CARNAP VISITS CANBERRA:
UPDATING THE LOGICAL POSITIVIST CRITERIA OF COGNITIVE SIGNIFICANCE

A thesis submitted to
Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

by
Andrew Whiteley Magrath

May 2012
Thesis written by
Andrew Whiteley Magrath
B.A., Oberlin College, 2004
M.A., Kent State University, 2012

Approved by
Deborah C. Smith___________________________, Advisor
David Odell-Scot__________________________, Chair, Department of Philosophy
Timothy Moerland___________________________, Dean, College of Arts and Sciences
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** .................................................................................. v

## CHAPTER

### I. INTRODUCTION .................................................................................. 1  
Logical Positivism and the Canberra Plan: A Brief Introduction ............... 1  
Guiding Question & Structure of the Thesis ........................................ 3  
My Motivation .................................................................................. 4  
Structure & Methodological Considerations ....................................... 6  

### II. CARNAP AND LOGICAL POSITIVISM ........................................... 8  
The Use of Logic in Positivism .......................................................... 8  
The Positivists' Reductionism & Expressibility .................................. 9  
A Theory of Meaning: Analytic a Priori and Synthetic a Posteriori ....... 11  
Ramsification: Observation Terms and Theory Terms ......................... 13  
Carnap's Four Criteria for Cognitive Significance ............................... 18  
Unpacking C1: Empirical Criteria ...................................................... 18  
Unpacking C2: Protocol Sentences ................................................. 19  
Unpacking C3: Truth-Conditions .................................................... 21  
Unpacking C4: Verification .............................................................. 22  

### III. THE COLLAPSE OF LOGICAL POSITIVISM .................................. 24  
The Trouble with C2: Protocol Sentences, *T-terms*, and *O-terms* .... 24  
The Trouble with C4: Verification .................................................... 29  
Confirmation: A Failed Replacement for Verification ......................... 31  
Where Does this Leave Carnap's Four Criteria? ................................. 36  

### IV. THE CANBERRA PLAN .................................................................. 38  
The Two Step: Goals & Explication .................................................. 39  
Step-One: Platitude Gathering & Analysis ....................................... 40  
The Lewis Tradition ..................................................................... 45  
The Jackson Tradition .................................................................. 49  
Step-Two: Location ........................................................................ 52  
Near-Realizers ............................................................................ 53  
The Unofficial Third Step ............................................................. 55
ACKNOWLEDGEMENTS

The transition from physicist to philosopher has not always been easy, and I have many to thank for lightening my burden, setting me straight, or simply not strangling me whenever I would hear an argument, roll my eyes, and insist on seeing some collaborating data. First, I would like to thank my parents for always encouraging me to keep as many doors in my academic life open as possible. While in middle school I might not have understood why anyone would actually want to read a non-science book, now I understand and am eternally grateful. I also would like to thank Tanigawa Junior High School and the other residents of Daito, Japan for expanding my horizons farther than I thought possible.

I am grateful to the staff, faculty, and my fellow students at both Kent State University and Oberlin College. Education is an activity best shared with others. I have had wonderful teachers and thoughtful peers. I thank you all. I would particularly like to thank Dr. Daniel Styer of Oberlin College for all his kindness over the years. I am also deeply indebted to Dr. Kwang-Sae Lee, Dr. Gene Pendleton, and Dr. Robert Twieg for acting as readers. As I went through the painful process of editing down the original drafts of this thesis I frequently quoted to myself a passage from the *Dao De Jing* that Dr. Lee first introduced me to:

*A man is born gentle and weak.*

*At his death he is hard and stiff.*

*Green plants are tender and filled with sap.*

*At their death they are withered and dry.*

*Therefore the stiff and unbending is the disciple of death.*

*The gentle and yielding is the disciple of life.*

I especially want to thank my advisor, Dr. Deborah Smith. She alone had the bravery to dive into earlier drafts of this document chainsaw in hand. It is thanks to her insightful cuts and edits that this thesis was slashed down to its current "nicer and conciser" form. From the modal logic courses she offered out of her home, to her thorough and thoughtful approach to obstacles, Dr. Smith has shown me not just what it is to be a philosopher but also what it takes to be a wonderful teacher and friend. Should I ever attain even some minor accomplishment, I owe it to her diligence, kindness, and luminary insights.

Last, I'd like to thank the members of the SSC—they know who they are. Thank you all.
Logical Positivism and the Canberra Plan: A Brief Introduction

The Vienna Circle, the progenitors of logical positivism,\footnote{Although some philosophers have made a distinction between 'logical empiricism' and 'logical positivism' (Carnap among them) it is now fairly common to use the terms interchangeably. I shall, therefore, continue in the contemporary tradition and treat the terms as synonyms.} were a collection of philosophers, scientists, mathematicians, and logicians. The early logical positivists saw the success of logic and the empirical sciences as emblematic of the direction in which philosophy should move. The goals of the logical positivist movement can typically be divided into the following main tenets:

A) A theory of meaningfulness that asserts that a term is only meaningful when the sentence containing the term is verifiable either deductively or through induction and sense data/direct observation. All other terms are meaningless.

B) The rejection of metaphysics and speculative philosophy. The positivists asserted that metaphysical terms were either meaningless or unnecessarily grandiose empirical claims.

C) A commitment to physicalism. The physicalism forwarded by the logical positivists was informed by their unique emphasis on language. The logical positivists' version of physicalism can be understood as the claim that all empirical claims are actually a claim of physics. All of the sciences
(including the human sciences) could be reduced to or translated into the language of physics.

D) Sentences of mathematics and logic are tautologies and therefore analytic\(^2\) a priori.\(^3\)

Although all four pillars will be discussed in this thesis, the main focus will be understanding (A).

Although logical positivism enjoyed a period of acceptance, problems began to erode the base of support for the project. One of the greatest questions was whether philosophical analysis remained a worthwhile project. Emerging from the vacuum left with logical positivism's collapse is the contemporary philosophy of the Canberra Plan. The Canberra Plan originated with David Lewis, Frank Jackson, and other philosophers primarily from the Philosophy Program of the Research School of Social Sciences at the Australian National University in Canberra. Practitioners of the Plan hope to reinforce the importance of a form of analysis in philosophy. Although the Canberra Plan shares many traits with logical positivism and can be understood as emerging from the positivist tradition, the Plan represents a more flexible approach to philosophical analysis.

\(^2\) The analytic/synthetic is a distinction relating to the truth of propositions. Analytic truths are true merely in virtue of the meaning of the component pieces and the way these components are combined. Synthetic truths are true not merely in virtue of the meanings of the component pieces and manner that they are combined, but importantly also in virtue of the way the world is.

\(^3\) The a priori/a posteriori is an epistemic distinction. Propositions that are a priori are justified in virtue of reason alone: where reason in this sense involves both rational insight and deductive methods. A posteriori knowledge, on the other hand, requires experience with the world. Knowledge of this sort depends on sensory experience, inductive reasoning, and empirical testing of the world.
Practitioners of the Canberra Plan, often called 'planners', seek to analyze concepts in two steps. The first step of the Canberra Plan is the assembling of concepts relevant to the notion under examination. These related concepts are weighed, ranked, and possibly eliminated in favor of other concepts in order to produce a descriptive sentence of the concept being analyzed. In step-two of the Canberra Plan, the descriptive sentence is handed off to a relevant science so that a referent of the concept can be discovered in the world.

Guiding Question & Structure of the Thesis

Of particular interest to this thesis is the manner in which tenet (A) is understood by the logical positivist, Rudolf Carnap. Carnap proposes a systematized theory of meaning broken into four criteria:

(C1) The empirical criteria for the term, 'a', are known.

(C2) It has been stipulated from what protocol sentences 'S(a)' is deducible.

(C3) The truth-conditions for 'S(a)' are fixed.

(C4) The method of verification of 'S(a)' is known.\(^4\)

Although the logical positivists' execution of their basic tenets and intuitions is questionable, I believe that the original ideals behind the logical positivist project remain worthwhile. I contend throughout this thesis that the major problem with logical positivism was the manner in which their ideals were cashed out: not a problem with the

ideals themselves. My question emerges: how to preserve the original intention of the logical positivists' project (particularly with regards to a theory of meaning) but also avoid the pitfalls?

I will argue that of Carnap's original criteria of meaningfulness, C2 and C4 are philosophically weak, and will need to be replaced, but C1 and C3 remain on strong philosophical ground. I assert that C2 can be directly replaced with step-one of the Canberra Plan, as the two play functionally similar roles. Yet, despite similarities, step-one of the Canberra Plan is less problematic than C2. I will replace C4 with a version of step-two of the Canberra Plan.

My Motivation

One of the unique qualities of logical positivism that drew me to the philosophy is that, unlike other philosophies that have as their locus an individual thinker, logical positivism was forwarded by a relatively unified group of philosophers and scientists. This group was known as the Vienna Circle. The Vienna Circle established a fairly unified, cogent, and popular philosophy even early in logical positivism's development. Each member could independently defend and expand the philosophy while still having access to the other members' aid as a sounding board. Although no philosopher works in a vacuum, it is rare and interesting to see philosophy crafted by a group.

Given that logical positivism arose from a dedicated group of philosophers, it can be difficult to choose a single voice to focus upon. Each member of the Vienna Circle borrows so heavily from the others it is often unclear exactly which ideas are unique to a
given philosopher. The synergy between group and individual may be intriguing, but it can also make the philosophy daunting to unpack. While acknowledging the group origin of logical positivism, I feel that I require an unequivocal voice for the Vienna Circle. To that end, I have chosen Rudolf Carnap to be the primary focus of this thesis. So, although other prominent logical positivists will necessarily need to be referenced in this examination, I will most often engage the writings and positions of Carnap.

