate the observable part of any intended model of  $\theta$  we can then introduce a set of observational predicates (O-predicates) into a reconstructed language for  $\theta$ . The intended model just is the unreconstructed theory, so the O-predicates are defined in terms of their being a subset of observable elements in the domain of  $\theta$ . The reconstruction is 'partially' interpreted because the interpretation of  $\theta$ ,  $I(\theta)$ , is extended only to observational terms, which are presumed to be given a complete empirical interpretation. The theoretical terms are left with only an *indirect* empirical interpretation provided by an implicit definition from within the axiom system. Like Russell, Carnap assumed that the logically primitive predicate expressions—ones we can justifiably assume to understand—are those restricted to genuinely empirical/phenomenal objects. Demopoulos (*ibid*:

386) observes that without such an assumption, "...the empiricist motivation for distinguishing between the O- and T-vocabularies—and, therefore, Ramsey's primary and secondary systems—would be lost." Thus, the factual content of  $\theta$  consists in its consequences expressed in its O-language, whereas claims involving the theoretical elements of the theory are understood as purely logical: their reference need only involve the appropriate logical category and arity.

The Carnap-Ramsey approach attempts to answer the question of how a theory's language is fixed according to its standard interpretation by altogether eliminating the theoretical vocabulary from a theory's rational reconstruction. So, despite its providing an interpretation for the reference of terms in a theory's observational vocabulary, it offers no interpretation for the reference of terms in the theoretical vocabulary; accordingly, the theoretical portion of Carnap's system is susceptible to Newman's criticism. This, of course, means that claims about the existence of specific theoretical entities are inferred from uninterpreted formal sentences. Consequently, critics of Carnap's programme have belabored the point that his system implies an instrumentalist, or antirealist, stance toward scientific theories—an unfortunate problem Carl Hempel (1958) dubbed, "the theoretician's dilemma".

The Carnap-Ramsey approach to theories is implicated in Putnam's model-theoretic argument against metaphysical realism, in that Putnam attempts to show that a realist variation on the Carnap-Ramsey strategy (i.e. one that extends an interpretation to the theoretical vocabulary) is susceptible to an argument parallel to that which Russell faced from Newman.<sup>10</sup> In the following section I discuss the apparent successor to the Ramsey-Carnap view of theories. The 'semantic', or 'model-theoretic', approach offers an altogether different strategy for formally reconstructing our scientific theories—one that some of its proponents believe permits for a less problematic treatment of the issues afflicting the received view (van Fraassen 1985: 289).

## 1.5 Models, Theory Modeling, and the 'Semantic' View

Recall that Newman observed a problem faced by any purely structural/ formal account of theories: since our knowledge of structure is clearly nontrivial—i.e.,

---

<sup>10</sup>It should be noted that Putnam does not cite Newman as influencing his formulation of the model-theoretic argument and, in fact, claims to have developed the argument independent of Newman's critique.

since we do not just *stipulate* the structures of our theories—it cannot be the case that our knowledge of the non-perceptible world is limited to knowledge of its structure. Acknowledgment of the preceding, and concession that it has proven virtually impossible to adequately capture a mature scientific theory in a formal, deductively closed, linguistic reconstruction,<sup>11</sup> has caused critics of the received view to coalesce around an alternative strategy for reconstructing scientific theories: the model-theoretic, or 'semantic', view of theories.

In science, 'modeling' a system is usually understood as involving the construction of a representation that helps describe and explain the system. Whereas, for example, engineers and architects design *a priori* models—e.g. schematic illustrations, scale replicas, 'blueprints', etc.—of a system yet to be constructed, 'modeling' a system is more often understood as a reversal of this process: an existing target system is examined relative to some general assumptions, and a model, or representation, of the system is created in abstraction and expressed in a language, or languages, deemed suitable to capture the system's salient features. One way to characterize the role of models in the practice of science is to view them as analogies. For example, the 'billiard-ball' model of gasses constitutes a depiction involving a set of assumptions about the motion and collision of an aggregate of gas molecules; hence, it constitutes a *representational analog* of the theory. From both an historical and practical perspective, the importance of analogical models in the *practice* of science cannot be overstated.

In the mid-1960's Mary Hesse advanced seminal work on analogical models by characterizing them as standing in a three-place relation involving a model *M* of a target system *X* that is relative to a physical system *Y*. Here, *Y* is understood as the source of a set of assumptions—sometimes called 'modeling assumptions'—on the basis of which the behaviour of *X* is to be investigated. If, for example, a modeler's target system is

*medium in which light waves are propagated*

the assumptions underlying *Y* will be, among other things, the associated physical laws governing optics. Hesse<sup>12</sup> characterized the relation between *Y* and *X* as involving,

1. some positive analogies, i.e. properties, or relations between properties, that both *Y* and *X* share;

---

<sup>11</sup>This claim was supported by Stephan Hartmann during his presentation "Unification and Coherence" at the Philosophy of Science Association Biennial Meeting at Vancouver, BC, 2006.

<sup>12</sup>The alphabetized points are lifted verbatim from Psillos (1999): 141, who paraphrases from Hesse (1966): 8-9.

2. negative analogies, i.e. properties, or relations between properties, in respect of which  $X$  is unlike  $Y$ ; and
3. some neutral analogies, i.e. some properties about which we do not yet know whether they are positive analogies, and which may turn out either positive analogies or negative ones.

The positive and neutral analogies between  $Y$  and  $X$  provide the conditions whereby a model of  $X$  might be constructed relative to a background of  $Y$ . The point of analogical models is to highlight the heuristic role played by  $Y$  in revealing important properties of a specified physical system  $X$ .

More recently, Ronald Giere (1988: 81) has developed and reformulated Hesse's analysis of the scientific model-building process in the following way. A model  $M$  of target system  $X$  based on source system  $Y$  *represents*  $X$  in the respects specified by the *positive analogies* between  $X$  and  $Y$ . In addition, the representational quality is graded in degrees that are specified by the conditions of approximation and the idealizations employed by the model. On this view, the assumptions that underlie  $Y$ —i.e. the so-called 'laws of nature'—are understood only as principles that guide the process of model-building.

Such a deflationary view of scientific laws stands in contrast to that of the received view of theories, where laws of nature are construed as *universal* laws meant to describe the *actual* behaviour of a designated class of *real* objects. Recall that under Carnap-Ramsey laws are expressed as nonlogical axioms that provide theories with most of their empirical content (or, from the antirealist's perspective, their instrumental value). But according to Giere, the Carnap-Ramsey view of laws as having empirical content is problematic given the degree of idealization<sup>13</sup> implemented in their statement—that is, either the intended class of objects described in the laws must be empty (in which case the laws apply to no real objects), or the assumed laws are just false. So, on the Gierean approach the question of a theory's truth is recast and expressed in terms of the respects in which the theoretical model and the real system are similar, and to what degree. Given the Gierean view

---

<sup>13</sup>Idealization here is meant to suggest that there are no 'real-world' systems answering to such laws. In expressing natural laws scientists assume theoretical models in which, for example, the energy of a moving body is held constant. This, of course, implies an absence of friction in the system. Scientists have to idealize as such for pragmatic purposes, since oppositional forces like friction are too complex to be captured by any even moderately complex formulae. So, whereas physical laws make explicit reference to things like Euclidean two-dimensional axes, perfect spheres, frictionless three-dimensional planes, resistance-free pendulums and the like, the class of such entities is nevertheless understood and assumed by proponents of this view to be empty, and the attendant laws are held to be, strictly speaking, vacuous.

of scientific theories, the question of *realism* is recast as one about the level of representative accuracy surrounding a family of models *that are the theory*, rather than over the truth of sentences or statements and how they correspond to the world.

The preceding discussion of analogical models is meant to highlight the important distinction between the sense of 'model' employed in a Hesse/Giere 'model-based understanding of theories', and the altogether different model-theoretic or 'semantic' view of theories. The sense of 'model' in the latter derives from the branch of mathematical logic known as formal semantics, or model theory. Model theory involves the interpretation of formal languages using set-theoretic structures and Tarskian 'T sentences' as a schema for truth. To a certain extent, the model-theoretic approach in the philosophy of science is motivated by the success mathematicians have had using model theory to clarify formerly obscure issues in both logic and mathematics, such as those surrounding the relationship between proof and truth. Patrick Suppes pioneered the model-theoretic approach to scientific theories in the late 1950's and developed it through the 1960's. On Suppes' account, a theory abstracts from a particular linguistic representation and focuses instead on what it is that makes the theory true.

The semantic view identifies theories with classes of model--types, where a theory is a class of fully articulated mathematical structure--types that are defined by the use of set--theoretical predicates (Suppes, 1967; da Costa and French 1990). Alternatively, theories might be defined by the use of the mathematical language that is dictated by the subject matter of the particular theory (van Fraassen 1980; Suppe 1989). The idea, in the former case, is to use the full apparatus of set-theory to specify the object, property and relation extensions of their linguistic counterparts in order to define the theory's terms. Hence, a theory satisfies some model once an effective mapping of the theory's vocabulary is linked to a set-theoretic representation of the target system's domain, with objects, properties and relations in a one-one correspondence via functions assigning subsets of objects to one-place predicates, two-place relations, and so on—such is the model's interpretation.

Standard interpretations are fixed via constraints on theories. *Theoretical constraints* are sets of sentences with fixed truth-values that stipulate the terms in question. They are usually expressed as the axioms or laws governing a particular theory, and they function as a central and explicit part of the theory. *Operational constraints*, on the other hand, function in the role of assertability, or experiential, conditions on empirical sentences; that is to say, they tell us what can be measured and observed. The idea is that empirical sentences have built-in test conditions that are not stipulated in order to provide the meanings of those sentences, yet the nature of those test-conditions constrain our judgment about when those



sentences are true. Our implicit understanding of these conditions constitute the pre-analytic and pragmatic background assumptions giving rise to a class of *intended* models, or intended interpretations, for theories. Operational constraints are not internal to, or part of, a scientific theory, but they function as necessary preconditions for a theory's creation and expression.

The preceding discussion of constraints should help to illuminate some of the difficulties faced by proponents of the set-theoretic version of the semantic view of theories—in particular, those who want it to bolster the case for metaphysical realism. Proponents of the view take an abstract set-theoretic representation as *isomorphic* with a theory, and the theory's truth is satisfied by such an intended model. However, if term-extensions are taken to be isomorphic with a set-theoretic structure, and if a theory's interpretation (intended or otherwise) is not an explicit part of the theory, the given structure is susceptible to Newman's complaint. Recall that Newman showed any such structure to be *ad hoc* without some *explicit* appeal to an empirical ground for choosing a specific structure.

Upon recognizing the failings of the foregoing account, Bas van Fraassen advanced, in the 1970's and 80's, an influential variation on the semantic view which abandons appeal to set-theory as a representative medium. Van Fraassen (1980; 1989) recognizes that an abstract, set-theoretic representation will not do;<sup>14</sup> one must instead be able to *identify* and specify examples of the theory as realized by the world. Sets, after all, should be ontologically suspect for any empiricist. So, van Fraassen's novel turn is the notion that *the world* models a theory, which affirms the classical view that theories are, to an important degree, 'read off' the world. Assuming such a view, Jupiter and its moons might constitute a model satisfying Newton's laws of motion plus the theory of universal gravitation.

The shift in emphasis to what *actually* satisfies, or models, the axioms of a theory constitutes a marked departure from both the set-theoretic and syntactic views, the latter of which emphasizes the axioms, their logical form, and what might be deduced from them. Recall that on the syntactic view one must first reconstruct the axioms of a theory into some appropriate formal language (introducing the relevant quantifiers and logical constants, etc.) and follow by introducing correspondence rules. In contrast, van Fraassen's semantic approach downplays concerns over the language in which the theory is stated. Any comprehensible language

---

<sup>14</sup>Van Fraassen implements a state space representation—a technique developed by physicists in the nineteenth century—as a formal alternative by which to express particular scientific theories. Since the details of the specific technique are immaterial to the philosophical points I mean to emphasize, a discussion of state space representation can be avoided here. Readers can refer to van Fraassen (1980: ch. 6) for a thorough discussion of state/phase spaces.

capable of making reliable distinctions between the objects satisfying the theory's axioms and those that do not will suffice. Hence, van Fraassen shifts the emphasis from the axioms—understood as linguistic entities—to the models, which are nonlinguistic, *physical* entities.

The semantic view of theories involves the implicit assumption,

$$\Theta \iff \Theta \text{ is true} \iff \Theta \text{ has a class of } \textit{intended} \text{ models,}^{15}$$

which means that a theory is true, where truth is understood as relative to some class of intended models. Stated somewhat differently, the semantic view *identifies* a theory with a class of intended models ( $\theta = m^{1\dots n-1}$ ), and the theory's truth hinges on the world being one of the class of intended models; which is to say, the theory just is the set of intended models.

## 1.6 Realism, Truth and the Semantic View of Theories

Recall from an earlier discussion that metaphysical, or orthodox, realism is captured in two central theses:

1. Sentences+ about particular entities, certain classes of entities, or some subject matter have (on suitable occasions) truth conditions which are mind/language-independent; and,
2. Some of the sentences+ about those entities or subject matter *actually are* true, and they are true independent of whether human beings, or any other knowers, take them to be so.

In an earlier discussion I rejected the part of thesis 2 which states that certain sentences+ are true independent of whether knowers take them to be so. The very notion of a truth condition for a sentence+, I suggested, implies grasping some minimally appropriate knowledge-condition for its warranted assertion or belief. The notion of truth, on this view, is inescapably tied to both knowledge and language. Given the preceding conception, truth is a property of sentences+ only relative to some linguistic/epistemic community.

In the mid-twentieth century, Polish logician Alfred Tarski offered a definition of truth for *formal* languages that is, by most accounts, still considered the paradigm. Philosophers quickly co-opted the Tarskian model and applied it to

---

<sup>15</sup>Demopoulos (2003): 391.

natural languages as a means of, among other things, expressing the conditions of truth for both everyday sentences and scientific propositions. When such sentences are recast under the Tarskian model they are known as 'T-sentences'. For example:

"The atomic mass of Barium is 137.327 amu" is true *iff* the atomic mass of Barium is 137.327amu.