The German philosopher, Rudolf Carnap (1981-1970), was one of the more prominent members of the Vienna Circle, and deeply influential in the analytic tradition. Carnap studied logic under Frege. Lacking the financial means to purchase the *Principia*, Carnap corresponded with Russell via letters. Russell was so impressed with Carnap's inquiry that he responded by copying long sections of his *Principia* longhand. Additionally, Carnap attended lectures by Edmund Husserl. Husserl's phenomenological stamp can be seen in many of Carnap's early attempts to reduce physics to phenomena. Yet, like many within the Circle, Carnap was not a byproduct of philosophy alone. In addition to his philosophical training, Carnap studied physics. One of his professors was Albert Einstein at the University of Berlin where Carnap studied relativity. Despite making contributions to physics, Carnap is best known as one of the prominent members of the Vienna Circle and one of the most outspoken and respected defenders of logical positivism.
Structure & Methodological Considerations

The reliance on science is a cornerstone of the logical positivist project, and one that I am sympathetic to. As I shall discuss in greater detail when examining possible objections to the Canberra Plan, there are certain practitioners of the Canberra Plan that assert that the Plan is inherently only able to draw conclusions about physical objects' properties and relations. For the positivists, the natural sciences (particularly physics), represented the best systems of explaining phenomenon in the natural world. Planners seem to share this special relationship with science. Additionally, both planners and positivists tend to see the role of philosophy as working closely with the other departments in the academy (particularly the natural sciences). I wish to preserve the special epistemic and metaphysical status of science which is held by both philosophies. Since the planners and positivists often rely on scientific examples, I too shall forward scientific examples when and where they are appropriate.

This thesis consists of six chapters. In the next chapter, I will explicate logical positivism and introduce the reader to the views of Rudolf Carnap, especially as they relate to his four criteria of meaning. Chapter 3 covers reasons for the decline of logical positivism particularly in relation to Carnap's criteria of meaning. Chapter 4 introduces the reader to the Canberra Plan with emphasis paid to both the Lewis and Jackson traditions. A potential problem with the Plan is introduced in Chapter 5. The final chapter will cover the unification of Carnap's criteria and the Canberra Plan, demonstrating how the addition of the Canberra Plan can solve the problems of Carnap's
original criteria. Additional attention will be paid to solving the problem with the Plan raised in Chapter 5.
The goal of this chapter is to outline the philosophical position of logical positivism as well as provide aspects of positivism that I see as valuable. I will pay particular attention to Carnap's four principles of cognitive significance (criteria for meaning) which, when properly viewed, represent a refined and honed version of the positivist's verificationist theory of meaning.

The Use of Logic in Positivism

The positivists contended that although logic makes no claim about how the world actually is, it is an invaluable tool to the philosopher. The role of logic is made explicitly clear in the positivists' division of knowledge into the a priori analytic and the a posteriori synthetic. Since, on most positivist views, there are no other acceptable combinations; logic could only fall within the a priori analytic form of knowledge, completely divorcing it from statements about how the world actually is (synthetic claims). Under the positivist's agenda, the sole aim of the philosopher was the careful study of logic both as a discipline itself and also as a system that acted as the underpinnings of scientific investigation. The logician, as envisioned by logical positivist, would need to come in two forms: what I shall call theoretical and applied.

The theoretical logician would study logic qua logic: studying logic as a language independent of any particular interpretation provided to terms and predicates. It seems
plausible that although some result of logic may not, at the time of discovery, be helpful in modeling existing natural language theories, but the new result may be useful in modeling a future theory.\textsuperscript{5} The applied logician, on the other hand, works within other disciplines in an attempt to formally axiomatize their theories. The process of axiomatizing a given theory is to identify the aspects of the theory that are to be taken as primitive: the non-logical axioms. Carnap notes, "[applied logic's use serves to] clarify meaningful concepts and propositions, to lay logical foundations for factual science and for mathematics."\textsuperscript{6} The applied logician was to enter other fields, master these fields, and then help bring order to these fields by applying a logical model. The applied logician needed to understand not just logic, but the system that was to be interpreted. In this way, the philosopher was to work with the scientist and mathematician each bringing their unique skills to bear to do the difficult work of understanding the world. The applied logician, if successful, would translate a physical or mathematical theory into first-order logic, allowing logical techniques to be more fully applied to scientific and mathematical matters.\textsuperscript{7}

**The Positivists' Reductionism & Expressibility**

A Theory T is said to be reducible to Theory T* if and only if T postulates nothing over and above T*. Theory T is expressible in T** if the language of T can be

\textsuperscript{5} Even a casual glance at the history of mathematics is filled with examples of mathematical results that where discovered with no practical purposes only to later be crucial in modeling real world phenomenon.
\textsuperscript{6} Carnap, "Elimination," 168.
\textsuperscript{7} The details of this sort of analysis will be discussed in greater detail later in my examination.
translated into the language of T**. That T can be reduced to T* tends to afford some ontological advantage since no additional primitives need to be accepted in order to accept theory T. The fact that T is expressible in T** often has the advantage of clarifying the content of T. The positivists hoped to use both relations to more clearly define theoretical terms in philosophy.

One of the strengths of first-order logic is that it is purely extensional. An extensional language has a semantics that does not rely on the sense of terms, predicates, and function symbols.\(^8\) Within first-order logic, terms, predicates, and function symbols have well defined semantic values. The semantic value of a term is an object in the domain of interpretation. N-place predicates have as semantic values sets of ordered n-tuples of objects in the domain, and n-place function symbols have functions from ordered n-tuples of objects to individual objects in the domain as semantic value. Ambiguity is removed because when a symbol is used in first-order logic, it has an exact interpretation provided by the extensional model. This level of precision within the language of first-order logic is what the positivists hoped to attain within all fields of putative knowledge. Thus, if science could be reduced to and/or expressed in a language of first-order predicate logic, the positivists could effectively create a language in which

---

\(^8\) Various philosophers have examined the distinction between what is meant and what is, roughly speaking, being pointed to. Although there are nuanced differences between their approaches, the distinction being made is that of Frege's sense/reference distinction, or Carnap's intension/extension distinction, or Mill's connotation/denotation distinction. Given the nature of my project, I will use Carnap's preferred terminology of intension/extension. With intension being what is meant by a term where extension means the object the term is meant to point/refer to.
all synthetic a posteriori knowledge could be expressed within the same system as analytic a priori knowledge.

Noting that much of physics could be translated into mathematics with certain key physical notions treated as axioms and that mathematics could be expressed in the language of logic, the early positivists believed that it might be possible to express physics in a language of logic as well. Such a translation would not be a true reduction. Even so, the positivists saw an opportunity to express all scientific knowledge in the language of first-order logic: first, reduce all of empirical knowledge to physics; then, express physics within the language of mathematics; finally, express that in the language of first-order logic. Through a series of translations and reductions, the positivists attempted to discover an axiomatic system that could effectively describe the world.

A Theory of Meaning: Analytic a Priori and Synthetic a Posteriori

One of the traditional problems associated with the synthetic a priori is that such a theory seems to commit one to some kind of mysterious epistemic ability of the mind. The mind must be able to apprehend, through reason alone (e.g. a priori), some necessary feature of the world (a synthetic fact). For most empiricists (and certainly the logical positivists), such a faculty of the mind was too mysterious and problematic. The solution that the positivists forwarded was that all a priori knowledge involves analytic/conceptual necessity. The mind is held to have a faculty that allows it to apprehend truths of mathematics and logic by reason alone, but not truths about the world such as the laws of chemistry or sociology.
This cleaves knowable propositions into two categories: All propositions must either be analytic a priori or synthetic a posteriori. This splitting allowed the logical positivists to develop a verification based theory of meaning. Under meaning theoretic verificationism, a meaningful sentence must have a means by which it could be verified/established as true. The process by which a notion is verified to be true depends on whether it is analytic a priori or synthetic a posteriori. A sentence that is analytic a priori will be verified through a mathematical and/or logical proof. Synthetic a posteriori claims require observation to be verified. Additionally, in contrast to historical empirical traditions, the logical positivists held that individual terms could not have a meaning independent of the sentences in which the term occurred. Within other empirical traditions, a sentence could have a meaning if each of its elements were meaningful and the sentence was grammatically constructed. The positivists found this insufficient given that a term can be meaningful in one context, but not meaningful in another. For example, the term 'absolute' is meaningful in the sentence, 'Absolute zero is the coldest possible temperature.' as the term is being used in a sentence that can be understood. Yet, 'absolute' is not meaningful in the sentence, 'The red line absolutes.' Thus, unlike previous empirical traditions, the positivists place a greater emphasis on the interplay between terms and sentences. For a term to be meaningful, the term must occur in the context of a sentence, and the sentence must have clear conditions under which its truth could be verified.
Ramsification: Observation Terms and Theory Terms

The goal of the positivists to translate all synthetic a posteriori sentences into sentences of physics raised an interesting problem given that the positivists wanted to link a posteriori to observation: many seemingly meaningful terms of physical theories refer to objects, properties, and relations that seem meaningful, but are not directly observable because they are too large, too small, or are simply not the type of thing that is visible. Yet, those theories have great predictive usefulness and explanatory weight. Let us call terms whose referents are not directly observable, 'theory terms' or 'T-terms'.

Theory terms are to be contrasted with observational terms (O-terms) that are terms whose referents can be observed. An example of a theory term would be 'electricity'. One cannot directly observe electricity. Even under the extreme (and potentially dangerous) circumstance of observing electricity arcing, one does not truly observe electricity. One observes a flash of light, hears a crack, and smells ozone. Due to this, 'electricity' is a T-term, as it is not directly observable. Yet, electricity has a myriad of observable effects referred to by observational terms. As just mentioned, in extreme cases one can observe lights, sounds, and smells associated with electricity. In more mundane cases, one might attach a voltmeter to a wire and observe a needle being deflected to the right. Under this formulation 'electricity' can be taken as a T-term, and the O-terms are 'needle', 'moves', 'right', etc.

---

9 An example of an object that is too small would be an electron. An example of an object that is too large would be the shape of the universe. Finally, an example of a physical concept that cannot be directly observed would be a field.
Theory terms present a serious problem to the positivists' project. The sentences of physics are clearly meant to be taken as synthetic. Thus, the language of physics cannot be verified by appeals to purely deductive or mathematical methods. Given the positivists' desire to link a posteriori verification and observation, it might seem as if the positivists would have to conclude that many terms of physics are not meaningful. In order to include theory terms, the positivists needed a way for the non-observable theory terms of science to be linked to observations that could be empirically verified. In order to allow theory terms to remain meaningful, Carnap devised a way to define all theory terms by way of \textit{O-terms}, which could meet the positivists' demands.

To accomplish the linking of \textit{T-terms} to \textit{O-terms}, Carnap drew inspiration from the earlier work of Frank Ramsey.\(^\text{10}\) According to Ramsey, the predicates or the members of one set of terms could be removed, leaving behind existential placeholders. For example, suppose that for some predicate \(\Phi\),

\[\Phi(V_1,\ldots,V_n;S_1,\ldots,S_m),\]

is a sentence such that the \(V\)s are some set of familiar terms and the \(S\)s are some other set of familiar terms. Let us further assume that one wished to eliminate the \(S\) terms (perhaps because they are theory terms). The corresponding Ramsey sentence would be:

\[\exists t_1,\ldots,t_m \, \Phi(V_1,\ldots,V_n;t_1,\ldots,t_m).\]

Here, \(t_1,\ldots,t_m\) are variables that stand in for the \(S\) terms, thus eliminating the \(S\) terms but preserving the predicate, \(\Phi\), and the other set of terms, the \(V\) terms. This process is now known as Ramsification.