The 'object language' portion of the T-sentence in quotations—left of the locution, 'is true *iff*'—stands in the relation of material equivalence to its 'meta-language' disquotation to the biconditional's right. But all the Tarskian algorithm ("*P*' is true *iff P*) tells us is that if it is legitimate to assert *P*, it is also legitimate to assert that "*P*' is true. So, if the Tarskian model is acceptable as a paradigm definition for truth—which the semantic view assumes—then it bolsters the view that the disquotation of a T-sentence merits substantial justificatory unpacking. Indeed, I assume this to be the case.

The question I intend to explore for the remainder of this section is the extent to which the semantic view of theories works to facilitate realism of theory. Specifically, I consider van Fraassen's *constructive empiricist* reformulation of the semantic view, since structural accounts—like both Russell's and the set-theoretic version of the semantic view—are susceptible to Newman's objection. Ramsey-Carnap, on the other hand, fails to support a robust realism by eliminating overt reference to a theory's theoretical vocabulary.

Van Fraassen (1980; 1989) uses empirical adequacy as an epistemic surrogate for truth of theory. That is, to the extent that the observable consequences of a given theory come out true, the theory should be considered empirically adequate. Constructive empiricists challenge both the existential and epistemic status of theoretical assertions; they doubt that one can ever be in a position to justifiably hold that the truth-conditions for certain theoretical assertions in fact obtain. Hence, constructive empiricists hold that the rational stance with respect to the truth of assertions involving reference to theoretical/unobservable entities is suspension of judgement, or agnosticism. Contrast this with the scientific realist, who stresses that there is ample warrant—at least in some instances—to believe that certain scientific theories have achieved a significant degree of theoretical truth. Alternatively, constructive empiricism motivates a picture of scientific practice where one need not be concerned over the issue of whether a theory gets its substructures right. Instead, the picture of scientific inquiry is one where the task of establishing the truth of certain theoretical assertions does not even feature among its ambitions. That is to say, van Fraassen offers an alternative *philosophical* view of sci-

ence: mature scientific theories are both successful and meaningful, but scientists neither aspire to, nor do they deliver theoretical truth by them—science aims only at achieving theories that are empirically adequate.

On van Fraassen's version of the semantic view of theories the distinction between what is and is not observable is not drawn on the basis of vocabulary (as with Carnap-Ramsey). Instead, the distinction is determined via the demarcation of empirical substructures.

To present a theory is to specify a family of structures, its models; and secondly, to specify certain parts of those models (the empirical substructures) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call appearances; the theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model. (van Fraassen: 1980, 64)

However, constructive empiricism shares with Ramsey-Carnap the view that to assert a theory as true is to say that there is, in fact, a model which realizes the theory. Recall that Ramsey-Carnap does so via an explicit—albeit non-specific—statement that the theory is realized by the world; van Fraassen does so by specifying the *empirical* parts of models as candidates for isomorphic "direct representation" by observers. We can therefore think of van Fraassen as a realist with respect to the unambiguously empirical parts of a theory, yet an anti-realist with respect to the sentences of a theory which make reference to posits that are in principle unobservable.

Earlier I indicated that the thrust of this section would be to investigate the degree to which the semantic view supports metaphysical realism. Under the constructive empiricist framework, a successful theory featuring substructures whose models cannot be described in experimental and measurement reports cannot be held as true. But since such substructures are, in principle, precluded from 'isomorphic direct representation', scientific realists view van Fraassen's criterion as austere. The unappealing result of such a rigid criterion is that a surfeit of established scientific theories must be taken as instrumental. Scientific realists reject the instrumentalist construal of mature theories because they take the unobservable/theoretical posits featuring in them as existentially quantified on account of reliable evidence. Moreover, scientific realists usually take a scientific community to be (in its corresponding sphere of investigation) the rightful arbiter of the conditions under which some posit might be justifiably granted existence. Hence, scientific realists reject constructive empiricism's canonical doctrine that it is always more rational to regard a mature theory as empirically adequate rather than true. In other words, they allege that constructive empiricists distort the intuitive

and conventional sense by which we attach ascriptions of truth to our theoretical claims.

Recall that a central motivation for scientific realism is Putnam's "no miracles argument" (1975: 73): it would be miraculous if our best scientific theories were not at least approximately true given their tremendous overall success. The argument captures a reasonable intuition but does little more than offer an abductive basis upon which to investigate the ways in which theoretical terms might actually refer, should we conclude, following Richard Boyd (1983, 1985), that a true theory's posits must do so. Van Fraassen, however, rejects the no miracles argument, despite its commonsensical force, because he rejects abduction<sup>16</sup> as a legitimate mode of inference.

The view that a theory's terms must determinately refer in order to uphold scientific realism has also been contested. The Gierean theory of scientific theories adopts a so-called realist attitude toward science while at once rejecting the notion that theories ought to be construed linguistically. His 'model-based' approach grades models—*qua* representations—in terms of similarity to their specific target systems, where the degrees of similarity are specified by conditions of approximation and the degree of idealization employed by the model. The relations between models and their targets are expressed in hypotheses that affirm, in phenomenal/cognitive terms, the relevant respects and degrees in which they are similar. The project can therefore be seen as an augmentation of Karl Popper's failed attempt to explicate theories in terms of verisimilitude, or truthlikeness, while still maintaining a realist *attitude* toward science.

Despite its billing, Giere's 'constructive realism' cannot be called a realist view if realism is understood in the traditional sense—that is, as involving a correspondence relation between what sentences+ say and certain conditions of the world. Giere rejects the notion that theories should be understood as truth-conditioned linguistic descriptions of their target systems, and he rejects the notion that theoretical posits must have determinable referents to uphold realism (Giere: 1999, 177). However, Chakravartty (2001) has observed that even constructive realism cannot—in its attempt to shift emphasis from correspondence between sentences and the world to similarity between features of models and their targets—avoid a retreat to sentences+ in the specification of similarities held to obtain. Moreover,

---

<sup>16</sup>Van Fraassen rejects abduction because he takes explanations to be interest driven; hence, a theory may be explanatory with respect to one set of aims but not with respect to another. For van Fraassen there is no independent perspective from which a theory might be pronounced explanatorily successful, since an explanation requires a context to fix interests and purposes. Only pragmatic considerations determine what will qualify as explanatory.

Giere argues that such 'auxiliary hypotheses' are precisely what give a theory its empirical content. Whereas, on the one hand, Giere is able to avoid Newman's criticism, on the other he and his supporters require an account of how auxiliary hypotheses (construed as sentences+) can give rise to any degree of significance whatsoever.

Indeed, there is a sense in which Giere is equivocal about 'realism', but a charitable reading suggests that he is in fact attempting to extend the 'goalposts' and perhaps redefine realism. While I am sympathetic to such a proposal—especially given the seemingly intractable task of explicating how sentences+ might be held true and correspondent with conditions of the world in some way independent of knowers—engaging such a discussion will, at this stage, lead matters too far afield.

Given the preceding survey, it should be apparent that no version of the semantic view of theories works to facilitate metaphysical realism solely in virtue of its framework of expression. That is, the varying frameworks within which theories are expressed fail to show how truth conditions might be credited to sentences+ in some manner independent of human judgment. In each case, the referential functions that compose a theory's intended interpretation are fixed via our pre-analytic intuitions about what kinds of things exist and how to ascribe truth to both our observable and theoretical domains; moreover, those intuitions are just assumed to be correct. Van Fraassen's variation on the semantic view manages to avoid some of the ontological challenges faced by both the received view and the set-theoretic approach, but under constructive empiricism we cannot countenance even some of our best theories as true.

In the next chapter I investigate a strategy in a family of arguments advanced by Hilary Putnam known as the 'model-theoretic argument against metaphysical realism'. Putnam shows that upon interpreting the theoretical vocabulary of a theory reconstructed in the Ramsey-Carnap, format, the theory—despite its now providing a, so-called, 'realist' referential scheme for a theoretical vocabulary—will still be susceptible to systematic reinterpretation along the lines Newman laid out in his objection to Russell. Moreover, Putnam attempts to show that no proposed naturalistic constraint can work to solve the Newman problem, since the proposed constraint will always be a candidate for reinterpretation given its membership in the global theoretical domain. My undertaking in the latter portion of the next chapter will be to show why Putnam's 'just more theory' argument fails.

# Chapter 2

"It's all the same, only the names will change."—Bon Jovi

## 2.1 Putnam's Model-Theoretic Argument

Putnam's notorious model-theoretic argument against metaphysical realism (henceforth, MTA) is the meta-logical counterpart to Newman's criticism of Russell. Demopoulos-Friedman (1985: 194) observes that, "Where Putnam argues from the consistency of a set of sentences, Newman argues directly from the existence of a relational structure satisfying the intuitive conditions of a model." The MTA is meant to set the stage for Putnam's Kantianesque 'internal realism', a system where words and thoughts cannot, even in principle, reach out to the external world.

Putnam views metaphysical realism as a constellation of several closely associated philosophical theses about the relations between language and reality, and between truth, knowledge and warranted belief. His depiction of metaphysical realism is captured here:

On this perspective, the world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs [what I've been calling, 'sentences+'.] and external things and sets of things. (1981: 49)

Putnam holds that the impressive success of science certainly warrants the view that our best theories cannot be grossly off-the-mark, but this in no way entails the foregoing version of metaphysical realism. Metaphysical realists hold that some sentences in our best theories are true and that their sub-sentential elements correspond with worldly objects, which is to say metaphysical realists believe that the sub-sentential terms of those true sentences determinately refer. If this is so,

according to Putnam, the onus is on the metaphysical realist to demonstrate how this might be so, and he attempts to show, via the MTA, that the realist cannot deliver on this requirement. As background for the argument Putnam appeals to an important proof of meta-logic—the Löwenheim-Skolem theorem—which merits some discussion.

When Leopold Löwenheim developed his theorem in the early twentieth century it seemed plausible to some mathematicians that mathematical objects, like sets, could be characterized via postulated axioms. The axioms could define the class of intended mathematical objects and would be expressed in what was, at the time, the only well-understood branch of logic: the first-order predicate calculus. At this time Georg Cantor had already proven several famous results, including the theorem which states that the set of all subsets (i.e. the power set) of a given set must be cardinally greater than that set. His proof implies there are infinities of different size, or *cardinality*. Since it is also possible to show that the cardinality of the set of real numbers is equal to the cardinality of the power set of the natural numbers, it follows that the set of all real numbers (i.e. all numbers expressed as following a decimal point) is larger than the set of natural numbers. Löwenheim, and later Thoraf Skolem, used Cantor's result to prove another theorem that seemed, at the time, rather paradoxical.

In Skolem's strengthened version, the 'downward' portion of the theorem shows that for countable, or finite, sets of sentences of standard first-order predicate logic, if there is any model with an infinite domain in which those sentences are true, there is another model in which they are true where the domain of the second model is a subset of the domain of the first model, and the new domain can be mapped one-one onto a set of natural numbers. Moreover, the new model is a 'submodel' of the original—that is, all the terms of the language have the same referent in both models, and for any  $n$ -place formula in the language and  $n$ -tuple of elements of the submodel, the  $n$ -tuple satisfies the formula in one model just in case it does so in the other. For example, take a theory  $T$  in which  $S$  denotes the set of all subsets of the set of natural numbers under interpretation  $I^1$ .  $S^1$ 's cardinality is therefore larger than that of the set of natural numbers. The Löwenheim-Skolem theorem shows that the true sentences of  $T$  can be reinterpreted under  $I^2$  such that  $I^1$  and  $I^2$  might say different things about  $S$ , yet the sentences of  $T$  come out true under both interpretations. Such reinterpretations are sometimes called 'proxy functions' and a systematic application of proxy functions is sometimes called a 'permutation'. Putnam takes the proof to apply to any domain of individuals—in particular, those held to exist by metaphysical realists—and claims to achieve the same paradoxical result.



The generalized model-theoretic version of the argument<sup>1</sup> asks that we consider a theory  $T$  construed in the Carnapian vein (i.e. descriptivist, partially interpreted and expressed in first-order logic). We should assume that  $T$  is an ideal theory; it is therefore:

...complete, consistent, and it predicts correctly (as far as we can tell) all observation sentences to meet whatever "operational constraints" there are [such that it is]... "beautiful", "simple", "plausible", etc. The supposition under consideration is that [ $T$ ] might be all this and still be (in reality) false.<sup>2</sup>

Since its observational consequences come out true we should extend the interpretation of  $T$  to its theoretical vocabulary  $L$ . We do so by assuming that the world,  $W$ , can be parsed into infinitely many pieces and have  $T$  claim there are infinitely many things. *Ex hypothesi*,  $T$  is consistent, so by the 'upward' Löwenheim-Skolem theorem<sup>3</sup> (which states that a model whose submodel is infinite has elementary extensions of arbitrarily large cardinality)  $T$  has a model of every infinite cardinality—that is, the sentences of  $T$  can be true under a different, albeit elementary equivalent, model. Choose such a model,  $M$ , of the same cardinality as  $W$ , and select a map of the individuals of  $M$  onto  $W$  that is in one-one correspondence. We can then use the mapping to define the relations of  $M$  directly in  $W$ . This gives us a satisfaction relation SAT that we can use to define—in Tarskian fashion—a truth predicate for  $T$ : TRUE(SAT). The SAT relation is a function mapping the terms of  $L$  onto sets of pieces of the world, so the 'correspondence' relation between terms and the world assures that  $T$  comes out true of  $W$ , which means any consistent theory is true 'of the world.' However—and this is the kicker— $T$  is true of the world only to the extent that we interpret 'true' as TRUE(SAT).

The problem, as Putnam sees it (and as Newman analogously saw it), is that TRUE(SAT) gives us an entirely *ad hoc* extension that trivializes the truth of  $T$ . The triviality is a direct result of  $T$ 's inability to reject any model meeting SAT, which is to say  $T$  commits us to the implication that  $T$  is true, *but only if*  $T$  is TRUE(SAT). Here we can see how the result mirrors Newman's concern:  $T$  is TRUE(SAT) is true as a matter of logic, not as a matter of empirical fact. TRUE(SAT) fails to distinguish between the intended model, with empirical content, and non-standard, albeit elementary equivalent, models. So the MTA and Newman's objection differ only to the extent that Newman shows that fixing the domain and models up

<sup>1</sup>I will focus primarily on the MTA as presented in Putnam (1981).

<sup>2</sup>Putnam (1976), 485.

<sup>3</sup>The upward Löwenheim-Skolem theorem is a consequence of the compactness theorem, which states that a theory  $T$  has a model if each of its finite subsets has a model.

to *isomorphism* does not fix the intended reference of the *T*-vocabulary<sup>4</sup>, whereas Putnam shows that fixing the domain and models up to *elementary equivalence* does not fix the intended reference of the *T*-vocabulary. Apart from this disparity, the only relevant difference between them lies in the specific purposes for which each theorist deploys his version of the strategy.

Putnam is able to generalize the first-order result—that any consistent theory can be interpreted in such a way as to make all its sentences true—by observing that all one needs to do so is to assume that there are enough things in the world. This, of course, renders his elaborate model-theoretic exposition of the argument superfluous. Nevertheless, according to Putnam, anyone who claims to be asserting the 'capital-T' truth of a particular theory about a theoretical domain accepts more than the fact that it is true according to some intended interpretation. The argument, in effect, challenges the realist assumption that metaphysical constraints provide a unique true interpretation, since a wealth of true interpretations obtain. Putnam then generates examples which he takes to show this (1978: 123-7; 1981: 29-48; 1983b: ix-xii, 1-25; 1989: 213-21). His results affirm that, given any domain of objects and an attendant theoretical or natural language *L* to use in reference, the designated extensions for the predicates of *L* can be permuted consistently without altering the truth value assigned to each sentence at each possible world. Such permutations are taken to show that *any* referential scheme we can give for a language is open to systematic reinterpretation, since each consistent permutation delivers the same truth-valued result at the level of sentences.

Putnam often prefaces his presentation of permutation arguments with a careful discussion of *operational* and *theoretical* constraints on reference. Recall from I.§5. that theoretical constraints are sets of sentences with fixed truth-values; they consist of the axioms that stipulate the meanings of the terms in question. Operational constraints, on the other hand, function in the role of experiential conditions on empirical sentences that tell us what can be measured and observed. Such empirical conditions are not stipulated in order to provide the meanings of those sentences (i.e. the conditions are considered pre-theoretic and intuitive), but the nature of those conditions are held to constrain our judgment about when empirical sentences are true. Therefore, in order to evade charges that his permutations are mere formal operations having no application to the concrete world, Putnam specifies that all theoretical and operational constraints on the reference of a language must be held constant.

The important conclusion Putnam intends to convey by way of his permutation examples is this: keeping theoretical and operational constraints invariant,

---

<sup>4</sup>Demopoulos-Friedman (1985), 194.

if the reference of sub-sentential terms may be varied in a systematic way without a shift in the truth values of entire sentences, then the putative truth conditions prescribed by  $W$  can neither determine the truth of sentences+, nor can they determine the reference of sub-sentential elements. As a corollary, metaphysical realism is shown to be false because the permuted worlds confound, according to Putnam, at least two of the metaphysical realist's core theses. First, and most explicitly, they are counterexamples to the claim that there can be one uniquely true theory/ description of the world. Second, they show that since truth can be made to obtain for sentences in a theory whose interpretation is radically nonstandard, the notion of a correspondence relation between what sentences+ say and the world is horribly obscured.

Permutations enable us to construct possible worlds in which, for example, the term "cat" might refer to dogs (or virtually anything else), while the truth values of the atomic sentences—hence, of all sentences—in those permuted worlds remain equivalent to those of our own. However, in each permuted world the specific relation between a given term and its world-indexed referent differs in some salient way. Cats and dogs, for example, are genetically distinct in ways that biological science has uncovered, but even without a genetic description we intuit that the animals are somehow distinct. So, if "cat" refers to cats in  $W$ , but in  $W^*$  "cat" refers to dogs, it should be obvious that the referential relation between the term "cat" and its corresponding extension—the set of all "cats"—in each world differs significantly. That is to say, the truth conditions for a sentence containing "cat" in each of these worlds will be different. For example, the conditions in  $W$  that satisfy the sentence "the cat is on the mat" are different from those that satisfy the same sentence in  $W^*$ , since in  $W$  "cat" refers to the set of all cats, and in  $W^*$  "cat" refers to, say, the set of all dogs. Yet those disparate conditions are—at the same time—truth conditions for the same sentence, only indexed to different worlds. Recall that for Putnam's metaphysical realist, truth is fixed through a *unique* correspondence between what sentences+ say and the way the world is. However, if systematic permutations of our referential scheme can produce sentences+ whose truth *values* are invariant across worlds, yet whose truth conditions vary from world to world with each instance of the same sentence+, the notion of correspondence is indeed obscured. Unless correspondence between a sentence+ and the world is relativised to a specific interpretation for that world, 'correspondence' holds of the relation between "the cat is on the mat" and every set of truth conditions (i.e. from each permuted world) that satisfy the sentence. Putnam takes this *true-relative-to-a-referential-scheme* result to confute the unique correspondence claim that is, according to Putnam, crucial for metaphysical realism.

In the following section I address some popular criticisms of the MTA by drawing them in general form. By doing so I hope, at the outset, to set aside some general misconceptions that have surfaced as putative grounds for the MTA's dismissal. I will show that these criticisms either miss its point, or apply only to certain expositions of the argument.

## 2.2 Putnam's Paradox

Putnam's MTA has struck many as counterintuitive and others as out-and-out paradoxical; hence, one well-worn critical strategy has been to attack it via *reductio ad absurdum*. Proponents of this strategy (e.g. Glymour: 1982; Lewis: 1984) reply to Putnam with a dilemma along these lines: (1) If we cannot pick-out the elements of the world,  $W$ , we cannot define, describe, or identify any function that assigns extensions to our terms and predicates in descriptions of  $W$ . (2) If, on the other hand, we *can* pick-out the elements of  $W$ , then we can distinguish between right and wrong assignments of extensions to our predicates in  $W$ . Proponents of this strategy usually assume Putnam to be offering the MTA as a counterexample to the antecedent of the conditional premise (2), thus affirming the truth of the antecedent in (1), which one can agree lands us squarely in contradiction. Those same critics take as self-evident the view that we are indeed capable of distinguishing between right and wrong assignments of extensions to our predicates in  $W$ , hence the antecedent of (2) must obtain, which shows that the MTA cannot apply to our language.

To restate somewhat less technically: it is clear we are capable of grasping the distinction between our intended interpretive scheme and non-standard (permuted) reinterpretations. This is why it strikes any competent English speaker as odd when someone attempts to use the term "bicycle" in reference to a jigsaw puzzle (or anything other than a bicycle, for that matter). The ability to distinguish our standard, intended interpretation from non-standard ones provides us the wherewithal for determining reference. That is to say, since we know which one among the scores of possible permutations is the intended interpretation, determining reference is merely a matter of stipulating that *our* standard interpretation is the correct one—namely, the one that squares with the way we actually use language. However, if we were incapable of making such an important distinction, it would be impossible for us to learn a language in the first place, let alone follow the reasoning in Putnam's argument. Hence, the MTA cannot possibly apply to our own language.

There is indeed something appealing about this argument. But this otherwise acceptable line of reasoning misconstrues the intended target of Putnam's denigration. Recall that Putnam takes the truth conditions for sentences to underdetermine the reference of their sub-sentential elements, and he supports this by showing how certain reinterpretations of those sentences leave truth-values invariant. Certainly Putnam relies on our ability to refer successfully—that is, to distinguish between right and wrong assignments of extensions to our predicates in the act of linguistic communication. So we can rightfully charge Putnam with question-begging if his aim is to indiscriminately undermine our *ability* to successfully refer; after all, he clearly presupposes a referential scheme by the very act of assuming his argument to have content. The important point to note, however, is that Putnam is not offering a skeptical thesis about our *ability* to refer successfully—what David Lewis (1983: 226) has called "a purely voluntaristic view of reference". Rather, Putnam means to dismantle a specific brand of metaphysical realism and its en suite theory of truth (as outlined early in §1). Hence, if the alternative Putnam advances (i.e. his internal realism) doesn't succumb to the criticisms he levels against the metaphysical realist (which I do not believe it does), it should be acceptable for him, at the very least, to assume some referential scheme in order to state the argument.

Another notable line of argument surfacing in the literature contra the MTA (e.g., Hacking: 1983, 105) takes aim at its first-order exposition sketched in Putnam's 'Models and Reality' (1977). The criticism, roughly put, is that, given the argument's first-order exposition, and given the Löwenheim-Skolem theorem's limited applicability to first-order languages, the MTA can only apply to first-order languages. Hacking explains:

The Löwenheim-Skolem theorem is about sentences in first-order logic. No one has ever shown that the commonplace language of physicists can ever be squeezed into a first-order format. So the argument is not known to be relevant to, say, quantum electrodynamics, and hence, not to scientific realism (1983: 105).

Strictly speaking, Hacking's criticism holds of the exposition against which he levels the charge, but the criticism must be restricted only to such first-order expositions. In other versions, however, concerns over its limited, first-order scope do not apply. For example, in *Reason, Truth and History* (Putnam, 1981) he asks that we consider permutations of a language whose reinterpretations preserve the truth conditions of the language's *atomic* sentences—that is, its *propositional* matrices. It should be clear to Hacking that the latter version (Putnam (1981)) preserves

the truth values of second-order generalizations (and also their modalizations<sup>5</sup>). So, whereas it seems likely that quantum electrodynamics will never be captured in first order logic, that the MTA doesn't apply to scientific realism does not follow. If, for example, certain legitimate physical theories actually are expressible at second-order (or higher), then the MTA—assuming for the moment that it *actually can* apply to scientific theories so expressed—would certainly pose a problem for scientific realism. If not already obvious, I will (ensuingly) argue that the MTA cannot apply to such theories.

In the next section I examine arguments by other authors who follow a strategy similar to that of Putnam. Their general approach is motivated by related, albeit distinct, worries regarding the determinacy of reference for terms in a given theory. My reasons for highlighting these other arguments are three. First, they clarify the gist of Putnam's strategy less technically; second, they share some assumptions about reference being, to varying degrees, independent and unconstrained; and thirdly, they falter for making the same error as Putnam.

## 2.3 Putnam and the Permutationists

### 2.3.1 Quine

In W.V. Quine's 'Ontological Relativity' he raises proxy functions in a non-mathematical context by rehearsing the "gavagai" thought experiment from his *Word and Object* (1960). To briefly recap: A rabbit scampers by and a native points and utters the word "gavagai". Assuming the role of field linguists, we face the problem of choosing a function that links the term to its referent. According to Quine, we cannot settle the matter by repeatedly querying the expression 'gavagai' for the native's assent or dissent while systematically varying stimulus conditions because, "...a whole rabbit is present when and only when an undetached part of a rabbit is present, also when and only when a temporal stage of a rabbit is present" (1969c: 30). Hence, we can never be certain whether, by 'gavagai', the native refers to the rabbit as, say, a type of species (like we tend to do in our conventional referential practice), or whether the native means to suggest, for example, more nuanced ideas like, "undetached rabbit-parts", "temporal rabbit stage", etc. According to Quine, translators will charitably assume that the native refers to the rabbit in the same manner we do, and therefore assign the 'Junglese' term "gavagai" as coextensive with "rabbit". However, we are just as warranted assigning "gavagai"

---

<sup>5</sup>Hale and Wright (1997) p. 433.

as coextensive with "temporal rabbit stage" or "undetached rabbit-parts"; hence, both meaning and reference are indeterminate.

Quine is, in fact, chiefly concerned with the indeterminacy of meaning/ inscrutability of reference with respect to members of a common linguistic community. He uses the example of radical translation to pump our intuitions that the disparate scenarios are indeed analogous. In spite of the obvious fact that we manage to use our languages successfully, Quine instead emphasizes the hazy individuation conditions for objects in our domain of ontological discourse. However, our ability to use language with great success—scientific language, in particular—should provide ample warrant for the view that the individuation conditions stipulated in our best scientific theories are, by and large, quite good, notwithstanding their defeasibility. Moreover, if human referential practices are in principle inscrutable, yet inscrutability is a matter of degree (which Quine would indeed permit), inscrutability ought then be conceived on a continuum, with, for example, radical, natural language translation at one extreme, and, say, theoretical physics at the other.

We can concede some amount of indeterminacy to Quine because the proxy functions in his own examples are (somewhat) plausible and offer salient alternatives. That is to say, the utterance "gavagai" is indeterminate with respect to 'rabbit', 'rabbit-stages/parts', etc. but not so of, say, 'Bradley tank' or 'moon of Jupiter'—we rule out such peculiar alternatives because they are, quite simply, too outlandish. For reasons related to contextual saliency we do not consider them legitimate referential candidates under any feasibly coherent scheme. Saliency and other use-related contextual considerations function as constraints on our interpretations of utterances; they limit referential inscrutability to the extent that most cases of indeterminacy are soluble and we ultimately achieve a successful translation. Moreover, most of us recognize that more exists than need be existentially quantified in our theories. As Field (1982) is apt to note, "it is rarely to the point to assert the existence of undetached rabbit parts as well as of rabbits." Evidently, a little common sense goes a long way in radical translation.

When Quine discusses proxy functions he has in mind the notion of a "cosmic complement", where predicates that apply to concrete objects can be reinterpreted to reflect different objects that might be assigned as the values of a theory's quantified variables. According to Quine, there is no way to resolve which function provides the genuine article. So, despite our ability to make reliable predictions under a given referential scheme, it is presumptuous to assume there is a uniquely correct referent.

However, when Quine observes that it is possible to permute the entire set-

theoretic universe by exchanging the referential roles played by sets and their cosmic complements without altering anything in our linguistic practice, he ignores the point that Newman stressed: there is indeed an empirical basis for the theoretical structures appearing in our best theories. So, whereas we can permit countless permutations of the abstract set-theoretic universe—where reference between mere terms and abstracta (like sets) is, by definition, unfettered—we cannot permit the same of cosmic complements if there are, in fact, naturalistic constraints that steer our practice of referring to them. Quine (1981) eschews such constraints by permitting radical proxy functions in support of ontological relativity. But the move puzzling in light of Quine's view that the difference between abstract mathematical objects and concrete objects is illusory—that is, there are, as a matter of empirical fact, only concrete objects. Hence, Quine would need to claim that a naturalistic constraint—e.g. causality—plays no role whatsoever in fixing the referential relations between the terms of sentences+ and their cosmic complements. But this is clearly not his view.

Quine, in fact, acknowledges the existence of causal constraints on reference as, for example, "...one's own inborn propensity to find one stimulation qualitatively more akin to a second stimulation than to a third" (1969c: 31). But on the Quinean picture the relevant causal constraints do not go far enough to provide clear individuation conditions for determining the referents of utterances: causal constraints only go so far as to give us likely candidates, like rabbits and rabbit stages. And he is quick to note that, given cases of deferred ostension, inscrutability runs deeper than we want to acknowledge. Deferred ostension occurs when, for example, one indicates the gas tank is full by pointing toward a gasoline gauge. Quine takes his examples of radical translation to amplify the indeterminacy exemplified by cases of deferred ostention. Hence, a native's utterance "gavagai"—rather than meaning 'rabbit'—could just as likely mean 'our main-course this evening', since we cannot be sure of the intended sense by which to interpret the utterance.