Carnap realized the potential of an adoption of Ramsey's techniques in solving the problem of the cognitive significance of theory terms. If all theory terms with non-observable referents could be replaced by sentences involving only observational terms, then the positivists would have a solution to their problem. Returning to the voltmeter example above, consider the statement, 'Electricity is present if and only if the needle of the voltmeter is deflected to the right.' While 'electricity' has a referent that cannot be observed, the movement of a needle on some voltmeter is observable and hence verifiable a posteriori. Generalizing this maneuver, Carnap formalized Nagel's notion of bridge laws\textsuperscript{11} to Ramsey's work. The bridge laws (also called correspondence laws/rules) were sentences consisting of $T$-terms and $O$-terms (including terms for temporal-spatial relations). The bridge law then acts as a kind of translational rubric to move between the $T$-terms and $O$-terms. Thus, the symbols of the theory, $T$, could be readily translated into empirically verifiable observation sentences. Theories could be stripped of all $T$-terms and left as a sentence containing only $O$-terms. Thus, by defining all theory terms by way of observational terms and using bridge laws, positivists were effectively able to provide for the meaningfulness of most scientific terms. Scientific theory terms could be translatable to observational sentences, which were, in turn, empirically verifiable.

Since bridge laws are a mixture of both $O$-terms and $T$-terms, they can be used to strip out either $O$-terms or $T$-terms depending on the direction of the analysis. Thus, bridge laws can be thought of as a function that takes sentences involving singular $T$-terms as arguments and outputs the corresponding $O$-term sentences (or vice versa). For

\textsuperscript{11} See Nagel's *The Structure of Science: Problems in the Logi of Scientific Explanation*. 
example, say the theory term \( t_1 \) is explained by a sentence containing three observation terms \( (o_1, o_2, o_3) \). If \( t_1 \) were to be input into the bridge law function, \( B \), the output would be \( \Phi(o_1, o_2, o_3) \).

In the case of the voltmeter example, the *T-term* ’electricity’ is correlated with a sentence that involves three observation terms: ’the needle’, ’moves’, and ’the right’. When ’electricity’ is input into the bridge, the corresponding observation sentence is output:

\[
B(\text{electricity}) = (\text{The needle moves to the right})
\]

The process of replacing individual terms could be scaled up, so that entire theories could be stripped of *T-terms*. Let the function \( B(x) \) be the set of bridge laws that are thought to govern the theory. When the entire theory \( T[t]^{12} \) is input into \( B \), each individual *T-term* is input into the translational rubric. Every *T-term* in \( T[t] \) is translated into a sentence involving the corresponding *O-terms*. The resulting theory will be expressed completely as *O-terms*. Thus:

\[
B(T[t]) = T[B(t_1), B(t_2), B(t_3), \ldots B(t_n)] = T[o_1, o_2, o_3, \ldots, o_m]
\]

---

\(^{12}\) For this usage of ’\( T[t] \)’, ’\( T \)’ represents the conjunction of all axioms/postulates of the theory and ’\([t]\)’ represents all of the individual T-terms, \( t_1, \ldots, t_n \), that occur in the theory.
As an example, assume there is a theory that, if there is too much electricity running through a wire, the wire will arc.\textsuperscript{13} This theory $T[t]$, has two theory terms: $t_1$, 'electricity', and $t_2$, 'arc'. These two $T$-terms have $O$-terms associated with them:

For $t_1$, 'electricity'

\[ o_1 = \text{needle} \]

\[ o_2 = \text{moving} \]

\[ o_3 = \text{right} \]

For $t_2$, 'arc'

\[ o_4 = \text{light} \]

\[ o_5 = \text{flash} \]

Inputting the theory into $B$ we get:

\[ B(T['electricity', 'arc']) \]

Next, apply the translation rubric to each theory term:

\[ = T[B('electricity'), B('arc')] \]

Finally, translate each theory term into relevant $O$-terms:

\[ = T[\text{needle, moving, right, light, flash}] \]

Thus, the theory terms 'electricity' and 'arc' have been completely replaced by observational terms by way of a bridge law. If desired, Ramsification can be formally completed by adding quantifiers for the (now) eliminated $T$-terms:

\[ = \exists t_1 \exists t_2 T[t_1, t_2, \text{needle, moving, right, light, flash}] \]

Formal Ramsification is not necessarily required as it is the bridge law that clearly demonstrates how seemingly non-meaningful theory terms can be tied to meaningful directly observable terms. By inputting theory terms through a bridge law translation

\textsuperscript{13} Obviously this is not much of a scientific theory, but it will make for an easy example.
rubric, Carnap is able to preserve the verification theory of meaning, yet still allow scientific terms to retain meaningfulness.

**Carnap's Four Criteria for Cognitive Significance**

Carnap's criteria of meaning are his attempt to cash out the verificationist theory of meaning and exclude what the positivists considered to be non-meaningful terms. Given any sentence, S, containing a term, 'a', the sentence can be expressed as 'S(a)'. From this, Carnap proposed four criteria of meaningfulness:

(C1) The *empirical criteria* for 'a' are known.  
(C2) It has been stipulated from what protocol sentences 'S(a)' is *deducible*.  
(C3) The *truth-conditions* for 'S(a)' are fixed.  
(C4) The method of *verification* of 'S(a)' is known.  

In what follows, I will unpack each

**Unpacking C1: Empirical Criteria**

(C1) The empirical criteria for 'a' are known.  

The positivists understood the empirical criteria for a term as a translation of the term into a new sentence only involving observational terms. The use of bridge laws and the process of Ramsification (outlined above) allow for such a translation of all theory. The *O-term* laden sentence that results is what Carnap has in mind with C1. With this *O-term* sentence one can be assured that the observational terms completely and unambiguously specify the empirical criteria of the original term.

---

14 Carnap, "Elimination," 158.
Unpacking C2: Protocol Sentences

(C2) It has been stipulated from what protocol sentences 'S(a)' is deducible.

Of primary importance to the logical positivists is grounding terms in such a way that sentences contain terms that are as epistemically strong as possible. Claims about sense impressions are often taken to be the strongest epistemic claims and the most resilient to many forms of skepticism. For example, I may be mistaken that I am seeing a blue object before me, but given my sense impressions, I cannot be wrong that I am having an experience of blue-ness. Sentences about sense data do not require the positing of any objects external to the perceiver that the skeptic may be able to question. In such claims about sense data, nothing is asserted beyond the existence of experience, thus, leaving these claims limited, but epistemically strong. It is this epistemic strength of sentences about sense data that Carnap hopes to utilize.

Historical empiricists, such as Locke, asserted that basic atomic ideas entered the mind through the sense, and then the mind compiled the atomic notions into complex sensations and ideas. The positivists' notion of a protocol sentence\textsuperscript{15} involves a similar conception. The protocol sentence is meant to express the empirical data provided by sensory perception of the world in its rarified raw form. For example, a red square held at arm's length may be described by a protocol sentence of the following sort: I am now having an impression of a patch of redness, an impression of a tactile smooth texture, an impression of distance, an impression of weight, an impression of firmness, etc. This example demonstrates a difference among the positivists. Although articulating these

\textsuperscript{15} The term 'protocol sentence' was coined by one of the founders of the logical empiricist movement, Moritz Schlick.
features may describe the square, one may be hard pressed to know that it is a square
being described: the sense data are not particularly illuminating. Additionally, if it deals
only with atomic sense data the length of a protocol sentence could be impractically
long. The unease certain positivists had with the viability of appealing to sense data led
some to assert that sentences involving reports of direct observation\(^\text{16}\) were better suited
for their project.

Direct observation involves the sensory experience of objects/events. For
example, when observing a red cube one only sees part of the cube, never the entire
object. Assume that one is looking at the cube in such a way that one can see two faces
of the three-dimensional object. The red faces visible are what are directly observed. We
infer that there are additional faces: a bottom, a top, additional sides, but what is directly
observed are the two faces (the additional faces are also directly observable, just not
currently being observed). The protocol sentence concerning observation of such a cube
would reflect these two directly visible square faces and the red color of each face. This
example again brings to light the distinction between a direct observation protocol
sentence and a sentence involving observations. Within most contexts, one could truly
report that, a red cube is observed. 'A red cube is observed', may be a sentence involving
observation, but it is not a direct observation protocol sentence given that only two faces

\(^{16}\) The term 'direct observation' is, in its self, is problematic, as it is often wondered what
direct observation is. Many observations inherently take place through instrumentation.
If, for example, one observes a moon of Jupiter though a telescope, are the criteria of
direct observation met? Or would that be a case of indirect observation, as the
observation of this sort cannot occur with the naked eye? While an interesting
philosophical problem, it is not necessarily germane to this examination. Keeping in
mind that there are areas of contention, I will limit my examples of direct observation to
those involving naked eye observations, which should not be controversial.
of the cube are directly visible; the remainder of the cube is inferred, and not directly observed. The direct observation protocol sentence only reports what is directly observed.

While many logical positivists accepted that the protocol sentence should be comprised of direct observation, others (such as early Carnap) feared that direct observations were not epistemically strong enough. Direct observations are more susceptible to skepticism than are protocol sentences involving sense data. Although still epistemically strong, the direct observations may be caused by e.g. Descartes's Evil Genius. Where sentences involving only sense data seem to escape such considerations by never referencing objects in an external world, direct observation protocol sentences could potentially come under skeptical attack. The sense data/direct observation protocol sentence debate represents a split in the ranks of logical positivism. This thesis will remain neutral on which types of terms should comprise a protocol sentence. This is, in part, because I will demonstrate problems associated with both direct observation protocol sentences and sense data protocol sentences in the next chapter. My neutrality is also based on the fact that Carnap starts his career pushing for sense data protocol sentences, but, in his later philosophy, reconsiders. Since Carnap is the primary lens through which I wish to view positivism, I see no need to privilege his earlier position over his later position (or vice versa).