But deferred ostension is far from intractable once we step 'outside' isolated cases and locate them in a wider causal network involving the causal histories of a speaker and the things to which she refers. In the preceding example, the *prima facie* inscrutable pointing gesture is resolved by appeal to causal accounts of both how we know what gauges are for and how they work. Making such an external shift works to disambiguate the gesture and get us to the 'full tank'-referent. So, whereas obscurities surrounding radical translation can make the task of specifying a referent very difficult, there is no reason to believe that such cases are in principle insoluble. Understood in this way, referential inscrutability can be viewed as the consequence of other 'social' constraints on the forging of explanations—i.e.



limitations on material resources, technology, individual perseverance, and time.

Indeed, failure to provide a satisfactory causal story at a certain point in time fails to rule-out the possibility of forging one at some future stage of inquiry. Deferred ostension and radical translation certainly pose profound challenges to our giving precise accounts of our referential system, but they fail to pose an in principle obstacle to the forging of an acceptable account of those practices. Rather, they pose precisely the kind of deep scientific challenges whose difficulty motivate further research and creative analysis.<sup>6</sup>

### 2.3.2 *Field*

Hartry Field (1975) employs a kind of permutation argument in an attempt to show the truth of a certain conventionalist thesis about reference. For Field, a scheme of reference for a language is an account that stipulates which words of the language are names, what the names denote, which words of the language are predicates, and what these predicates are true of. The thesis runs roughly as follows. If a given account of reference is adequate, then many such accounts are adequate, and the choice of one account over another is only a matter of convention. Field attempts to show that, given our adequate account of reference, another account can be had by considering a systematic permutation of the physical universe. So, for any reference scheme  $S$  for a language  $L$ , a very different scheme  $S_\varphi$  can be constructed for the universe with a permutation that satisfies the same sentences. The permutation runs along these lines: if  $S$  says a name-symbol refers to  $x$  then  $S_\varphi$  says the same name-symbol refers to  $\varphi x$ ; if  $S$  says a predicate-symbol holds of just those things with property  $Q$  then  $S_\varphi$  says the same predicate-symbol holds of those things  $x$  such that  $Q$  holds of  $\varphi^{-1} x$ . Whereas Putnam's permutations involve operational constraints that are fixed, Field does not stipulate their invariance, so causality (or any similarly conceived naturalistic constraint) is immediately up-for-grabs. Assume, for example, that under our own scheme of reference the utterance of a name is causally connected with  $x$  and not  $\varphi x$ . Field claims that, despite the role causal relations happen to play in  $S$ , in  $S_\varphi$  it would be  $\varphi$ -causality underwriting the reference of names, which amounts to a scheme of  $\varphi$ -reference.

Field arrives at the preceding conclusion because he holds the view that when we appeal to causality in the context of reference it is, in some sense, arbitrary—we have, in his words, "chosen" to employ the notion of causality in order to find the

---

<sup>6</sup>Here I mean to suggest recent work in theoretical neuroscience aimed at giving a causal/ informational account of reference in the context of representational mental content. See, for example, Eliasmith (2000).

"...physical underpinnings of the concept of reference that people actually employ." (*Ibid*: 384) Furthermore, Field claims that we might have arrived at the same (or better) results by having chosen  $\varphi$ -causality over causality, and that we have chosen causality only because an "...*arbitrary decision* is already built into the ordinary usage of 'refer'." (*Ibid*: 379; my emphasis) Moreover, the causal theorist owes us more argumentation in order to ground that causality is superior to  $\varphi$ -causality. The upshot is that, for Field, proper names and natural kind terms cannot be fixed by causality because 'causality' is just another term in the theoretical domain that is susceptible to permutation. This is indeed both a peculiar and radical view, and one echoed by Putnam.

The causal relations we should feel safe to employ in our analyses of reference are the ones successfully applied in sciences like biology, chemistry, physics, as well as those appealed to in 'softer' sciences like economics and psychology. Causal explanations in these sciences generally tell us why some particular outcome or some general pattern of outcomes occurs. A distinguishing feature of such explanations is that they show how what is explained depends on other distinct factors, where the dependence in question has to do with some matter of empirical fact, rather than for logical or conceptual reasons. When we consider scientific explanation in the context of the semantic view of theories, function mappings are made according to a commonsense framework where an array of causal relation-types are pre-theoretically assumed to obtain. These same relation-types are assumed to be conditions that constrain our judgment about when empirical sentences are true—that is, they feature among the 'operational constraints' on theory. Field's view is that one can permute the referential assignment of those pre-theoretic terms that give us purchase on empirical content. The problem is that he wants access to the very constraints that steer our interpretive practices to the effect that commonsense causal terminology like 'touches', 'sees', etc. become ' $\varphi$ -touches', ' $\varphi$ -sees', etc., and this implies that our accounts of biology, chemistry, etc. are likewise conventional.

Convention, construed in the strong sense, pertains to decisions made among members of a community. But in both commonsense and scientific contexts it is far from clear that causal explanation is conventional in this strong sense. Indeed, if it makes sense to say that we can choose a language, the language choice will always be relative to some aim; for example, to communicate, convey specific information, state facts, etc. Hence, relative to the aim for which we want a language, not all language candidates are suitable or on par. Grover Maxwell (1962) has suggested that when it comes to the adoption of a linguistic framework suitable for developing scientific theories, the language should, as a *condition of adequacy*, allow for the

development of *explanations*. Field clearly conceives of theories as linguistic, and he considers the attempt to determine a causal scheme of reference for such theories an attempt to legitimate the relation as a physical one. But, he also contends that the choice to specify the relation between some relata as causal is *arbitrary*. That is to say, he takes our intuitive notion of what constitutes a cause to be a mere matter of convention.

A problem for Field is that we take such causal specifications to convey genuine information about the *kind* of relation into which the designated relata fall, and not merely about the way in which we use the term. That is to say, we take such designations to be indicative of features of the world, and not merely about the way in which we happen to perceive it. Our designation of certain relations into kinds thus elicits considerations surrounding both the context of explanation and the nature of our sensory apparatus.

In Part III I examine the extent to which a causal theory of reference provides a satisfactory account of reference for natural/ theoretical languages. In the following subsection I detour momentarily to introduce the important notion of a causal context of explanation, as well as the view that we are physiologically/ psychologically disposed to construe interpretive contexts in science as causal.

## 2.4 Causal Contexts

The view that causality cannot be jettisoned from certain explanatory contexts is bolstered when we consider a different thought experiment from Field (1982). Consider an alien species that has evolved radically different sensory/ perceptual modalities from our own. As a result, the aliens recognize joints in the world that we do not. Moreover, their ability to predict and explain features of the world is on par with ours, yet they do so without employing concepts like 'cause and effect', 'molecule', 'quark', etc. Clearly the alien is a being who interprets the world using a fundamentally different interpretive scheme than we do—one involving concepts we cannot comprehend due to constraints on our contingent physical constitution.<sup>7</sup> Assuming we can grant the possibility of such a being, does its existence do any work to show that, in certain contexts of explanation, invoking causality is,

---

<sup>7</sup>Field (1982) raises the possibility of such beings in order to challenge Putnam's specific characterization of metaphysical realism; namely, the premise that realists believe there is exactly one true and complete description of the world. There seems to me no problem with my transposing the alien thought-experiment into this context, since Field clearly takes their existence as a genuine possibility.

for us, just a matter of convention and one we might have otherwise passed over? Clearly it does not.

To the contrary, what Field's story implies is this. The concept of causality is an artefact of our physical constitution that arises with the attempt to give a sense to certain phenomenal features of the world. It is unproblematic to claim that the decision involving which languages to adopt in stating a theory might involve elements of convention (e.g. we might have alternatively decided to explain the world in a language associated with supernatural design) but we cannot say the same of causality if the theory is meant to explain some phenomena in a relevantly *causal context*.<sup>8</sup> Causal contexts involve perceived asymmetries between purported causes and their outcomes; that is to say, If *A* causes *B*, then *B* does not cause *A*. To borrow a famous example, consider a flagpole casting its shadow on a sunny day. We can explain the length of the shadow by way of the flagpole's height relative to the sun's position at a given time, but we cannot explain the flagpole's height by way of the shadow's length relative to the sun's position. Indeed, we can *derive* the flagpole's height with the requisite data, but this does not *explain* the flagpole's height in the sense whereby we explained the shadow's length. And this is so because it is obvious and intuitive that the sun and flagpole together *cause* the shadow and not *vice versa*. So, even if Field were to formulate an alternative mode of explanation jettisoning causality in the context of interpretation, it does nothing to shore up the conventionalist view of causality in any strong sense.

Rather, Field's hypothetical extraterrestrials imply the following. Given our physical constitution, we are disposed to represent certain features of the world in certain ways, just as those features we represent are themselves disposed to behave in certain ways. We infer everyday instances of causality due to varying degrees of palpable continuity in our representations of the world; specifically, where those continuities suggest discernable causes and apparent effects relative to our explanatory aims. Causality is thus anchored to both our sensory / perceptual faculties and our specific aims (i.e. explanation). Hence, causality is conventional, but

---

<sup>8</sup>Here I certainly do not mean to suggest that causal contexts constitute the only means by which we do our scientific explaining. There are a host of legitimate non-causal modes of explanation; e.g. explanation via identification. This is exemplified by the discovery that temperature is identical with mean molecular motion. Explanation enters the picture when we consider that increasing the temperature of a gas increases the level of pressure. There are certainly underlying causal processes at play, and there is a sense in which the identity might alternatively be explained in causal terms, but given the context of explanation the 'heavy lifting' is done by the less cumbersome and symmetrical identity between temperature and mean molecular motion. In addition, it should be noted that in explanation we also employ things like models and analogies which, likewise, are not straightforwardly causal.

only in the weak sense that had certain features of the world been different, those differences might have precluded any *knowers* from using the concept. However, causes are epistemically inherent to creatures like us whose explanations assume the world to consist of spatiotemporal, property-bearing objects that fall into relations.

In the section below I discuss Putnam's notorious argument that, like any other concept-term in our language, 'cause' is "just more theory" and hence susceptible to reinterpretation. We have seen this view implicit in Quine's arguments for referential inscrutability from proxy functions, and overt in Field's conventionalist view of reference. I will survey and critique some of the objections made by other authors to Putnam's 'just more theory' argument in order to both clarify where they err, and to marshal support for the view I offer and defend in Part III.

## 2.5 The "Just More Theory" Gambit

The 'just more theory' argument surfaced as a rejoinder to criticism from proponents of causal theories of reference (e.g. Devitt, 1980), where Putnam claimed that "cause" is just another term in our domain of discourse and thus susceptible to reinterpretation. In later works Putnam augments his argument to cover any supposed natural constraint on reference. He sketches the matter as follows:

Suppose there is a possible naturalistic or physicalistic *definition* of reference...Suppose

(1)  $x$  refers to  $y$  if and only if  $x$  bears  $R$  to  $y$

is true, where  $R$  is a relation definable in natural science vocabulary without using any semantical notions (i.e. without using 'refers' or any other words which would make the definition immediately circular). If (1) is true and empirically verifiable, then (1) is a sentence which is itself true even on the theory that reference is fixed as far as (and only as far as) it is determined by operational plus theoretical constraints...

If reference is only determined by operational and theoretical constraints, however, then the reference of ' $x$  bears  $R$  to  $y$ ' is itself indeterminate, and so knowing that (1) is true will not help. (1981: 45-6)

Putnam believes that knowing all instances of (1) are true does not answer the problems raised by the MTA because the argument shows that in every permuted world the truth-value of (1) is true. And this is particularly troublesome when,

after permutation, the logical role played by  $R$  in (1) is supplanted by an entirely different relation (e.g.  $x^2$  refers<sup>2</sup> to  $y^2$  if and only if  $x^2$  bears  $R^2$  to  $y^2$ ). Moreover, there will be just as many dissimilar relations  $R^1 \dots R^n - 1$  as there are permutations to the universe of discourse. Hence, according to Putnam, it is illusory to think a special reference relation has been defined because  $R$ 's explanatory power is equivalent to  $R^2$ 's.

Putnam's claim that any available (indeed, any possible) naturalistic, reductive account of reference amounts to 'just more theory' and is thus subject to systematic reinterpretation is the issue of primary concern here—in fact, it is the deal-breaker. To allow the claim would be to provide Putnam with shelter against *any* constraint on reference whatsoever, provided it features in the theoretical domain of discourse. Fortunately, we need not allow the claim for reasons made clear by David Lewis (1984: 225).

Constraint  $C$  is to be imposed by accepting  $C$ -theory, according to Putnam. But  $C$ -theory is just more theory, more grist for the mill, and more theory will go the way of all theory... $C$  is not to be imposed just by accepting  $C$ -theory. That is a misunderstanding of what  $C$  is. The constraint is not that an intended interpretation must somehow make our account of  $C$  come true. The constraint is that an intended interpretation must conform to  $C$  itself.

Lewis' point is that there is a crucial distinction between, on the one hand, modeling a proposed constraint via some intended interpretation—i.e. making a statement about the constraint come out true—and on the other, the interpretation *itself* conforming to the very constraint that appears in the theory. The move—an *externalist* shift—sees the problem as a consequence of the internalist point of view from which Putnam argues, where each new theoretical term is co-opted by global theory. Lewis' idea is that the constraint is not something we impose on ourselves, from within; rather, it is imposed by what we find in the world, where we, as knowers, feature among the things we find.

The point also highlights, albeit indirectly, how theories expressed in the 'semantic' format are, in a sense, insulated from Putnam's maneuver. The semantic view of theories does not address the question of how an interpretation is fixed: the intended interpretation is *already assumed to be correct*. As Demopoulos (2003: 392) observes: "[The semantic view] presupposes the pragmatic background within which the models that *are* the theory are given and within which the semantic view is expressed." So, whereas a theory stated in the Carnap-Ramsey format is susceptible to permutation—since the empirical ground for its interpretation has

been reduced to a set of formal sentences—the model-theoretic format anchors terms to the world via an interpretation that is both governed by the constraints in question, and *ex hypothesi* assumed to be correct. So, to accept natural constraints on reference is to accept that the constraint in question is, in some important sense, *theory transcendent*, since it forms the basis upon which theory can develop in the first place. In Part III I will examine this notion of theory transcendence in greater detail.