**Unpacking C3: Truth-Conditions**

(C3) The *truth-conditions* for 'S(a)' are fixed.
The demand that the truth conditions for \( S(a) \) are fixed would seem to follow from the overall positivist project. The truth-conditions of a sentence must not be conflated with the truth-value of the sentence. The truth-conditions are the conditions that would have to obtain in order for the sentence to be true. The truth-value of the sentence, on the other hand, is the truth (or falsity) of the statement. Thus, if the truth-conditions obtain, then the truth-value of the statement is true. For example, the truth-conditions for a sentence such as 'Shakespeare is the author of *Waverley*' are the conditions that must obtain for the sentence to be true, namely: there must have been a human named 'Shakespeare' and this person must have written a text called *Waverley*. The truth-value of 'Shakespeare is the author of *Waverley*' is false, as the truth-conditions do not obtain (Scott wrote *Waverley* not Shakespeare). Given that to verify something is to establish that its truth conditions obtain, it is essential that a sentence have fixed truth conditions if it is to be verified. Carnap held that without verificationism, one may hold that ambiguous or vague statements had some content.

**Unpacking C4: Verification**

(C4) The method of verification of 'S(a)' is known.

Where C3 establishes that there is fixed conditions for 'S(a)' to be true (or false), C4 emphases how one would determine whether those conditions obtain. The logical positivists demanded that meaningful claims must be capable of being verified. Since the positivists asserted that all propositions were either a priori analytic or a posteriori synthetic, the options of verification were limited (by design). An important aspect of C4
is the emphasis on knowing a method of verification, rather than having a requirement of verification. One must be able to show how 'S(a)' could be verified for 'S(a)' to be meaningful; one need not actually verify 'S(a)'. Thus, 'S(a)' could be meaningful but not verified. There are, for example, many meaningful sentences within the sciences that current technology simply cannot yet verify, but the method by which these sentences could be verified (if the technology were available) is known. Given that the positivists only recognized the analytic a priori and the synthetic a posteriori as potential categories of knowledge, the positivists' methods of verification included logical/mathematical proof and the notion of experimentation so crucial to the empirical sciences.

C1-C4 yield a fairly stringent standard for meaningfulness. As Carnap notes late in "The Elimination" there is no room for terms that have extra-sensory referents within the Circle's metaphysics. It would therefore seem that whole corners of philosophy would likely have to be abandoned including much of metaphysics because Carnap's criteria of meaning possibly could not be met (or could not be met in a way that did not place the concept beyond scientific inquiry). Carnap sums this up in his assessment of the duties of philosophy: "[W]hat, then, is left over for philosophy, if all statements whatever that assert something are of an empirical nature and belong to factual science? What remains is not statements, nor a theory, nor a system, but only a method: the method of logical analysis."[18]

[18] Ibid. 168.
Blame for the eventual decline of logical positivism can be placed on both the requirement for protocol sentences (as dictated by C2) and the positivists' commitment to a strong verification (as outlined by C4). While both C2 and C4 are problematic as individual criteria, the synergy between the two is devastating to the overall project of logical positivism. Although the overlap between the two is at times difficult to unravel, I shall examine C2 and C4 separately, discussing how both stand on philosophically weak ground.

The positivists were also aware of some of the flaws that I will outline—particularly with respect to the notion of verification. The positivists attempted to replace the notion of verification with the notion of confirmation. With that in mind, I will also show that changing C4's focus on verification to confirmation does not salvage Carnap's project.

**The Trouble with C2: Protocol Sentences, T-terms, and O-terms**

The second criterion (C2) involves the notion of a protocol sentence. Recall that the appropriate content of protocol sentences differed among the positivists. Within some positivists' philosophy, such as early Carnap, protocol sentences were comprised of terms that refer to sense data. Within other positivists' philosophy, such as later Carnap, protocol sentences were broadened to include terms that refer to direct observations. The
dismissal of sense data protocol sentences is fairly straightforward. Additionally, most of the problems with direct observation will be a problem for the sense data version as well. Thus, after critiquing sense data protocol sentences, the majority of this examination will consider direct observation protocol sentences.

The positivists initially asserted that all terms could be translated into the language of sense data. Unfortunately, it is not apparent that this is possible. Imagine a common object such as a chair. It is not clear that one could describe a chair using only terms that refer to sense data. The length of such a sentence that would be produced just to capture the experience of the softness of the chair, the color impressions, the act of sitting and experiencing the chair cradling one's weight, etc. is almost beyond comprehension. It is likely one could always add another sense data term to the protocol sentence to more completely or more accurately describe the experience such that the sentence would never be complete, or would be so complex as to obscure the very notion of a chair. If producing a sense data protocol sentence is unlikely or impossible for something as mundane as a chair, it seems even more dubious that such a protocol sentence could be created to describe more complicated notions.

The logical positivists want to export a great deal of philosophy's explanatory project to the natural science. The use of sense data greatly compromises this project. A hallmark of good science (which the logical positivists are committed to believing is also the trait of good metaphysics and epistemology) is public reproducibility. Public reproducibility is the notion that anyone, given the proper equipment, can reproduce any claim and/or result of science. The reliance on terms referring to sense data seems to
violate this basic tenet of science. While observations have the possibility of being public, sense data are always private. For these reasons, most positivists (Carnap among them) abandoned protocol sentences using only the austere language of sense data. Therefore, I shall address the more plausible notion of a direct observation protocol sentence throughout the remainder of the examination.

Many positivists, including Carnap in some of his later works, adopt direct observation as the foundation of the protocol sentence. While sense data relate to what is experienced qua subjective experience, direct observation involves sensory experience of external objects and/or events. Direct observation is typically supplemented by the intellect. Thus, when one sees a cup resting on a table, we infer that the cube has a bottom, an interior space, sides and a back, yet all that is directly observed is the immediate face of the cup. Although we infer a great deal from our direct observations, the positivists contend that we must build protocol sentences from only what is directly observable.

Recall from Chapter 2 that one of Carnap's projects was an attempt to replace all theory terms ($T$-terms) with observation terms ($O$-terms). Theory terms were any term that played a valuable role in a theory but whose references could not be directly observed. By lacking referents that are directly observable, these theory terms run the risk of failing the positivists' criteria of meaning. Carnap attempted to solve this problem by linking all theory to direct observation.

Ideally Carnap wanted a relation of the form: For anything $x$, $x$ satisfies a given theory term $T$ if and only if $x$ satisfies the agreed upon set of observation conditions $\Phi$. 
For example, assume there is some new proposed particle in nuclear physics called the 'pretendon'. Further suppose that positing the pretendon has a great number of theoretical virtues and allows useful predictions to be made, but the pretendon is not directly observable. Under the positivists' criteria, 'pretendon' would be considered a theory term because the object it references cannot be directly observed (a requirement of C2).

Assume there is a machine that can indirectly detect pretendons. Whenever a pretendon enters the machine, a blue light is illuminated. The blue light is directly observable, thus the positivists could freely replace the theoretical particle term with terms relating to the direct observations associated with the pretendon detector, in this case, the blue light:

$$\forall x(\text{Pretendon}(x) \iff \text{Bluelight}(x))$$

Although more reasonable than the reliance on only language of sense data, protocol sentences involving direct observation seems to result in conflating the notions of what is being observed and how it is being observed. Assume there are two groups of scientists: a group of human scientists and similarly advanced group of Martian scientists. The human and Martian scientists have no contact with each other, but have identical mathematics, sense organs, neurological systems, and scientific theories. Both the Martian and human scientists have reason to postulate the existence of the pretendon

---

19 For obvious reasons, we shall forgo the exercise of reducing the illumination of the blue light down to a direct observation protocol sentence level, although a strong reading of the positivists' philosophy would require that it be done.

20 In this contrived example the only observation associated with the pretendon is the illumination of the blue light, but there could have been any number of observations that had to occur for the machine to be said to have detected pretendons. These additional observable requirements could be added as conjunctions. For example, if there must be a blue light and accompanying buzzer for pretendons to be present:

$$\forall x(\text{Pretendon}(x) \iff (\text{Bluelight}(x) \& \text{Buzzer}(x)))$$
using identical axioms, mathematical assumptions, and experimental results. The only difference between the two groups is how their respective pretendon detectors signal detection.

When a pretendon is detected, the human scientists' machine produces a blue light, while the Martian detector sounds a chime. These different detectors would generate completely different direct observation translations of the term 'pretendon':

\[ \forall x (\text{Pretendon}(x) \leftrightarrow \text{Bluelight}(x)) \]

\[ \forall x (\text{Pretendon}(x) \leftrightarrow \text{Chime}(x)) \]

Eliminating the \textit{T-term} 'pretendon' using observation terms from the human detector, results in utilizing the \textit{O-term} 'blue light'. Yet, the Martian detector replaces the \textit{T-term} 'pretendon' with 'chime'. Given that the two theories are more or less identical, it seems that the term 'pretendon' should be translated into the same expression in the language of \textit{O-terms}. Under C2, this does not seem to be the case.

C2 seems to commit the positivists to the belief that the Martian pretendon is a different particle than the human pretendon. Recall that the only difference between the two groups of scientists' notion of pretendons pertains to the machine that detects them. However this generates a huge difference in the protocol sentences. In the human version's protocol sentence, the directly observed phenomenon is light, while in the Martian protocol sentence a chime is directly observed. A human scientist would have no trouble understanding what a Martian pretendon was because functionally the human pretendon is identical. Yet, because of the difference between the light and the chime, the positivists seem to be committed to the result that the Martian and human scientists are
not working with the same theoretical object. The role of direct observation inherent to C2 is too limiting, absurdly resulting in a case were intuitively identical objects must be considered to be of different kinds.\textsuperscript{21}

**The Trouble with C4: Verification**

Carnap's fourth criterion for cognitive significance commits the positivists to verificationism. The positivists defined verification as, the means by which one establishes that the truth conditions for a given sentence obtain. Even without consideration of C2, C4 is problematic. The positivists' demand for verifiability creates difficulties particularly when combined with their stance that the only types of propositions that can exist are the synthetic a posteriori and the analytic a priori. The analytic a priori posses relatively few problems, as proofs of mathematics and logic can appeal to a priori rules to verify universal claims. The problem with verification comes about with respect to claims about the synthetic a posteriori.

One of the most troubling aspects of verification is related to synthetic a posteriori universal claims, such as: all $x$ are $y$. It seems that the only method that could be used to verify that all $x$ are $y$ would be to have access to every $x$ in existence and observe that each $x$ is $y$. This presents two interesting potential scenarios: one in which the number of $x$s is known, and a scenario in which the number of $x$s is unknown.

\textsuperscript{21} Additional problems for C2 are raised by Quine's arguments against the analytic/synthetic distinction. See "Two Dogmas of Empiricism." In the absence of a coherent notion of analyticity/synonym it seems impossible to produce a translation of $T$-terms into the language of direct observation. An examination and response to Quine's argument is beyond the scope of this thesis.
The ideal scenario is one where it is known that there are a finite number of objects to be examined. If one had some prior knowledge of the number of xs that existed, and presumably had access to each x, one could potentially observe each x in an attempt to verify that it is y (although it could be a difficult task). For example, assume x is a particular type of bottle produced at a factory. This particular factory specializes in unique limited edition bottles. If there were meant to be one-hundred bottles of a particular shade of blue, one could go through and check each, thereby verifying that: bottle\(_1\) has the proper hue, bottle\(_2\) has the proper hue, bottle\(_3\) has the proper hue, …, bottle\(_{100}\) has the proper hue. Since, under the above scenario, one is assured there are only one hundred bottles of this sort in existence, one could conclude, having checked them all, that all of the bottles are this particular shade of blue. In such a case, the method of verification would be known: check each bottle's shade.\(^{22}\)

Another scenario is when the number of elements in the set is unknown. Even if there were only one bottle made at the factory, if the individual checking the bottle had no knowledge of the total number of bottles, that individual could never conclude that all bottles were blue, merely that all observed bottles were blue. This type of generalization (where the total number of xs is unknown) is far more likely than the scenario where the total number of xs is known. Verification is problematic because there is no way to assert universal statements of the sort we take to be supported by the types of observations typically made. It may also strike one as intuitively strange to assert that there is no

\(^{22}\) Even with exceptionally large sets, so long as the exact number of members is known, the positivists' demand for verifiability could theoretically be satisfied, but it may be practically impossible to accomplish.
sample size (short of the whole) from which one can draw universal conclusions. Yet C4 commits the positivists to the position that where the number of xs is unknown, no amount of observations can verify a universal claim.

**Confirmation: A Failed Replacement for Verification**

With the role of verification in C4 leading to unacceptable, non-intuitive results, the positivists instead embraced the notion of confirmation. Rather than providing decisive evidence for a claim (the goal of verification), confirmation merely strengthens belief in a claim. The notion of confirmation was outlined by Hempel.²³ He held that in the case of a statement that all A type things are B type things, every instance of an empirical observation of an A type thing being a B type thing, strengthens the universal claim, and every instance of an A type thing not being a B type thing weakens the belief in the universal claim. While confirmation cannot prove something to be the case, confirmation can build a very strong case for believing a universal claim.

The move to replace verification with confirmation seems at times a strange one. As I shall demonstrate, confirmation can better handle justifying belief in generalizations, but it does not seem suited to form the basis of a theory of meaning. As we saw above, under Carnap's criteria with verification, the vast majority of universal empirical claims would have to be thrown out as meaningless since they would be unverifiable.

---

Confirmation, being weaker than verification, makes it seem at least plausible that many empirical propositions would be confirmable.

Unlike verification, confirmation of a universal generalization can occur even when the number of objects in the subject class is unknown. If one were to see a black raven it would be hasty to conclude that all ravens are black. Yet if, no matter where the observer went, when she saw a raven it was black, each new sighting of a black raven would strengthen (confirm) the belief that all ravens were black. This is in accordance with many statistical theories and also our natural intuitions about sample sizes. The more representative of the larger population the sample, the better the generalizations that can be made. Although we can never verify that all ravens are black, the number of sightings strengthens belief in the likelihood that all ravens are black. Thus, confirmation seems to better handle generalizations than verification.

For all that the move to confirmation solved, it introduced serious philosophical problems as well. The French logician and philosopher Jean Nicod's work was used by the logical positivists, particularly Hempel, when discussing confirmation. Nicod's criterion states that, a hypothesis of the form "A entails B" gains support (is confirmed) whenever an A state of affairs is followed by a B state of affairs. The hypothesis of the form "A entails B" is weakened (disconfirmed) if an A state of affairs occurs but a B state of affairs does not follow. If one discovers an x that is both an A and a B, then Nicod's criterion states that it gains support. The size of the sample relative to the population is relatively unimportant, so long as the sample is relatively randomly generated. For example, a randomly generated sample of 2,000 potentially yields a conclusion with a set confidence interval and margin of error whether it is taken from a population of 5,000 or 50 million.
criterion states we should believe more firmly in the assertion. Yet, if one discovers an x that is an A but not a B, then Nicod's criterion states our believe in the hypothesis should be weakened or the hypothesis has been outright refuted.  

Returning to the black raven example above, the hypothesis would be that all ravens are black. Each black raven would strengthen belief in the universal claim. If the budding ornithologist were to begin spotting green ravens, then the belief that all ravens are black would be weakened. Nicod's criteria are meant to capture the tug of war between data and belief: confirming data instills and strengthens belief, and disconfirming data results in doubt. Unfortunately, Nicod's criterion, Hempel asserts, leads to the paradox of confirmation. The theory that all ravens are black is translatable into logic as:

\[ \forall x (R_x \rightarrow B_x) \]

---

26 To simply assert that an instance of an A that is not also a B refutes the hypothesis that all A are B is too strong. There may be times when observations are not ideal, and where things that are merely assumed to be A type things (but may not be) are observed to not also be B type things. For example, perhaps a partial fossil is discovered and believed to be a certain species, but also has properties that are not associated with that species. This may cause us to doubt certain assumptions, but not abandon them outright. Additionally, certain long held beliefs may require more than a single/few instance of disconfirmation to refute the believed hypothesis. For example, at the time of this writing, there is some experimental evidence that neutrinos can travel faster than the speed of light, but this data does not instantly cause the scientific community to abandon certain assumptions of relativity.

27 It should be noted that the issue here is not entailment in the sense of logical entailment. The concept being examined is much closer to nomological entailment of the sort captured in more contemporary modal logic. No rule of logical entailment demands that all ravens are black. Yet, one may be inclined (as Hempel seems to be) in indicating that a necessary condition of being a raven is being black. Thus, Hempel is really talking about nomological entailment. Taking these concerns into account, it would, perhaps be more accurate to note that Hempel is really concerned with the statement, in all possible
Using contraposition, (1) can be converted into the logically equivalent statement, 'all non-black things are non-ravens':

(2) ∀x(~Bx→~Rx)

The paradox begins by examining what would logically satisfy sentences (1) and (2) and comparing that to intuitions about what would confirm sentences (1) and (2). The only state of affairs that would fail to logically satisfy (1) is if an object, x, were a raven but not black. Confirmation is a different claim than logical satisfaction as confirmation is epistemic where logical satisfaction is purely semantic. To illuminate the paradox, let us take two objects: a black raven and a green pencil. In the case of a black raven, (1) is logically satisfied: it is the case that x is a raven and x is black. A black raven also confirms (1), as we have come into contact with an instance of a raven that is black, giving a little more credibility to the probability that all ravens are black. Now let us take a black raven and statement (2): all non-black things are non-ravens ((∀x(~Bx→~Rx)). A black raven logically satisfies (2). Yet, intuitively, a black raven does not confirm (2). Statement (2), ∀x(~Bx→~Rx), is a statement about all objects that are not black. It seems intuitively strange to hold that a fact about a black object strengthens our beliefs about objects that are not black. Therefore, although a black raven is able to logically satisfy (2), it does not seem to confirm (2). However, if Nicod's criteria of confirmation are correct, and the positivists assert that it ultimately is, then the black raven should confirm (2). This, in itself, is troubling, but the situation worsens when considering our other object, a green pencil.

worlds with a sufficiently close accessibility relation, if x is a raven, then x is black, formalized: □∀x(Rx→Bx)
If $x$ is allowed to be a green pencil, (2) is logically satisfied. Additionally, a green pencil intuitively provides some confirmation to (2). Statement (2) is about non-black things; our pencil is a non-black thing and the pencil is not a raven. When $x$ is allowed to be a green pencil, (1) is logically satisfied. However, a green pencil does not seem to confirm (1). Statement (1) is about all ravens being black; so it does not seem that a non-raven object should strengthen our beliefs about ravens.$^{28}$

A black raven and a green pencil can logically satisfy both (1) and (2), these objects cannot intuitively confirm both (1) and (2). The paradox arises because the positivists desire two equivalence conditions.$^{29}$ The first desired equivalence condition is between logical satisfaction and confirmation, and the second is between confirmation of logically equivalent statements. The first equivalence condition is particularly important to the logical positivists. The positivists desired any condition that logically satisfied a given statement $\phi$, to also confirm $\phi$. Additionally, any condition that does not logically satisfy $\phi$, should also disconfirms $\phi$.

The second equivalence condition is quite plausible. Intuitively, it seems that we would like it to be the case that any condition that confirms a statement would also confirm all logically equivalent statements. Thus, if some condition, $p$, confirms $\phi$, and $\phi$ is logically equivalent to $\psi$, then it would be advantageous if $p$ also confirmed $\psi$.

---

$^{28}$ I find the problems associated with the green pencil more problematic than I do the problems associated with the black raven. This is due to 'raven' being a natural kind term. I tend not to think about non-black things as a natural kind whereas black things can, at least prima facie plausibly, be taken to be a natural kind. If non-black things are taken to be a natural kind or if the notion of a natural kind is not ultimately coherent, then the green pencil example gives rise to the same problem as the black raven.

$^{29}$ Hempel, "Logic of Confirmation," 39.
Similarly, if condition \( p \) disconfirms \( \phi \), and \( \phi \) is logically equivalent to \( \psi \), then intuitively it seems that \( p \) should also disconfirm \( \psi \). Unfortunately the paradox of confirmation suggest that these two equivalence conditions cannot be jointly satisfied. In order to resolve the paradox, most would conclude it is the first that must be abandoned.

The solution to this paradox proposed by Hempel and embraced by Carnap is that our intuitions about these paradoxes are wrong and that the two equivalence conditions actually hold. Carnap presents his case that we should abandon our intuitions about confirmation that drives the paradox by adopting a Humean theory of causation, according to which metaphysical causation does not exist.\(^{30}\) The abandonment of any form of metaphysical causation is too much for many philosophers. Many will hold that Carnap fails in his attempts to dissolve the paradoxes of confirmation. This leaves the theory of confirmation seemingly no better off than the theory of verification.