Putnam's 'just more theory' manoeuvre indicates that he has overlooked Lewis' key observation.

"Causal connection" is attached to *R* by causal connection, not by metaphysical glue, they write. But this is, in fact, just to say that *R* (causal connection) is self-identifying. This is to repeat the claim that a relation can at one and the same time be a physical relation and have the dignity (the built-in intentionality, in other words) of choosing its own name. (Putnam, 1984: 7)

I certainly do not mean to suggest that naturalistic constraints have "built-in intentionality", so Putnam's rejoinder suggests (to me, anyway) that he has overlooked the far-reaching consequences of his own stricture; namely, that of holding operational constraints invariant. Recall that the pre-analytic and pragmatic basis for an intended interpretation involves, among a host of other considerations, the existential assent of an array of metaphysical categories; namely, objects, properties and relations, all of which feature in the operationalization of theory. But Putnam's stipulation that we hold operational constraints invariant entails invariance of these pre-analytic notions (and their corresponding terms) *prior to* permuting reference. However, Putnam's methodological restriction provides sufficient constraint to preclude radical reinterpretation of theory.

The invariance of operational constraints so severely limits Putnam's legitimate permutative options that it is difficult to imagine a permutation that could produce importantly different interpretive results from the scheme we presently employ. Indeed, we still face the Quinean problem of referential inscrutability, where the proposed constraints only limit us to salient referential alternatives. But this sort of tame inscrutability is, for the most part, an inexorable consequence of having to transmit the content of thought using vague language. Nevertheless, we still manage to communicate and successfully interpret the content of our thoughts and theories using language. Moreover, the inscrutability brought about by deferred ostension poses no especially stubborn problems once we avail ourselves of the relevant causal/ explanatory resources (in addition to ones from language

pragmatics and commonsense). The burden therefore falls on Putnam to model an alternative referential scheme that shows his desired result while *actually* holding operational constraints constant.

Azzouni (2000) has also observed that Putnam's notion of operational constraints is ill-conceived and, ultimately, responsible for his arriving at the paradoxical view that naturalistic constraints are themselves subject to reinterpretation. Specifically, he observes that Putnam runs together the notions of 'operational constraint' and 'observational constraint', supporting the charge by pointing to cases (in specific papers) where Putnam uses the phrases interchangeably without drawing any distinction between the two. Azzouni takes the equivocation to indicate that, for Putnam, operational constraints in fact *do* stop at the observational level. If this is indeed Putnam's view it is pretty obviously false. Consider, for example, a practicing physicist who causally interacts with particles using devices designed specifically for such purposes.

At the proximate end of the causal chain, "observational" predicates apply, but at the other end they don't. Operational constraints, on the causal picture, cut deep into the world—they don't stop only at what we "observe," and, consequently, to honor their restraints we must accept that they fix more than just appearances. (Azzouni, 2000: 166)

On this picture, 'operational constraints' subsume causes that run so deep as to carve nature at joints that elude our straightforwardly observational vocabulary. This picture is consonant with results in modern physics, where technology allows us to track causal interactions with unobservables and collect empirical support for inferences about their microstructures.

In the following, closing chapter I examine the prospect for an externalist account of reference for natural kind terms. In addition, I examine reference in the context of naturalism and discuss features of the ontological picture inherent to the externalist picture of the world.



# Chapter 3

"Mine's not an idle cause." - Shakespeare

## 3.1 Externalism and Internalism

The causal, or *external*, approach to semantic content has been influential in terms of providing a plausible account of the mechanism of reference for proper names and natural kinds. That is to say, it accords with some basic intuitions about the way the reference relation seems to work in natural/ theoretical languages. Saul Kripke's (1980) landmark case for a causal theory of reference for names and natural kinds evolved from perceived inadequacies afflicting the description theory of reference. Kripke points to specific cases that seem to be intuitive examples of successful reference, then purports to show that we do not have the conceptual resources to refer in a way that accords with the traditional descriptivist account. In other cases he argues that the descriptive analysis implies the intuitively wrong result—that the thing we are referring to via description is, in fact, something other than the thing we initially seemed to refer to. Rather than focus on Kripke's specific arguments, more important for my purposes are the considerations that motivated his external shift in strategy.

The contrast between Kripke's externalist move and that of *internalists*, or descriptivists, involves important disparities in their respective philosophical projects. They disagree over which assumptions are problematic, and over how to best characterize the relevant philosophical problems facing a theory of reference. Recall that descriptivists like Russell and Carnap begin with what they take to be the most basic constituents of thought: our phenomenal experience. They consider our phenomenal knowledge to be both fundamental and unproblematic, and they each provide (albeit, problematic) accounts of how we might move beyond mere phenomena to form a conception of an external world. Stalnaker (2007)

has aptly characterized this approach as an attempt to explain our representational resources for a wider domain in terms of given resources for a narrower domain. When viewed in this way we can see how such an approach could give rise to a framework consisting in an observation/ theory dichotomy, and how an answer to the quandary might be assumed to reside in a clarification of the logical and semantic relations between the two.

Externalists, on the other hand, hold that, given the internalist's starting point, they cannot hope to achieve satisfying answers to skeptical challenges—that is, they cannot evade what we earlier saw Hempel call the 'theoretician's dilemma'. Thinkers and speakers cannot, according to the externalist, have the perfect and complete acquaintance with properties and relations that is required by the internalist to grasp propositions expressed in a descriptive analysis. Externalists therefore allege that further reduction would be required of those phenomenal properties and relations. To escape the skeptical dilemma the externalist proposes that we start with the world in which we reside and use the resources offered by commonsense and our best theories to tell us about it. That is, we do not begin with 'sense data' or 'percepts' and attempt to bridge the divide between observation and theory. Instead, the externalist emphasizes our *objective* conception of the world: the commonsense view of a world consisting of objects, properties and events that fall into relations.

Externalists are motivated by the problem of how to reconcile this commonsense picture with the evident fact that the world contains knowers that operate from a purely *subjective* perspective. At the hub of the externalist project is the view that general terms, along with names and other singular referring expressions, depend for their semantic values on environmental conditions. This view owes primarily to arguments advanced by Putnam (1979b) and Burge (1979), and has come to dominate the overall philosophical debate over the nature of mental content in recent years. Externalists hold that if it is indeed the case that certain terms depend for their semantic values on external conditions, our intentional relations to their referents cannot have the kind of foundational status that the internalist project requires.

The debate between these contrasting viewpoints is characterized by their charging one another with unjustifiably assuming something that requires further explanation. The internalist is skeptical of the commonsense picture sketched by the externalist because it begins with entities, events, properties and relations to which we are not, on their view, directly acquainted. They take any assumption involving the theory-independent existence of such things to be either an unwarranted slide or an article of metaphysical faith. Externalists, on the other hand,

see the internalist's rejection of the commonsense ontological picture as perverse, given its intuitive force and—they argue—its centrality to our scientific picture of the world. Chiefly, they see internalists as having at the core of their project an uncritical acceptance and a flawed conception of our knowledge of what we experience. Externalists argue that internalists take for granted phenomena—e.g. sense data, percepts, etc.—which themselves demand explanation. Moreover, externalists argue that the best way to investigate such phenomena is under the aegis of science, which already assumes the externalist picture.

My sympathies—if not already obvious—are with the externalist, so my aim in the remainder of this concluding chapter is to argue for an externalist approach to the problem of reference/ interpretation. However, I conclude that, whereas the externalist project obviates the skeptical challenges that beleaguer internalist/ descriptivist accounts, reference remains elusive even on the externalist picture.

### 3.2 Natural Kinds and the Causal Theory of Reference

Natural kinds (and/or natural properties) are sometimes touted as the 'building blocks' of the sciences. To assert the existence of natural kinds is to accept a metaphysical thesis about the way the world is structured—specifically, that it has a distinct and mind-independent natural kind structure, where the world consists of objects, events and relations that are structured in such a way as to make those structures discernable to us.

Consider a sample of something we can genuinely claim to causally interact with; say, an element, like phosphorus, or a biological kind (exemplified by two creatures capable of producing offspring together.<sup>1</sup>) Alternatively, a sample might consist of events; say, an egg-hatching or a volcanic eruption. The first two samples are each presumed to instantiate a corresponding class of object, or entity. The latter two samples involve events, where a cluster of conditions give rise to an event-kind. We also refer to kinds that are the *causes* of events that we use to identify the kind in question, such as when we say someone who exhibits a distinctive cluster of symptoms has 'the flu'.

Causal interaction with samples has enabled for the rigorous systematization of kinds in the natural sciences, and classifying samples has made it possible for us to make ampliative and reliable generalizations about the way they will behave under certain conditions. Such is the process from which scientific laws are born.

---

<sup>1</sup>The obvious exceptions include those hermaphroditic biological kinds (e.g. worms, plants, etc.) that reproduce by themselves, as well as bacterium, etc.

But in order to state scientific laws we require a predicate that holds of everything that is *the same* as the samples we have on index. Hence, a problem prior to the demarcation of kinds involves determining what it means for one sample to be *the same as* another.

To determine sameness we might, for example, enumerate a list of properties exemplified by our samples in-hand and use a conjunction of those properties as necessary and sufficient conditions for picking out a kind's extension. However, it may be the case that the properties appealed to are neither necessary nor sufficient to determine a kind. Whereas it may *prima facie* seem that two samples are the same at a certain stage of investigation, it may later be revealed that they are, in fact, very different. In the case of scientific classification, such problems are handled via stipulation, where the best available methods for investigating a subject matter are relied upon by experts, who then tell us, to the best of their reckoning, what the requisite conditions for membership in a kind are. Causal theorists usually stress that terms denoting natural kinds should be taken to reference classes of property-bearing objects and events, as opposed to mere clusters of properties, which we then associate with specific samples. That is to say, such terms should be understood as referring directly to the 'real' members of a kind.

When we encounter a novel sample and assign it a name, the process is sometimes called a dubbing, or baptismal, ceremony; sometimes it is called reference-fixing. The process subsequent to reference fixing is reference borrowing, where ensuing speakers—ones often ignorant of the details surrounding the kind's dubbing—refer to kinds by the name assigned at its dubbing. The idea is that underlying a borrower's use of a term is a causal chain of borrowings extending back to the events surrounding the kind's baptism. That is to say, we borrow the term from others through linguistic communication. Putnam (1979b), who has developed this idea in great detail, analyzes the meaning of a natural kind term into four components: a syntactic marker, a semantic marker, a stereotype, and an extension. On this picture, a term like 'tiger' has as a syntactic marker, "noun"; as a semantic marker, "animal"; as a stereotype "large, furry, orange and black-striped cat"; and as an extension, the set of creatures that have the same internal 'tiger-structure' as specified by experts. Hence, Putnam takes the meaning for natural kind terms to obtain in a "division of linguistic labor" (1979b), where the borrowing process mainly involves transfer of a kind's stereotype, and where the demarcation and extension of a kind is specified by the relevant experts.

Perhaps the most serious challenge to the causal theory of reference involves the supposed problem of reference change. Gareth Evans (1982) famously points to the island of Madagascar as providing an example of such change. The island is

located off the southeastern coast of Africa, but was mistakenly named 'Madagascar' by the explorer Marco Polo. 'Madagascar' was originally the name given to an area on the mainland coast of Africa. However, Polo misconstrued the name's referent and his error was able to flourish due to subsequent reference-borrowings by readers of the maps and navigational charts he produced based on his travels. The problem is that Polo did not intend to introduce a novel use of the name. Rather, he believed himself to use 'Madagascar' in accord with the intentions of the person from whom he borrowed the name. However, his error constitutes a genuine case of reference change because, despite our knowledge of the mistake, today nobody is taken to be promulgating error when they refer to the island as 'Madagascar'.

A causal theory of reference handles the preceding problem by recognizing that a name is typically grounded in its bearer via successive perceptual confrontations. That is, despite our knowledge that reference change occurred due to an error in reference borrowing, we should take Polo's initial error to be semantically significant and thus view it as a kind of revisionary dubbing ceremony. The idea is that, given a sufficient number of erroneous borrowings over a sufficient period of time, reference change is a possible, even likely, outcome. Hence, reference for 'Madagascar' shifted from the mainland to the island once perceptually-based groundings in the island (via borrowing) were established.

Reference change for natural kinds poses a different problem for a causal theory. In science, reference change usually indicates errors made in the naming process of a kind, and usually results in the kind's abandonment. New technology and novel discovery sometimes show that what we once took to belong to a kind in fact belongs to another kind, or that we were altogether wrong about the requisite conditions for a sample's membership in a specific class. For example, once technology enabled us to analyze the structural composition of rock we learned that the term 'jade' had been used in reference to two distinct kinds of silicate rock sharing an uncanny ostensible resemblance. The error therefore occurred at the purported kind's dubbing, prior to any borrowing, where it went unrecognized that the rocks contained distinct aggregations of silicate minerals. Hence, the subsequent dubbing of 'jadeite' and 'nephrite' constitute baptismal ceremonies for new kinds, while jade is no longer taken to specify a genuine natural kind, despite sustained and pervasive use of its stereotype.

Examples like the foregoing motivated Devitt and Sterelny (1987: 72) to argue that the reference of natural kind terms must involve a structural component. On this view, the scope of a kind-term is taken to range over all and only those objects that have the same *internal structure* as an available sample. The causal theory, it is held, offers an account of causal interaction with samples where ostensible

properties form the basis for reference. Yet, according to the foregoing view an account appealing solely to ostensible properties cannot be sufficiently robust to demarcate the scope of a kind. Relevant similarities in the ostensible behaviour of samples give rise to intuitions about what *might* constitute a kind, but their designation as falling under a specific kind will ultimately involve an appeal to internal, structural similarities uncovered and/or inferred by rigorous scientific investigation. To borrow a well-worn example, ice and water are the same kind not due to ostensible properties they share; rather, despite contrasting appearances, their being members of the same kind owes to their shared internal structure.