**Where Does this Leave Carnap's Four Criteria?**

As I transition into the Canberra Plan, let me first take a moment to re-examine Carnap's four criteria for meaning:

- (C1) The *empirical criteria* for 'a' are known.
- (C2) It has been stipulated from what protocol sentences 'S(a)' is *deducible*.
- (C3) The *truth-conditions* for 'S(a)' are fixed.
- (C4) The method of *verification* of 'S(a)' is known.

---

In this chapter, I outlined reasons why logical positivism failed. Fault for the collapse, I have argued, can largely be placed on criteria C2 and C4; leaving C1 and C3 in relatively good standing. I shall, in time, propose replacements for C2 and C4 by way of the Canberra Two-Step.
CHAPTER IV

THE CANBERRA PLAN

The concerns giving rise to the fall of logical positivism left the relationship between analysis and naturalistic philosophy unclear. For a time, it seemed as though naturalism and analysis were simply meant to go their separate ways. Yet, not all analytic philosophers saw the separation as permanent. Believing that meaningful analysis could still be accomplished and play a valuable role—particularly within naturalism—practitioners of the Canberra Plan forward a potential solution to the problems that weakened the positivist's project and help restore the role of analysis.31

As of the writing of this thesis, the Canberra Plan is only beginning to emerge as a philosophical school of thought and methodology. Although many of the techniques and philosophical intuitions inherent in the Canberra Plan are borrowed from earlier analytic traditions, the exact nature of the Canberra Plan is only just being codified in contemporary philosophy, and many of its consequences just beginning to be explored. Where the Vienna Circle allowed the early positivists to advance a fairly unified and well-formed version of logical positivism, practitioners of the Canberra Plan are not as monolithic. Nevertheless, philosophers that practice the Canberra Plan, often called 'planners', present a compelling and cogent philosophy.

The Two Step: Goals & Explication

The goal of the Canberra Plan is to analyze a group of related concepts which, when taken together, form what is referred to by planners as a folk theory. Planners use "folk" to refer to a collection of specified individuals or a community of speakers, rather than in the sense of the philosophically unsophisticated. The folk theory being analyzed could be highly technical in nature (if the community were a group of, say, astrophysicists) or more mundane (for example a community trying to discover the best hamburger). The analysis of the folk theory is accomplished, broadly speaking, by the "Canberra two-step".  

The first step of analysis consists of gathering the various ideas, notions, and putative truths (which are all lumped together under the blanket term, 'platitudes') surrounding the topic of inquiry. With the relevant platitudes gathered, the concepts contained therein are analyzed. Analysis takes a different form for various planners depending in part on their end goals (obviously the type of platitudes and analysis surrounding the concept of an undiscovered subatomic particle will be different than those related to the concept of happiness). A restriction placed upon analysis is that the platitudes are to be understood in such a way that there is some confidence that, in the next stage of the two-step, a suitable satisfier of the concept can be found. Once analyzed, this analysis is handed off to an appropriate group of empirical researchers so that the item being described can be "located" in the world.

---

The goal of location is to find an actual or near realizer of the concept analyzed in step-one. The nature of how a term is or might be realized often depends on the nature of the concept that is analyzed. I will discuss both steps in greater detail in the following sections. For example, if the notion of a theoretical particle in physics were what had been analyzed, then the location step would likely involve attempting to isolate the particle in a lab. On the other hand, if the notion being analyzed is a concept such as happiness, then location may take the form of discovering a suitable realizer in the brain or central nervous system. In this way, location is context dependent.

**Step-One: Platitude Gathering & Analysis**

As I shall demonstrate, the planners conception of analysis is fairly open ended, so, for the time being, I will allow the term to remain loosely defined. I will focus first on platitude gathering, and then tackle analysis. It is important to realize that there is not a hard order to these two elements. Obviously one must gather platitudes before there is anything to be analyzed, but step-one should not be understood as mandating that the platitudes are *first* gathered and *then* analyzed. There is not a strict linear relationship between gathering and analysis. Often it is necessary to return to the gathering of additional platitudes during the analysis, or to accomplish some preliminary analysis while in the initial stages of gathering.

The gathering of the platitudes could be very restrictive or very inclusive (as Nolan puts it, "kitchen sink" in nature).\(^{33}\) For example, suppose a planner were interested

---

\(^{33}\) Ibid. 275.
in understanding the concept of happiness. An inclusive gathering of platitudes concerning happiness would include notions from a wide range of sources: common usage of related terms, scientific, historical, philosophical sources, etc. A more restrictive gathering might only consider platitudes relating to neurophysiologic aspects of happiness. Both the inclusive and restrictive approaches to platitude gathering have strengths and weaknesses. One consideration when deciding on the appropriate level of restrictiveness is a reflection on the typical usage of the terms relating to the notion being examined. For example, if the notion being examined were an economics folk theory, then it is likely that the planner would restrict the platitudes to those relating to other economic notions. At present, the level of restriction applied to platitude gathering is left to the individual planner’s discretion. In this way the Canberra planners, unlike the logical positivists, can cast both a very fine and very broad net when needed.

An important aspect of all stages of the Canberra method is the weighing of the platitudes against each other. During the course of the gathering of the platitudes (or later during analysis or in the location phase), it may become apparent that platitudes are in tension with each other, or even inconsistent. For example, if a planner were to take an inclusive approach to gathering platitudes about the notion of happiness, it is possible that the planner would identify as a platitude that happiness is a subjective feeling that may transcend the physiological. Since the planner is taking an inclusive approach to gathering, the planner may also identify another platitude that happiness is merely a physiological/neurological event. This seems to reveal an inconsistency. In one platitude

34 Ibid. 274-278.
happiness may transcend the physiological, in the other platitude happiness is necessarily physiological. In such a case, the planner may weigh the two platitudes, deciding which is more valuable to the theory and, thus, deserving of the greatest credence. The weighing process may include a ranking of the platitudes in order of confidence in their truth, or the process may involve dismissing platitudes that are responsible for the apparent inconsistencies. It is up to the planner to determine which platitudes to retain, dismiss, and value when there is an inconsistency or internal tension.

Due to the nature of the Canberra Plan, conflicts in platitudes may arise during both the gathering and analysis phases. This is the case because, like the positivists, planners tend to hold that sentences can often be unpacked into more distinct claims. Just as Russell's famous examination of the sentence "The present king of France is bald" could be unpacked into three distinct claims,\(^{35}\) the notions that the planners are seeking to understand may unravel into component claims as well. Once all the platitudes have been unpacked through analysis, it may be the case that inconsistencies or tensions are revealed. Just as was the case with an overly inclusive gathering, it is up to the planner to weigh the platitudes and decide which to keep, which to discard, which to give the greatest epistemic weight, etc.

Where the positivists had a rigid notion of analysis, the planners are less committed to any one technique. Planners typically point out that the Plan ascribes to no

\(^{35}\) (1) There is an x such that x is a present King of France: \(\exists xFx\). (2) For every x that is a present King of France and every y that is a present King of France, x is y: \(\forall x\forall y(Fx \rightarrow (Fy \rightarrow x=y))\). (3) For every x that is a present King of France, x is bald: \(\forall x(Fx \rightarrow Bx)\)
singular methodology vis-à-vis logical analysis. That said, the process of analysis for the planners typically takes the form of "armchair" a priori philosophy—examining intuitions about possible cases. Planners typically rely on the types of thought experiments involving possible worlds and Twin Earth experiments made famous by Kripke and Putnam. Twin Earth thought experiments posit another planet that is typically identical in every way to Earth except for one key feature—the feature being analyzed. For example, Earth and Twin Earth are identical with respect to the total functional role played by what the inhabitants of each world call 'water' (that water is sustainer of life, located in ponds and streams, falls in precipitation, etc.). On Earth this role is played by H$_2$O; on Twin Earth the role is played by XYZ. As such, Earth inhabitants use the term 'water' to refer H$_2$O, while on Twin Earth inhabitants use the term 'water' to refer to XYZ. The thought experiment concerns whether or not XYZ is water given that it plays a functionally identical role on Twin Earth as H$_2$O plays on Earth. In this way, the concept being analyzed (in this case that being denoted by the term 'water'), must stand against the trial of possible cases with each version of Twin Earth further restricting some feature or another in hopes of discovering the essential features of the usage of the term.

Planners approach the gathering and analysis of platitudes in a similar manner. Possible situations are considered and examined. Thought experiments, typically of a Twin Earth nature, act as a way to generate the original platitudes and/or to sort the

---

36 See Kripke's *Naming and Necessity* and Putnam's "The Meaning of 'Meaning'." Although the thought experiments that both Kripke and Putnam employ are often adopted by planners, planners do not necessarily, nor even typically, subscribe to the semantic theories adopted by Kripke or Putnam.

37 Putnam took the thought experiment to demonstrate that XYZ was not water.
platitudes should there be internal conflict. A planner asks himself or herself what types of scenarios would make the platitudes true/false, more or less relevant, etc. For example, suppose a planner is attempting to analyze the concept water and is inclined to view the world through chemistry. It is likely that such a planner would adopt the platitude that water is composed of base elements in a certain proportion and bonding arrangement: $\text{H}_2\text{O}$. This planner may devise a Twin Earth experiment where $\text{XYZ}$ behaves identically to $\text{H}_2\text{O}$ in all chemical reactions. This will present an interesting problem for this planner, compositionally $\text{XYZ}$ is not $\text{H}_2\text{O}$, but the two compounds' chemical reactions are identical. The planner will have to ask herself which matters more: the chemical behavior of the substance or the atomic composition of the substance? The answer to that question will help the planner better refine the initial platitude set.

As the Canberra Plan has begun to take shape as a coherent philosophical technique, two competing branches of step-one have emerged. While the end goal of the two techniques remains the same (construct a descriptive sentence or set plausibly having a realizer that can be located in the world), the two approaches differ in the form of the descriptive sentence. The first approach was taken by David Lewis and seems to be a direct successor to the notion of Ramsification used by Carnap. The other approach is taken by Frank Jackson. I shall call these two versions of step-one the Lewis and Jackson Traditions respectively.
The Lewis Tradition

David Lewis sees himself working within the Ramsey-Carnap tradition. As such, Lewis follows a process similar to that of Carnap. In his article, "How to Define Theoretical Terms," Lewis provides an extensive outline of the end goal of his version of step-one of the Canberra Plan which ends with Ramsification of the list the platitudes. Recall from Chapter 2 that Carnap adopted and slightly modified Ramsey's approach to the elimination of theory terms. In doing so, Carnap hoped that any theory term (T-term) could be replaced by a Ramsified sentence involving only direct observation terms and predicates (O-terms). This fits the logical empiricist's overall project of grounding all concepts in (some form of) observation. Lewis' construction of O-terms and T-terms is far more flexible than Carnap’s hard-line direct observation/sense data conception. Where Carnap hoped to eliminate non-directly observable terms with directly observable terms, Lewis is hoping to eliminate vague/contentious terms with well-understood predicates.

Like Ramsey and Carnap, Lewis is fundamentally interested in terms corresponding to key concepts in the folk theory to be analyzed. Carnap's reliance on bridge laws and need to reduce O-terms to the level of protocol sentence (be it the level of direct observation or the even more austere level of sense data) greatly limited the use of Carnap's technique. Lewis saw this as a failing of Carnap's approach. To avoid a similar mistake, Lewis views O-terms differently than does Carnap. Where Carnap

---

39 Ibid. 78-95.
specifically attempted to ground $O$-terms in the realm of sense data/direct observation, Lewis asserts that the $O$-terms need only be: "[A]ny other term, one of our original terms an old term we already [understand] before the new theory $T$ with its new $T$-terms…". $^{41}$ Under this conception, $O$-terms can take the form of any previously understood notion that can be represented by a predicate.

Due to the nature of the Lewis tradition, the gathering phase is guided by picking out $O$-terms that stand in some relation to the $T$-terms that are to be eliminated. Analysis, for Lewis, takes the form of Ramsification. For example, given a theory, $T$, the theory postulate '$T[t]$' is a single conjunctive sentence that says each of the $T$-terms $t_1,t_2…t_n$ that it occupies a certain functional role expressible entirely in $O$-terms. The goal of the Lewis approach is to replace the $T$-terms in the postulates with variables $x_1,x_2…x_n$. Once accomplished, the resulting open sentence, '$T[x_1,x_2…x_n]$' only contains $O$-terms. The $T$-terms have been eliminated. The sentence is then Ramsified in order to obtain the generalized statement:

(1) $\exists x T[x]$ $^{42}$

Within Carnap's work, Carnap allows that there are any number of n-tuples that could satisfy the Ramsified sentence. This reflects the stance that theory terms have no definite meaning. Under such a view, one need not be committed to one realizer of the theory term over any other because, so long as the sentence is satisfied, then the theory term has accomplished what it was created for. The positivists, not caring about the status of $T$-

$^{41}$ Lewis, Papers, 79.
$^{42}$ For simplicity, I shall address a theory with a single T-term. Where there are n T-terms, there will be n existential quantifiers governing distinct variables $x_1-x_n$. Thus, given two T-terms the goal would be: $\exists x_1 \exists x_2 T[x_1, x_2]$. 
terms, did not worry that multiple realizers might satisfy the Carnap-Ramsey sentence. This is another area where Lewis parts ways with Carnap, for, unlike Carnap, Lewis asserts that there must be a unique satisfier of the Ramsified sentence.\textsuperscript{43}

\[\exists_1 x \Phi[x] \equiv \exists x (\Phi[x] \& \forall y (\Phi[y] \rightarrow x=y))\]

Lewis, hoping that the Plan can actually be used as an engine of discovery and mindful of the second location step, requires that the sentence be satisfied by one and only one realizer. If multiple realizers exist, then the analysis is a failure. Continuing with the outline of Ramsification, a Lewis-approach produces the conditional meaning postulate for T:

\[\exists_1 x T[x] \rightarrow T[t]\]

If T is uniquely realized, the T-terms name the components of the realization. Since 'T[x]' contains only O-terms, this allows an explicit definition of T-terms by O-terms.

\[t = \text{the unique } x \text{ such that } T[x].\]

Thus, the analysis is complete when sentence (3) entails sentences of form (4). Thus, given a theory with its T-terms, a unique set of O-terms can be used to define each of the T-terms. Once the Ramsified sentence (3) is constructed, the sentence is handed off to an empirically minded set of experts for potential location.

The Lewis approach is a powerful change in the nature of the Ramsey-Carnap sentence. Carnap's conception allowed theory talk to be eliminated as desired, but Lewis' technique allows for discovery. Carnap's version worked as an explanation of theory terms to provide them cognitive significance (under the strict definition of the positivists).

\textsuperscript{43} Lewis, Papers, 88-90.
where Lewis' notion provides not only definition, but can also point scientists in the right direction of an actual search for a realizer. For example, assume that a new baryon is being proposed in particle physics. Both Carnap and Lewis would begin the process of Ramsification by gathering *O*-terms. In the case of Carnap, *O*-terms would be related to sense data/direct observation. It would not be sufficient merely to say what types of quarks compose the baryon. Lewis, on the other hand, could gather *O*-terms that are already understood and part of a particle physicist's vocabulary. On the Lewis-approach, the predicates involving *O*-terms would likely express what types of quarks that comprise the new theoretical baryon. Thus, on the Lewis approach, the theoretical baryon could be described in terms of its composite quarks, because 'quark' is an understood term in theoretical physics. It is an "old" term and the planner need dig no farther. By defining the particle in terms of known properties (quantum numbers, mass, type of reactions it interacts in, etc.) an experimental rubric could be created to know exactly what the scientists were looking for and how to go about looking for it. Rather than simply move from observation to observation (as in Carnap's conception of Ramsification), Lewis hopes to create an engine of discovery by providing an account of ways in which the referents of analyzed terms might be identified on the basis of empirical investigation.

---

44 Baryons are elementary particles that are, by definition, comprised of three quarks. The ratio and types of quark determines the type of baryon. The two best-known baryons are the proton (two up quarks and one down quark) and the neutron (one up quark and two down quarks).
The Jackson Tradition

While it is clear that Lewis saw the Plan as continuing Carnap's work into the philosophy of science and strongly theory based endeavors, other planners, such as Frank Jackson, see the Plan as working for nearly all potential inquiries. Indeed, Jackson's view on the breadth of problems the Plan could tackle is summed up nicely by the title of his book, *From Metaphysics to Ethics*. Unlike Lewis, who focused on the axioms and logical systems of first order theories, Jackson focuses on *concepts*. Shifting the focus allows Jackson to avoid the formal process of Ramsification and instead work with natural language platitudes that articulate the concept. As such, Jackson's notion of analysis does not involve the translation into a formal first order theory that the Lewis' tradition requires. While Jackson's approach may not yield a reduction of *T*-terms to *O*-term, Jackson's approach nevertheless is reductive in nature.

Jackson's treatment of reductionism is apparent in his ambitious conception of the Plan. Jackson sees the goal of what he calls "serious metaphysics" as "[Acknowledging] that we can do better than draw up big lists, [and instead] seek comprehension in terms of a more or less limited number of ingredients, or anyway a smaller list than we started with." The goal of Jackson's reductionism is to create a smaller and smaller list of objects, properties, and relations that can account for the physical world, with no feature of the physical world standing over and above everything we observe in the small list of objects, properties, and relations. Jackson is unsure whether such large-scale reduction is

---


46 Where a first order theory is a set of axioms expressible in first order predicate logic.

possible, but it is clear that Jackson asserts that philosophy should actively pursue such a goal.

An example of Jackson's treatment of the Plan is the folk theory of density. Jackson notes that some things have mass, some things have volume, and some things have density. Density is clearly a different concept than mass and/or volume, but, at the same time, density is closely related to mass and volume. Reviewing relevant platitudes connected to each, Jackson concludes that, "Density is not a feature of reality over and above mass and volume." From this, we get a clear sense of what Jackson is doing with his conception of the Plan. Through analysis, one identifies the set of platitudes associated with a given concept. If such a set of platitudes can be identified, then a reduction to the items referred to in the platitudes occurs. In this example, the folk theory of density has platitudes expressing facts about mass and volume. Thus, the concept of density requires nothing over and above what is required by the concepts of mass and volume.

Jackson stresses the need for the conceptual analysis during step-one to be modest. A modest analysis consists of a purely conceptual sorting process. Platitudes are gathered and ranked, resulting in a weighted list of platitudes and an identification of the relationship between the platitudes. Jackson's modest analysis flushes out relations (either between platitudes and other platitudes or between concepts and platitudes). Whereas Jackson's notion of immodest conceptual analysis purports to make claims about the world as it is. Modest conceptual analysis can rightly conclude that proposition $\Phi$ does

---

not accord with the folk theory. Immodest conceptual analysis, on the other hand, would conclude that $\Phi$ is false.\textsuperscript{49}

Where the Lewis tradition takes a concept and describes its functional role using only old terms in constructing the theory postulate, the nature of Jackson's reductive approach seems to require a greater commitment to weighing the initial platitudes. While it is certainly the case that some weighing goes on in the Lewis tradition (particularly if no realizer is found in step-two), Jackson's tradition inherently involves the planner asking the question "what is the minimal set of platitudes that this concept can be reduced to?" The emphasis on the minimal set is absent from the Lewis tradition. Although the Lewis tradition requires some notion of translation in order to generate sentence (1), this translation need not be austere: there may be many $O$-terms that are extraneous or duplications in relation to sentence (4) defining a given $T$-term.

These distinctions spark a great deal of internal debate within the planner community, with compelling arguments for and against both traditions. Although my preference is for the Lewis tradition, both versions of the Plan seem viable. Despite the differences and nuances between the Jackson and Lewis approaches, the end goal is the same. Jackson's platitudes play a conceptually similar role to Lewis' sentence (4) involving $O$-terms. The end goal of both Jackson's and Lewis' interpretations remains the generation of a descriptive sentence (or set there of) of some folk theory that can be handed off in an attempt to locate a unique realizer in the world.

\textsuperscript{49} Ibid. 43-44.
**Step-Two: Location**

If the planner has done her job properly, the folk concept or theory being examined can now be completely described either as a Ramsified *O-term* sentence (in the Lewis tradition), or as a set of related platitudes (in the Jackson tradition). The analyzed concept can be now handed off to the appropriate department to attempt to identify a unique realizer. Who does the locating is dependent on both the type of theory that the planner has analyzed and the method and details of the analysis. If, for example, a planner of the Lewis tradition carried out analysis to define some as of yet undetected particle in nuclear physics, a completely Ramsified sentence containing a description of the potential particle's properties could be given to nuclear physicists to search for something of that sort. If, on the other hand, a planner of the Jackson tradition analyzed the concept of beauty, then the ranked platitude list might be given to aestheticians\(^5\) to locate beauty in the world. Once the descriptive sentence is handed off, the second step begins.