The problem this poses for the externalist is considerable: a comprehensive causal theory for natural kind terms must provide a distinctively causal account of reference for *all* theoretical terms. This is so because, if the preceding view is correct, the reference for a natural kind in science is conceived as a function from its concept and expressed using other theoretical terms to specify its internal structure. Certain samples belong to the extension of a kind-term because they are judged the same as the samples giving rise to the kind-term's baptism; and sameness of internal structure in a range of samples is what shows the extension of a kind-term to be non-arbitrary. So, if Devitt and Sterelny are correct, the burden of reference for natural kinds is shared by both an explanatory account delivered in terms of causal interaction with samples during reference-fixing, and a conceptual clarification of the term using extant observational/ theoretical language.

Indeed, the challenges facing the external approach are substantial, but the point an externalist wants to emphasize is that, even if one were to concede that singular reference to natural kinds in science is conceptually direct, it does not follow that there is no causal/explanatory story to be told about what it is in virtue of which a theoretical term refers. That is to say, it would be a mistake to confuse an account of *how* terms come to be used in reference with a conceptual analysis of *what* that term expresses. Clearly the 'theoretical environment'<sup>2</sup> in which a new kind-term is subsequently embedded bears some of the burden for specifying its reference in the actual practice of science. The important point for the externalist is that, despite our ability to specify reference in science via other theoretical terms, our use of each theoretical term must be underwritten by some causal chain of events, whether accounted for or not. Internalists, however, take the reference relation to be exhausted by an account of the logical and semantic relations between observation and theoretical languages, and the outcome is a globalist picture of language/theory, where each new theoretical term is just another element in our domain of theoretical discourse.

---

<sup>2</sup>The phrase 'theoretical environment' owes to Psillos (1999): 285.

### 3.3 Naturalism and Constraints on Reference

Advocates of permutation arguments—usually internalist/descriptivists—attempt to show that constraints on the interpretation of language are weak due to the apparent global character of language/theory. Global descriptivism, on David Lewis' (1984) rendering, runs roughly as follows: The intended interpretation of a theory's language is the one that assigns things, etc. in some part of reality to the terms of the language such that the theory comes out true of that part of reality. Under global descriptivism each new term is part of the global theoretical language, and this, according to Putnam et al, renders theoretical terms susceptible to systematic reinterpretation. Conversely, semantic externalists take the commonsense realist interpretive framework—the one that comfortably assumes the objective existence of objects, events, properties and causal relations—to preclude the possibility of radical referential reinterpretation.

Some semantic externalists take the priority of the commonsense framework to imply a kind of philosophical naturalism. Naturalism comes in various stripes, and involves a variety of theses, but central to it is an avowal of the existence of natural kinds/ properties. Lewis (1984), for example, rejects global descriptivism for a metaphysical naturalism that purportedly answers the question of how newly introduced terms are to be understood in a growing language—that is, how to get more reference when we have some already. Lewis calls his view of language 'local descriptivism': new theories give rise to new terms, so we should understand such new terms as referring to things that are as described in the old terms of that new theory so as to make the new theory come out true (if possible). For Lewis, theoretical terms refer only if there is a *unique* realization of a theory's Ramsey sentence, so there cannot be multiple realizations of entities that satisfy the observational consequences  $O$  of a theory. 'Uniqueness', on the Lewisian view, should be understood as an *exclusive* extensional relation  $ER$  that structures the relevant domain such that the theoretical terms existentially quantified in a theory's Ramsey sentence stand in  $ER$  to  $O$ . The uniqueness requirement mirrors Carnap's method for showing how reference can be made to theoretical entities, but Lewis departs from Carnap by extending genuine metaphysical status to those existentially quantified classes. Hence, the Lewisian uniqueness requirement translates as a requirement that the domain is already carved into genuine ontological kinds.

Recall that, for Lewis, there can be no answer to the problem of interpretation if we regard all theoretical terms as equally eligible for providing content in a sentence+. But to regard virtually any proxy function as equally eligible—as is held to be the case by Putnam *et al*—is to ignore apparent constraints on our

referential practices. Lewis continues:

Only if we have an independent, objective distinction among properties, and we impose the presumption in favor of eligible content a priori as a constitutive constraint, does the problem of interpretation have any solution at all. If so, then any correct solution must automatically respect the presumption. (Lewis, 1983: 55)

In other words, in order to avoid falling prey to permutation arguments we must assume that only *certain eligible sets of kinds* constitute the extensions for predicates denoting our theoretical/ natural kind terms.

However, Lewis' solution is bound to face criticism from certain other naturalists because his "a priori constitutive constraint" is really just a stipulation, since only *after* scrutinizing our referential practices is it evident they are governed by constraints. An ontological naturalist like Lewis—one who holds that things that exist must exert some *causal* influence in the concrete world—should cry "foul" over the mere presumption of eligible kinds. Realism entails that the eligible sets exist, and that they do so independently of knowers, but for Lewis to stipulate a natural distinction among properties/ kinds without offering a positive causal account of our ability to access them violates the strictures of his own naturalism. In other words, Lewis needs to explain—in causal terminology—why we should respect the presumption in the first place, or be willing to face charges of occultism. Putnam (1981), for example, pejoratively characterizes the strategy of postulating objective joints as one of conjuring a "magical theory of reference".

The problem for Lewis is that, if one asserts that natural constraints govern our referential practices, we should require a naturalistic account of those practices that explains our ability to access the eligible sets and the relations they fall into. That is to say, Lewis needs to provide a positive causal theory of how we come to know the kinds (or at least appeal to an established one) that warrants postulating the ones he does. In view of this, it should come as no surprise that *causal-informational-representational*<sup>3</sup> theories are in vogue among theorists attempting to naturalize mental content. However, despite his failure to honour the strictures of his own naturalism, we should acknowledge Lewis' observation that natural constraints indeed govern our interpretive practices. In II. §4 & 5. I offered arguments in favour of the view that permutation arguments which aim to deflate the determinacy of reference arise from a failure to acknowledge the genuine primacy of such natural/ causal constraints on reference. I then argued that such constraints

---

<sup>3</sup>Examples of causal theories can be found in Dretske (1981, 1988, 1997) and Cummins (1996); and for a two-factor approach in Eliasmith (2000).



indeed obtain and suggested that this is evident where causality transcends theory. In the next section I digress somewhat to introduce an ontological framework which, on my view, helps to illuminate the sense of such theory transcendence.

### 3.4 Internal and External Questions

In his famous paper "Empiricism, Semantics, and Ontology", Carnap (1956) outlines a complex set of distinctions that separate ontological questions as either 'internal' to a language or 'external' of a language. Since its introduction Carnap's view has been regarded by many as flawed, and this is due, for the most part, to criticisms leveled by Quine (1976). Despite the extensive opposition to Carnap's view, I nevertheless concur with the likes of Michael Friedman (1999), Graham Bird (1995) and David Chalmers (2006), that a careful examination of Carnap's position reveals deeper distinctions that, when acknowledged, deflect Quine's criticisms. Assuming the Carnapian framework<sup>4</sup> I intend to show how the notion of causality is indeed theory-transcendent.

Bird (1995) observes and emphasizes that Carnap's internal/external divide actually consists of a four-fold distinction of types of ontological questions. Among internal questions Carnap separates 'particular' from 'general' questions. The idea is that if we develop a language replete with syntax, vocabulary, rules, and test procedures, we can then, in principle, settle any questions of the internal-particular type. Carnap provides examples from what he calls the 'thing language' for physical objects, and from the 'number language' associated with basic arithmetical operations. The former constitutes an 'empirical' system, and the latter, a 'logical' system. When such languages are available, answers to questions from the thing language such as, "Is there a stapler on my desk?" can be resolved. Similarly, we can answer questions from the number language like, "How many prime numbers less than 10 are there?". Both languages provide an apparatus with which to answer such questions—the former by way of empirical *procedures* and the latter via *a priori* reasoning.

General-internal questions, by contrast, are exemplified by such broad questions as "Do physical objects (in general) exist?", and "Do numbers (in general) exist?". Carnap makes two observations about such questions. First, as internal questions they can be answered by appeal to the answers of certain particular-internal questions. So, if there is indeed a stapler on my desk, and staplers con-

---

<sup>4</sup>In particular, to the one as appears in the revised version of "Empiricism, Semantics and Ontology" appearing in Carnap (1956).

stitute physical objects according to the given linguistic system, then it follows logically that there exist physical objects; *mutatis mutandis* for the general-internal question about numbers. Hence, general-internal questions pose no more of a problem than their particular counterparts. The idea is that such questions *presuppose* the internal apparatus of either the thing-language or the number-language.

The second of Carnap's observations about general-internal questions is this. When philosophers raise questions like "Do numbers exist?", they tend not to do so in the sense characteristic of the general-internal question. Rather, they ask the question in a sense independent of the internal apparatus. So, whereas general-internal questions might be restated as, "Does X feature in affirmative existence claims in sentences of the linguistic system?", the sense in which such questions usually arise, according to Carnap, is more akin to, "Are there principled reasons for X's inclusion in the linguistic system (or, more generally, are there principled reasons to accept the linguistic system as a whole)?" Carnap calls such questions 'external', and he points out that, despite the potential for such sentences to share the same linguistic form as certain general-internal questions, the content of the question when conceived externally is importantly distinct from its content conceived internally.

Carnap takes a further step, separating external questions into two categories: first, 'practical' and second, 'theoretical'. Practical-external questions can be formulated as follows: "Are there good reasons to adopt the X-language?". Carnap takes such practical-external questions to be legitimate ones arising 'outside' the linguistic system, with the system (or specific aspects of it) featuring as their subject matter. These are the appropriate sorts of questions to examine *prior* to the adoption of a language by assessing and weighing the benefits of acceptance against the disadvantages of non-acceptance. That is to say, answering practical-external questions involves making practical decisions, and the process of doing so, according to Carnap, should not be taken to involve notions like 'truth' and 'evidence', which only apply to questions that arise internal to a linguistic system.

The final quadrant of Carnap's ontological mosaic consists of theoretical-external questions: for example, "Are universals objective features of the world?" Carnap considers the other three types of question— particular-internal, general-internal, and practical-external—to be the intelligible species of existence question; but he maintains that, in spite of its notable intellectual heritage, there is no adequate interpretation for this fourth type of question. For Carnap, the requirement for identifying a legitimate ontological question is that it have a clear interpretation. The theoretical-external form, according to Carnap, has no such interpretation—it has nothing to which one might appeal in order to prove a determinate

answer. Chalmers (2006) has characterized such theoretical-external questions as, in general, asking whether there are objective, theory-independent proofs for the basic questions of what exists. Consider that, with respect to particular-internal questions, we can turn to either the system's empirical or logical apparatus to provide answers. General-internal questions can be answered by appeal to the answers of specific particular-internal questions; and practical-external questions are decided via a pragmatic cost-benefit analysis. Theoretical-external questions cannot, according to Carnap, be given a sense in any of the preceding ways. Bird (1995; 44) has remarked that such questions are, to some extent, *defined* by their exclusion from the group of interpretable, ontological question-types.

The preceding Carnapian framework for ontological questions is, for the most part, a suitable guide by which to assess ontological questions. Indeed, there have been important objections to Carnap's scheme—ones I cannot devote space in this essay to—but Friedman, Bird and Chalmers persuasively defend Carnap's analysis, showing its immunity to the popular criticisms. My reconsideration of his model attempts to clarify the primacy of our commonsense conceptual scheme, emphasizing its priority to theory. In other words, I exploit Carnap's system to show how one might decide answers to ontological questions regarding the more general elements of this scheme.

### 3.5 The Ontological Status of Causality

Philosophers like Putnam (1983a); van Fraassen (1980; 112-16), Field (1975) and many others, have objected to the assignment of any special ontological status to causality. This deflationary stance harkens Hume's (2000) classic discussion of causality, where he famously argued that we cannot observe causes at work but, rather, theoretically impose them onto situations we observe. For Hume, the problem that causality poses is a skeptical one: how might one justify the inference that causes exist as part of the external world, and do so from one's own evidence, when the evidence to which one can appeal is limited to the contents of one's immediate sense experience, where causes simply do not feature? Hume's *solution* to the problem—a forerunner of Kripke's strategy and one largely ignored by Humean skeptics—involves making a shift that assumes the individuals in this skeptical quandary to be constituents of the world who make inferences about it, and to ask how they do so, why they infer as they do, and why they are as successful as they

are.<sup>5</sup> Internalists, however, see any such shift as rendering the externalist's causes as both interest and context-sensitive. Hence, internalists generally take causality to be merely conventional.

In II. 4. & 5. I conceded that causality is conventional but only weakly so, in that a successful natural science without causality might have been possible had characteristics of our physical constitution been different. But causality, I argued, should not be taken as conventional in the strong sense—that is, as being grounded by agreement in society. Our *knowledge* of causes is indeed dependent upon our contingent neurophysiology, but this in no way precludes them from having genuine ontological standing. Just as one can benignly assert that objects like mountains and rivers exist, one can likewise assert that certain events exist.

Consider that, just because an explanation *E* specifies event-*C* as a cause, and *C* happens to be context and interest-sensitive to *E*, it does not mean that *C*'s place in the broader complex of possible causal agents is strictly context and interest-sensitive. Despite the epistemic and pragmatic importance of *C* to *E*, our reasons for singling out *C* as just one aspect in a larger causal complex has no bearing on *C*'s ontological status. It certainly does not follow from this that *C* does not exist because any theory in which *C* features will already assume that it does. Of course, it may be the case that *C* does not *really* exist, but *C*'s non-existence does not follow from the fact that with other interests and in another context one might not be inclined to specify *C* as salient to, say, *E*\*. Hence, the question of *C*'s existential status is orthogonal to whether *C* in fact plays the role it is specified to play in *E*. To put the matter somewhat differently, if one is at all willing to grant ontological status to events, then an appeal to *C* in an explanation (*qua* cause) should be understood as a specification of that event's genuine influence in bringing about a specific outcome.

Critics of the causal theory of reference sometimes claim that many of the intuitive, commonsense causal relations appealed to by externalists are scientifically illegitimate. Specifically, in order for causality to underwrite reference one should be able to supply necessary and sufficient conditions for acts of reference in 'scientifically respectable' terms. But how is this requirement anything more than an echo of what the internalist takes to be an acceptable answer to the problem of reference? Indeed, the internalist critic is only attempting to reassert the terms of

---

<sup>5</sup>Credit for this observation goes to Stalnaker (2007), who in his recent John Locke Lectures points to the analogy between Hume's solution and Kripke's shift. Stalnaker remarks that, whereas Hume raises this as a possible escape from the skeptical dilemma, he indeed takes the skeptical problem more-seriously than most externalists are inclined to, and also expresses some reticence about employing it.

debate. The externalist, on the other hand, sees no chance of escaping the skeptical quandary on the internalist's terms and instead offers an account of the conceptual resources we use to form beliefs—an account that assumes causality and deploys it to form an explanation of how we acquire and use those resources. In addition, the externalist emphasizes our obvious capacity for finding our way about reliably in our environment and offers an explanation for that capacity. Hence, externalists use the very methods they are assessing to arrive at the conclusion that the world is conducive to the success of those methods, and this is taken by externalists to be the only way to escape the skeptical quandary. The internalist's rejection of this strategy amounts to a rejection of Hume's answer to the skeptical problem without offering a comparable, or satisfying, 'exit strategy'.

Azzouni (2000) has suggested that the internalist's objection can be answered by noting that the causal relations appealed to by externalists are often complex *gross regularities*<sup>6</sup> that are epistemically independent of the natural sciences and which naturally admit of exceptions. Externalists will usually agree there is no unproblematic definition of 'refers' in scientific causal terminology, so they take the term to be a kind of shorthand for the mechanism underwriting our conception of how to use a given natural kind term. Some have expressed this by claiming that the commonsense causes appealed to by the externalist supervene on 'scientifically respectable' microcausal events. But even proponents who frame the matter as such acknowledge that the supervenience conditions are too complicated to specify (for now, anyway). Recall that the model-theoretic approach assumes theories to be underwritten by a commonsense framework that guides the interpretation of a theoretical language. In light of this, Azzouni's gross regularities should be seen as the kind of causes (or some subset thereof) that are assumed and implemented in reference fixing and the operationalization of theory. From the methodological point of view, we implicitly assume the commonsense framework to be correct—indeed, it must be assumed correct in order for us to successfully interpret a language/theory. Externalists can therefore answer charges that commonsense causal relations are scientifically suspect by pointing to the vital role they play in

---

<sup>6</sup>In I.6 of Azzouni (2000: 63), he outlines a theory of "gross regularities" which he describes as a "boundary layer" between scientific theories and their domain of application. He continues, "The irreducible status of (many of) the laws of the special sciences to physics, and of (many of) the laws of other special sciences that (metaphysically speaking) underlie them has its analogue in the irreducible status of a body of knowledge about middle-sized objects that, strictly speaking, does not belong to any science; we, as lay-persons and as specialists, rely on a body of knowledge about such objects that is only partially reducible to *any* science what hands-on experience and craft teach are...gross regularities. Some of these regularities are expressible in language, but many are items of kinesthetic "know-how".

the interpretation of scientific theory.

In the preceding section I introduced the Carnapian scheme as both an acceptable outline for assessing legitimate ontological questions and a framework by which we might affirm ontological commitment. When we assess causality in light of this framework it helps to illuminate the sense in which causality might be considered an event-kind that transcends theory. Recall the four-fold distinction of question-types: particular-internal, general-internal, practical-external and theoretical-external. Internally construed, causality features pervasively in our scientific language. Consider a particular-internal question like, "What are some of the common causes of HIV infection?", or a general-internal question like "Does causality play a role in medical science?". Determinate answers to the former are provided via the deductive or inductive apparatus internal to the relevant language, and once answers to the former (and other questions like it) are provided, an affirmative answer to the latter is obvious.

On the other hand, ontological questions surrounding causality become more interesting when assessed in light of the external perspective. Practical-external questions concern the acceptability of entire languages or specific features of those languages. Earlier I described such questions as being decided via practical cost-benefit analysis. But prior to that I argued that causality should not be seen as being grounded in communal agreement. Specifically, my claim was that causes are events in the natural world, the *knowledge* of which are contingent on there being creatures like us (or other creatures suitably endowed) to infer them; moreover, we cannot help but invoke causality in certain contexts (also, given certain assumptions about what exists, such as concrete objects, events, relations, and time). So, whereas causality may not be a metaphysically necessary constraint on reference—i.e. it need not be invoked as the mechanism of reference in all possible worlds—it is, nevertheless, a constraint on the way in which we take the referential relation to work for non-formal languages. Causality, so understood, should be seen as only weakly conventional.

Clearly we did not, at some point in time, entertain practical considerations over whether or not to adopt causality as part of our conceptual framework. We do, however, entertain such pragmatic considerations when questions arise over the ontological legitimacy of certain postulates (e.g. the luminiferous aether), or in the selection of alternative frameworks of explanation for special phenomena (e.g. subatomic events, where causality does not feature as an *explicit* part of the explanatory framework). Setting such special cases aside, causality seems a fixture for medium-sized interpreters like us who cannot help but intuit the world as spatiotemporal. Cartwright (1983) has persuasively argued that, despite its apparent

dispensability in certain branches of physics, abandoning causality would cripple the vast majority of the sciences. She concludes, on this basis, that until physical theories account for the causes we employ in other sciences, they are incomplete. Recall, as well, Lewis' observation that causality oversteps theory by constraining our interpretation of the very languages/ theories in which the term 'cause' is used to denote a genuine kind of event or relation. These considerations suggest that causality is indeed theory/ language transcendent, which is to say that it features among the most general, pre-theoretic structures at work in our thought about the world.

This proposed transcendence is explicit when we examine explanatory frameworks assumed in the study of our means of access to the world. Take, for example, the aforementioned externalist strategy for naturalizing mental content, where an explainer invokes the very notion she is explaining. In causal-informational-representational accounts of mental content, the causal exchange of conserved energy between 'bodies' is assumed as part of the physiological explanation of our ability to infer high-level, commonsense causes (i.e. gross regularities). It is a case of an explainer assuming in their explanans the same relation which features as the explanandum. Indeed, there are important differences between the kinds of microcauses that occur at, say, the molecular level, and those that are invoked at the level of gross regularities. The former are usually construed in aggregate, labeled causal *processes*, since each salient microcausal event in a process plays roughly the same role in an explanation of our ability to perceive some macro-level causal event. However, when abstracted and recast in the general form of 'C caused E', there is no important difference between causes at the micro and macro level.

Causes—like objects, events, certain properties, and certain other relations—unambiguously feature in our best theories about the world. Recall, however, that under the Carnapian framework one cannot point to the elements of a theoretical domain as providing theoretical-external questions with a sense. Carnap is correct that a term's mere appearance in a language cannot bestow metaphysical reality on its specified referent in our theoretical domain. But causality's place 'outside', or antecedent to, theory can be taken to provide such a sense. This stance certainly runs afoul of Carnap's ontological views, but in his presentation he fails to acknowledge the distinctive status of general ontological categories like objects, events, properties and (causal) relations. Such categories precede a specified domain of scientific discourse, and they function as the basis upon which such theories are constructed. To acknowledge as much, and to make the external shift in perspective is a legitimate way to avoid Hempel's dilemma. Regrettably, internalists are unlikely to concede the objectivity of causality since, for them, 'objective' is

term applicable only to those properties with which we are *acquainted*, in the Russellian sense. Hence, causality will not count because we are not acquainted with causes—we theoretically impose them into the domains where they feature. Externalists, in contrast, emphasize that pre-theoretic commonsense leaves no ambiguity over the objective existence of causality and its metaphysical cohorts. Moreover, they see both the practice of natural science as methodologically dependent on their supposition, and the impressive success of our best theories as their vindication.

### 3.6 Natural Kinds and Realism Revisited

Contemporary causal-informational-representational theories offer compelling theoretical accounts of mental content assuming a causal theory of reference. However, that a causal approach can be made to work as successfully in an account of the semantic content of our scientific theories is not as obvious. Indeed, our ability to refer to many natural kinds is undoubtedly the result of causal interaction between samples and those who investigate their inner structures. However, even if we set aside the conceptual problems surrounding the determination of sameness among samples—problems that are 'resolved' by stipulation—causal theories of reference continue to face another set of problems.

In science, the naming process for a natural kind involves various stages of investigation that include, among others, detection, classification, confirmation and, finally, naming. Consider this in the context of Azzouni's earlier example involving the physicist who causally interacts with individual atoms using modern technology. A thoroughgoing naturalist treatment of natural kinds and, more broadly, theoretical terms would need to involve operationalizing the theories the physicist deploys in her investigations, since operationalization gives an account of the procedures followed in the measuring of each property attributed to a kind. Operationalization would therefore track the various causal process involved in the investigative stages leading to the naming of, for example, some new particle.

However, even operationalizing theory cannot entrench a given causal account of semantic reference for natural kinds. On the one hand, operationalization works to systematize the overarching causal process and offers a scientifically legitimate means of tracking the relevant causal processes responsible for our interactions with samples. But giving an account in terms of extant scientific procedures is problematic for a metaphysical naturalist due to the ever-changing nature of such procedures: old ones are continually abandoned for new ones and this affects the



status of preceding accounts if we are to assume, as science does, that newer procedures are superior to older ones. Moreover, newer procedures often produce different results from older ones. So, just as science undergoes continual revision, so will any naturalistic rendering of semantic reference for a natural kind term or theoretical term in causal terms. Of course, this poses no serious problem for the methodological naturalist of the Quinean stripe, but it illustrates why a metaphysical naturalist/realist like Lewis sees the causal approach, on its own, to be inadequate.

Indeed, the gap between the referential scope of a term and our ability to underwrite that scope via causality will vary from term to term, and given a linguistic division of labour it will also vary over time. Furthermore, as Lewis (1984: 227) observes, sometimes a causal theory is just ill-suited to explicate a particular case of reference.

The causal theory often works, but not as invariably as philosophers nowadays tend to think. Sometimes an old-fashioned descriptivism works better; sometimes there are puzzling intermediate cases in which causal and descriptive considerations seem to tug in opposite directions.<sup>7</sup>

Lewis' discontent with a purely causal approach results in his looking elsewhere for a bona fide constraint on interpretation, and he finds one in a 'natural hierarchy' of eligible kinds (*qua* natural properties), where the world is ordered such that only an elite minority of objects can be given tidy individuation conditions. Things like fundamental physical particles reside at, or near, the apex, while less-elite objects—ones much harder to individuate—like persons, bicycles and basketballs fare worse in the natural bureaucracy, but remain related to the fundamental joints in nature. Earlier I claimed that Lewis' failure to provide a positive account for his considering those specific sets as eligible, or as hierarchically arranged in the way he takes them to be, would lead to criticism from other naturalists who require a naturalistic explanation for choosing the joints he does. However, Lewis seems quite content to leave the specification of elite kinds to the physicists and take their claims at face value. This is certainly a reasonable strategy, but given the problems of theory change and changing procedures, any appeal to extant physics as an arbiter renders Lewis' naturalism a species of methodological/ epistemological naturalism. And this, of course, renders his claim of objective reality for such joints an article of metaphysical faith.

---

<sup>7</sup>Peter Unger examines a host of such examples in sections 6 - 10 (pp. 21-43) of his, "The Causal Theory of Reference", *Philosophical Studies* 43 (1983), pp. 1-45.

Causal theories of reference can aid in showing the extent that causal resources can work to underwrite the reference of natural kinds and theoretical entities. But given the stated problems—science's apparent reliance on descriptions involving theoretical terms, theory change, and the ever-changing procedures that enable for substantive causal interaction with samples—the externalist must acknowledge that the reference relation for natural kinds can be only partially and provisionally underwritten by causality. Moreover, externalists must concede that the causes to which they appeal in referential accounts of names and natural kinds bear on our *knowledge* of those causes.

The issues surrounding realism that I engaged in earlier chapters are, to some extent, orthogonal to questions surrounding the mechanism of reference. Externalists, as should be clear by now, are all realists to the extent that they are willing to countenance, at the very least, the objective existence of some things that fall under the metaphysical categories of object, event, property and relation. Lewis' realism involves ranking these things according to our ability to specify clear individuation conditions for them. Hence, on the Lewisian picture, things that are smaller and simpler have, in some sense, a higher degree of 'reality' than others. Causes, too, seem amenable to a similar ranking hierarchy, where, say, the *measurable* exchange of conserved energy between bodies constitutes an elite kind of causal interaction, while the vague, difficult-to-measure causes taken to obtain between less-elite objects and events rank lower. However, some—myself among them—prefer to think of the two as distinct *kinds* of causes, given the problematic and counterintuitive results that the hierarchical approach gives rise to.

Consider that hierarchists like Lewis rank everyday, medium-sized objects lower on the hierarchy than ones much smaller and easier to individuate, like fundamental particles. The problem with this view is that the former are both intuitively and empirically less-contentious—that is to say, we know they exist about as well as we know anything. Constructive empiricists can therefore be seen to endorse the inverse, since they reject sentences involving reference to unobservables as truth valued. Rather than opt for constructive empiricism's own counterintuitive error-theoretic (or *fictionalist*, depending on who you ask) treatment of unobservables, the sort of *proto-realism* I advocate takes the preceding tension as a failing of the ranking system itself. Instead, realist/ antirealist border disputes are solved by granting genuine reality to the posits appearing in well-established scientific theories, with special exclusionary provisions made for mathematical objects and those idealized objects which figure in scientific laws and axioms. Ultimately, individual commitment to the degree of objective reality for an object, property, event or relation will vary among theorists—even among those assumed to be working

within the seemingly parochial confines of either 'ism'.

### 3.7 General Conclusion

A curious feature of the reference relation is that, in certain contexts, it functions as an *ideal* relation that hangs on an *intended* interpretation for the theory/language in question. In mathematics, for example, the natural numbers can be given a set-theoretic reduction whereby each natural number refers to its set-theoretic counterpart.<sup>8</sup> The reference relation is thus perfect and unambiguous, and this is characteristic of domains whose elements are wholly abstract. The internalist approach mirrors such accounts where the reference relation is conceived to be conceptually direct. The important difference, of course, is that internalists speciously assume this model to suffice in contexts where the relevant domain consists of concrete objects and events. The problem with any such approach is that it cloaks the important story that accounts for constraints on our referential assignments *via* some mechanism. Hence, the internalist approach gives rise to supposed problems of referential reinterpretation, like the ones emphasized by permutationists.

The point I mean to accentuate is this. Even where a referential relation is conceived as a function from a term to a concept for some worldly thing, underwriting that concept is a complex causal interaction (whether accounted for, or not) that allowed us to detect (in some cases, *derive*), classify and ultimately name the item, and hence enable us to continually specify members of the term's extension. Putnam's unwillingness to recognize the forcefulness of this point is the model-theoretic argument's central failing.

A similar problem might be said to afflict *truth*, which also functions ideally in formal contexts once an interpretation is given, yet staggers when applied to natural language sentences whose content is meant to capture features of the concrete world.

The preceding considerations suggest that notions like reference and truth are 'at home' in a purely formal context. Their deployment in languages used to describe the world should therefore be conceived as closely analogous to that of applied mathematics: specifically, to assume a role similar to that of geometry in the study of physical objects. Given this, any question that implicitly treats truth and/or reference as a purely non-formal notion is sure to encounter problems. Hence, we should expect attempts to define reference and truth in terms of causal-

---

<sup>8</sup>Admittedly, this is a contentious point in the philosophy of mathematics, but it is illustrative of a context where the reference relation is ideal.

ity and correspondence to encounter the pitfalls they do. However, if we conceive of reference and truth in the recommended manner—as paradigmatically ideal relations—a host of philosophical problems surrounding these issues dissolve.

The type of proto-realism I endorse makes no brazen metaphysical claims about there being a *fact* in nature about how the things we refer to must be individuated in order for claims to be true. Rather, contingencies like the specific nature of human physiology and the state of a scientific community at a given stage of human development are taken to influence the causally relevant units we use to specify an interaction. Both our ability to successfully communicate in ordinary language and the impressive success of modern science suggest that, to an important degree, we have indeed identified some of nature's joints. Are we not therefore *referring* to them? Perhaps. But the history of science admonishes that we should expect to discover a good deal about nature's joints that is not already implied by even our best theories.

## Bibliography

- Anderson, David Leech. (1993) "What is the Model-Theoretic Argument?" *The Journal of Philosophy*, 90: 311-322.
- Azzouni, Jody. (2000) *Knowledge and Reference in Empirical Science*. International Library of Philosophy. London; New York: Routledge.
- Berry, George D. W. and John R. Myhill. (1951) Symposium: "On the Ontological Significance of the Löwenheim-Skolem Theorem" In White, M. (ed.), *Academic Freedom, Logic, and Religion*. Philadelphia: University of Pennsylvania Press, pp. 39--55.
- Bird, Graham H. (1995) "Carnap and Quine: External and Internal Questions." *Erkenntnis* 42: 41-64.
- Boyd, R. (1983) "On the Current Status of the Issue of Scientific Realism". *Erkenntnis* 19: 45-90.
- . (1985) "Lex Orandi est Lex Credenti" in Van Fraassen, Bas C., Paul M. Churchland, and C. A. Hooker. *Images of Science :Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*. Science and its Conceptual Foundations. Chicago: University of Chicago Press.
- Burge, Tyler. (1979) "Individualism and the Mental" in French, Uehling, and Wettstein (eds.) *Midwest Studies in Philosophy IV*, Minneapolis: University of Minnesota Press, pp. 73-121.
- Carnap, Rudolf. (1956) *Meaning and Necessity : A Study in Semantics and Modal Logic*. 2d , enl ed. Chicago: University of Chicago Press.
- Carnap, Rudolf and Rolf A. George. (1967) *The Logical Structure of the World and Pseudoproblems in Philosophy*. Translated by Rolf A. George [Der logische Aufbau der Welt. English]. Berkeley: University of California Press.
- Cartwright, Nancy. (1983) *How the Laws of Physics Lie*. Clarendon Paperbacks, Oxford: Oxford University Press.
- Chakravartty, Anjan. (2001) "The Semantic or Model-Theoretic View of Theories and Scientific Realism", *Synthese* 127: 325-345.

- Chalmers, David J. (2006) "Ontological Anti-Realism" (forthcoming). URL = <http://consc.net/papers/ontology.pdf> (accessed 03/07).
- Chomsky, Noam and Alexander George. (1989) *Reflections on Chomsky*. Oxford; New York: B. Blackwell.
- Cummins, Robert. (1989) *Meaning and Mental Representation*. Bradford Book; Cambridge, Mass.: MIT Press.
- Da Costa, Newton and Steven French. (1990) "The Model-Theoretic Approach in the Philosophy of Science", *Philosophy of Science*, Vol. 57, No. 2: 248-265.
- Demopoulos, William. (2007) "Carnap on the Rational Reconstruction of Scientific Theories" in *The Cambridge Companion to Carnap*, Richard Creath and Michael Friedman (eds.), forthcoming, Cambridge University Press.
- . (2003) "On The Rational Reconstruction of Our Theoretical Knowledge", *British Journal for the Philosophy of Science*, no. 54, 371-403.
- Demopoulos, William and Michael Friedman. (1985) "Bertrand Russell's the Analysis of Matter: Its Historical Context and Contemporary Interest". *Philosophy of Science* 52, no. 4: 621-639.
- Devitt, Michael. (1997) "On Determining Reference" in *Language and Thought*, Alex Burri (ed.); New York: Walter de Gruyter: 112-121.
- . (1983) "Realism and the Radical Putnam". *Nous*, 17: 291-301.
- . (1980) "Critical Notice of *Meaning and the Moral Sciences* by Hilary Putnam." *Australasian Journal of Philosophy* 58: 395-404.
- Devitt, Michael and Kim Sterelny. (1987) *Language and Reality :An Introduction to the Philosophy of Language*. Cambridge, Mass.: MIT Press.
- Dretske, Fred I. *Naturalizing the Mind*. (1997) The Jean Nicod Lectures; Cambridge, Mass.: MIT Press.
- . (1988) *Explaining Behavior :Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.
- . (1980) *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.

- Dunham, William. (1990) *Journey Through Genius: Great Theorems of Mathematics*; John Wiley & Sons Canada, Ltd.
- Eliasmith, Chris. (2000) *How Neurons Mean: A Neurocomputational Theory of Representational Content*. doctoral dissertation. URL = <http://watarts.uwaterloo.ca/~celiasmi/Papers/Eliasmith.2000.How%20neurons%20mean.Phd.pdf> (accessed 03/07).
- Feigl, Herbert and Grover Maxwell. (1962) *Scientific Explanation, Space and Time*. Minnesota Studies in the Philosophy of Science ;. Vol. 3. Minneapolis: University of Minnesota Press.
- Field, Hartry. (1982) "Realism and Relativism". *The Journal of Philosophy* 79, no. 10, Seventy-Ninth Annual Meeting of the American Philosophical Association Eastern Division: 553-567.
- . (1980) *Science without Numbers :A Defence of Nominalism*. Princeton, N.J.: Princeton University Press.
- . (1975) "Conventionalism and Instrumentalism in Semantics". *Nous* 9, no. 4: 375-405.
- Friedman, Michael. (1999) *Reconsidering Logical Positivism*. Cambridge, U.K.; New York: Cambridge University Press.
- George, Alexander. (1985) "Skolem and the Löwenheim-Skolem Theorem: A Case Study of the Philosophical Significance of Mathematical Results". *History and Philosophy of Logic*, 6: 75-89.
- Giere, Ronald N. (2004) "How Models are Used to Represent Reality". *Philosophy of Science*, 71: 742-752.
- . (2000) "Theories" in Newton-Smith, W.H. (ed.) *A Companion to the Philosophy of Science*. Blackwell Companions to Philosophy; Massachusetts, Oxford, Victoria: Blackwell Publishing.
- . (1999) *Science without Laws*. Science and its Conceptual Foundations. Chicago: University of Chicago Press.
- . (1988) *Explaining Science :A Cognitive Approach*. Science and its Conceptual Foundations. Chicago: University of Chicago Press.

- Glymour, Clark. (1982) "Conceptual Scheming, or Confessions of a Metaphysical Realist", *Synthese* 51: 169-180.
- Hacking, Ian. (1983) *Representing and Intervening :Introductory Topics in the Philosophy of Natural Science*. Cambridge Cambridgeshire; New York: Cambridge University Press.
- Hale, Bob and Crispin Wright. (1997) "Putnam's Model-Theoretic Argument Against Metaphysical Realism" in *A Companion to the Philosophy of Language*; Hale & Wright (eds.) Blackwell Companions to Philosophy; Massachusetts, Oxford, Victoria: Blackwell Publishing.
- Hempel, Carl. (1962) "The Theoretician's Dilemma" in Feigl, Herbert, Grover Maxwell, and University of Minnesota. *Scientific Explanation, Space and Time*. Minnesota Studies in the Philosophy of Science ;. Vol. 3. Minneapolis: University of Minnesota Press.
- Hodges, Wilfrid. (2005) "Model Theory", *The Stanford Encyclopedia of Philosophy (Winter 2005 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2005/entries/model-theory/>.
- Hume, David. (2000) *A Treatise of Human Nature*. David Fate Norton and Mary J. Norton eds., Oxford Philosophical Texts, Oxford, UK, Oxford University Press.
- Khlentzos, Drew. (2004) "Semantic Challenges to Realism", *The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2004/entries/realism-sem-challenge/>.
- Kripke, Saul A. (1980) *Naming and Necessity*. Oxford: B. Blackwell.
- Lewis, David K. (1970) "How to Define Theoretical Terms", *Journal of Philosophy*, 67: 427-45.
- . (1983) "New Work for a Theory of Universals". *Australasian Journal of Philosophy*; 61: 343-77.
- . (1984) "Putnam's Paradox". *Australasian Journal of Philosophy*; 62: 221-236.
- . (2001) "Forget about the 'correspondence theory of truth'", *Analysis* v.61, n.272: 275-280(6).



- Miller, Alexander. (2005) "Realism", *The Stanford Encyclopedia of Philosophy (Fall 2005 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2005/entries/realism/>.
- Newman, M.H.A. (1928) "Mr. Russell's Causal Theory of Perception", *Mind*, 37: 137-48.
- Newton-Smith, W.H., ed. (2000) *A Companion to the Philosophy of Science*. Blackwell Companions to Philosophy; Massachusetts, Oxford, Victoria: Blackwell Publishing.
- Psillos, Stathis. (1999) *Scientific Realism :How Science Tracks Truth*. Philosophical Issues in Science. London; New York: Routledge.
- Putnam, Hilary. (1983a) *Realism and Reason*. Philosophical Papers, Vol. 3. Cambridge Cambridgeshire; New York: Cambridge University Press.
- . (1983b) "Models and Reality" in *Realism and Reason*. Philosophical Papers, Vol. 3. Cambridge Cambridgeshire; New York: Cambridge University Press.
- . (1981) *Reason, Truth and History*. Cambridge Cambridgeshire; New York: Cambridge University Press.
- . (1979a) *Mind, Language, and Reality*. Putnam, Hilary. Philosophical Papers, Vol. 2. Cambridge Eng.; New York: Cambridge University Press.
- . (1979b) "The Meaning of 'Meaning'" in *Mind, Language, and Reality*. Putnam, Hilary. Philosophical Papers, Vol. 2. Cambridge Eng.; New York: Cambridge University Press.
- . (1979c) *Meaning and the Moral Sciences*. International Library of Philosophy and Scientific Method. London: Routledge and Kegan Paul.
- . (1976) "Realism and Reason". *Proceedings and Addresses of The American Philosophical Association*, 50, 6: 483-498.
- . (1975a) *Mathematics, Matter, and Method*. Philosophical Papers, Vol. 1. London; New York: Cambridge University Press.
- . (1975b) "What is Mathematical Truth" in *Mathematics, Matter, and Method*. Philosophical Papers, Vol. 1. London; New York: Cambridge University Press.

- Quine, W. V. (1981) *Theories and Things*. Cambridge, Mass.: Harvard University Press.
- . (1976a) *The Ways of Paradox, and Other Essays*. Rev. and enl. ed. Cambridge, Mass.: Harvard University Press.
- . (1976b) "On Carnap's Views on Ontology" in *The Ways of Paradox, and Other Essays*. Rev. and enl. ed. Cambridge, Mass.: Harvard University Press.
- . (1969a) *Ontological Relativity and Other Essays*. John Dewey Essays in Philosophy; Vol. 1. New York: Columbia University Press.
- . (1969b) "Natural Kinds" in *Ontological Relativity and Other Essays*. John Dewey Essays in Philosophy; Vol. 1. New York: Columbia University Press.
- . (1969c) "Ontological Relativity" in *Ontological Relativity and Other Essays*. John Dewey Essays in Philosophy; Vol. 1. New York: Columbia University Press.
- . (1960) *Word and Object*. Studies in Communication. Cambridge: Technology Press of the Massachusetts Institute of Technology.
- Russell, Bertrand. (1954) *The Analysis of Matter*, New York: Dover Publications Inc. (originally published in 1927).
- Soames, Scott. (2003) *Philosophical Analysis in the Twentieth Century, vols. I & II*, Princeton University Press; Princeton, New Jersey.
- Stalnaker, Robert. (2007) "Starting in the Middle"; *Our Knowledge of the Internal World: The John Locke Lectures*; Faculty of Philosophy: Oxford University, Oxford, UK; 5--02-2007.
- . (1990) "Narrow Content" in Anderson and Owens (eds.) *Propositional Attitudes* CSLI Publications.
- . (1993) "Twin Earth Revisited" *Proceedings of the Aristotelian Society*, Vol. XCIII, pp.297-311.
- Suppe, Frederick. (2000) "Ramsey Sentences" in Newton-Smith, W.H. (ed.) *A Companion to the Philosophy of Science*. Blackwell Companions to Philosophy; Massachusetts, Oxford, Victoria: Blackwell Publishing.

- 
- . (1989) *The Semantic Conception of Theories and Scientific Realism*, Urbana, IL: University of Illinois Press.
- Suppes, Patrick. (1967) "What is a Scientific Theory?" in *Philosophy of Science Today*; Sidney Morgenbesser (ed.), pp. 55-67, Basic Books.
- . (1960) "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences"; *Synthese* 12: 287-301.
- Unger, Peter. (1983) "The Causal Theory of Reference", *Philosophical Studies* 43, pp. 1-45.
- Van Fraassen, Bas C. (1997) "Putnam's Paradox: Metaphysical Realism Revamped and Evaded". *Philosophical Perspectives*, v. 11, Mind, Causation and World.
- . (1989) *Laws and Symmetry*. Oxford: Clarendon Press.
- . (1985) "Empiricism in the Philosophy of Science" in Van Fraassen, Bas C., Paul M. Churchland, and C. A. Hooker. *Images of Science :Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*. Science and its Conceptual Foundations. Chicago: University of Chicago Press.
- . (1980) *The Scientific Image*. Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.
- Van Fraassen, Bas C., Paul M. Churchland, and C. A. Hooker. (1985) *Images of Science: Essays on Realism and Empiricism, with a Reply from Bas C. Van Fraassen*. Science and its Conceptual Foundations. Chicago: University of Chicago Press.