The details of the location phase are left to the individual set of relevant experts. In the case of an analyzed concept being handed off to a psychology department, brain scans may be done to locate a realizer of the analyzed concept, or the same psychology department might instead conduct testing on animal subjects. The nature of how location is accomplished not only varies based on the type of concept being analyzed but also on the results of the analysis. The realizer of an aesthetic concept is not likely to be located

---

\(^5\) This demonstrates the flexibility of the Plan given the nature of the analysis. If the planner analyzed beauty, one might give the platitude list over to a geometrist, architect, psychologist, etc. It all depends on the nature of the platitudes gathered.
in the same manner as a nucleotide sequence. The strength of the Plan remains the flexibility of both analysis and location.

The literature produced by the planners on step-two is fairly sparse. This is likely in part due to the role the planners see philosophy playing in the academy. Canberra planners, like the positivists, see philosophy playing an integral role in collaborating with other fields of study. Where the positivists only saw philosophers able to play a very limited role as logicians, the planners have expanded the role of the philosopher. In both logical positivism and the Canberra Plan it seems that philosophy is meant to do the analysis while other departments are meant to accomplish the more empirical research. I will return to an earlier quote by Carnap: "[W]hat, then, is left over for philosophy, if all statements whatever that assert something are of an empirical nature and belong to factual science? What remains is not statements, nor a theory, nor a system, but only a method: the method of logical analysis." Although planners likely do not agree with the details, it seems most would concur with the spirit of Carnap's sentiment. With the completion of both stages of the Canberra Plan, philosophy accomplishes its method: not as science's overlord or handmaiden, but as a specialized equal.

Near-Realizers

An interesting problem arises in the second step of the Canberra Method. Both treatments of step-one result in a descriptive sentence (or set there of) to be used in the location phase. Ideally, a unique realizer is discovered that perfectly satisfies the

---

51 Carnap, "Elimination," 168.
descriptive sentence. If no realizer is discovered, then either analysis is a failure on the Lewis view and in doubt on the Jackson view, or we must acknowledge the falsehood of the folk theory. If multiple perfect realizers are identified in step-two, then a choice must be made whether to consider analysis a failure or acknowledge that there is a disjunctive satisfier. The former choice seems required on the Lewis tradition; the Jackson tradition seems to allow for greater flexibility in potentially allowing for disjunctive satisfiers.

What remains unclear, on both views, is how to address situations where no perfect realizer is identified but a near-realizer is located. Both traditions are silent on the treatment of near-realizers. For example, what if, instead of locating a thing with features A, B, C, and D as indicated by the step-one analysis, the located thing only has features A, B, and D. Is this enough to satisfy the location step? Unfortunately, there has been very little literature on this topic. This may be due in part to the relatively young age of the Canberra Plan. The Plan has not been put into practice a sufficient number of times to draw the needed conclusions about near-realizers.

Jackson's treatment may initially seem more susceptible to the near-realizer problem. Since the goal of the Jackson tradition is to establish a list of minimal plitudes, if any one of those minimal elements is missing, it seems that either the analysis has failed (the list of plitudes is not minimal) or the location has failed (the platitude list has not actually resulted in the location of a realizer). However, the notion of platitude weight can play a role in solving the problem (for the Jackson approach). If each platitude in the list generated during step-one was weighted, the difficulties of near-realization would be reduced. If plitudes A and B were given a great deal of weight and
C and D were given little weight, it may not be a problem if the location phase could only find an object satisfying platitudes A, B and D.

There is some debate among planners just how close a realizer needs to be found to count as a successful use of the Plan's methodology, but one can, to some extent, understand the motivation behind the planners producing so little literature on this topic. The point of step-one is to hand off analyzed folk notions for location by another group. Different folk groups may have different intuitions about what properties are most important, or how close the potential realizer needs to be. The level of precision demanded by a researcher in the physical sciences may be too stringent for a researcher in the behavioral sciences; conversely, the standards in the behavioral sciences may be too imprecise for the physical sciences. It could be viewed as inappropriate for philosophy to attempt to create a blanket rule for how close is close enough, particularly given the vision of the Plan. Just as there is very little literature on how the second-step is to proceed (presumably because such methodologies are left to the groups receiving the analysis from step one), so too it may be with near realization. The issues of near-realization may be, quite literally, another group of expert's problem.

The Unofficial Third Step

There is an unofficial third-step in the Plan. If either of the two steps seems to fail—if the analysis comes up incomplete, or after handing the results off to experimentalists a unique (near) realizer is not located—then the platitudes must be examined anew. For example, the analysis may fail because a defined list of platitudes
may not be generated. Returning to the chemistry minded planner and the \( \text{H}_2\text{O}/\text{XYZ} \) problem. Recall that the planner devised a thought experiment in which XYZ behaved in chemical reactions identically with \( \text{H}_2\text{O} \) while having a distinct atomic composition. The problem for the planner is whether the behavior of the chemical or the composition is what matters. Suppose that this planner simply cannot decide on this issue. This would be a case of a failed analysis. An example of a failed location step might be that a descriptive sentence is created but multiple realizers are found. I have often used technical examples, so I wish to here stress the flexibility of the Plan. Suppose, for example, a planner were trying to pick the best car for her needs, she completes the analysis and generates the descriptive sentence of her ideal car, only to find that, once on the car lot, multiple cars realize her requirements. In any application of the Plan, grandiose or mundane, the process of both examining and subsequent revisiting of the platiitudes will likely not be a clean one: the weight given to platiitudes, the amount of change to the platiitudes that is acceptable without changing the subject, etc. are all factors that are not easy to provide blanket guides to and potentially depend on the empirical facts on the ground.

The unofficial third-step also reveals an important strength of the Plan. Where many philosophies purport to be fully formed with sweeping explanatory power of the world as is, the Plan, by focusing on a process of analysis and the empirical search for realizers, acts as an evolving description of the world rather than a stagnant set of conclusions. This allows the results that the Plan generates to be adapted to shifting empirical conditions or theoretical developments. By reexamining platiitudes, attempting
location anew, or simply learning from failed attempts to accomplish either, the Plan, pushes knowledge forward even when confronted with apparently false analyses or failed attempts at location.
CHAPTER V

PROBLEMS WITH THE PLAN

The Canberra Plan offers an interesting and powerful approach to doing philosophy, yet the methodology is not without faults. In this chapter, I will explicate a problem concerning a tension that can arise between step-one and step-two of the Canberra Plan raised by David Braddon-Mitchell in his "Naturalistic Analysis and the A Priori."  

Braddon-Mitchell's Critique: Two Sorts of Dispositions

In his article, "Naturalistic Analysis and the A Priori," Braddon-Mitchell explicates and attempts to solve a potentially deeply troubling problem of the Canberra Plan. The creation of a descriptive sentence in step-one involves the identification and weighing of platitudes. The process of weighing the platitudes may take the form of trial by possible cases (such as the Twin Earth-like thought experiments outlined in the previous chapter). These experiments are meant to reveal or predict how the planner would be disposed to apply a given concept in various hypothetical situations. The dispositions related to how the concept would be applied in these hypothetical situations are labeled by Braddon-Mitchell the theoretical dispositions. However, there may be

---

53 Ibid. 23-43.
54 Although Braddon-Mitchell is clearly working within a more Jacksonian tradition and interpretation of the Plan, it is not inconceivable that similar problems may arise in a Lewis treatment.
times when we find ourselves in a situation in which we are able to see how we actually apply a concept in a given cases. Such cases reflect what Braddon-Mitchell labels the real dispositions. Braddon-Mitchell notes that there is a tension between these theoretical dispositions (those revealed a priori via thought experiments in step-one) and the real dispositions (those revealed a posteriori in actual experience). Where the theoretical judgment and a real judgment diverge, which disposition should be privileged? For example, while engaging in a thought experiment from the armchair, a planner may conclude that she is disinclined to judge that such and such on object is a thinking thing, but when she later actually encounters such an object, she finds herself judging that it is a thinking thing. Was it the theoretical disposition or the real disposition that most accurately reflected the concept being analyzed in step-one? This problem is magnified by the fact that the second step of the plan, the potential location of a unique (near) realizer, involves the attempt to bring about situations in which the real disposition would be revealed.

The planner assumes that his/her theoretical dispositions to apply the given concept accurately reflect the appropriate analysis of the concept and thereby determine the nature of any realizer of the concept. At least *prime facie*, it seems reasonable to conclude that, so long as we are sufficiently careful, the judgments about how we would apply the given concept arrived at from the armchair reflect our understanding of the concept and would be reflected in actual judgments if the actual conditions were

---

55 It is unfortunate, perhaps, that Braddon-Mitchell chooses to call these dispositions the "real" dispositions. The use of the term 'real' here is not meant to imply weight or ascribe these types of dispositions a higher status. The real dispositions are simply those that govern how one actually behaved, judged, etc.
experienced. Jackson's preference for modest conceptual analysis looms large in this assessment. The conclusions that one draws from the armchair relate to our dispositions to apply a given concept (to make judgments about its applicability) in various situations. Conclusions are not being drawn a priori about the nature of the world; to do otherwise would be to engage in immodest conceptual analysis. Since armchair analysis only allows one to conclude how a concept may be applied (rather than how the world actually is), we may reasonably take the theoretical judgments to be both modest and instances of a priori knowledge. Braddon-Mitchell notes that in the writings of planners (particularly Jackson) there is a theme that, while our concepts "may not be explicitly introspectible in the form of definitions or even clusters, what we do have access to is what we would say in various circumstances."\(^56\) Braddon-Mitchell asserts that planners, if not outright assuming that armchair thought experiments are an infallible guide to our dispositions to apply concepts, at minimum rely on armchair thought experiments to be sufficiently close to infallible as to allow us to rely upon theoretical judgments with great confidence. The problem arises if one begins to wonder whether this confidence is misplaced.

Despite certainly having *prima facie* assurance of the reliability of armchair conceptual analysis, we must also be skeptical of any claim that the theoretical dispositions would always coincide with the real dispositions. Sometimes we have a real world opportunity to observe how we apply a given concept, only to find that our real disposition to apply the concept does not result in a judgment that would have been accepted from the armchair (or vice versa). Sometimes we actually change our minds

about what to say about a given kind of case based on experience. In such cases, the real
disposition does not accord with the theoretical disposition. For example, I have a friend
whose father was deeply homophobic. The father would often make claims that he
would disavow any child of his that was a homosexual. These claims reflect his
theoretical dispositions with respect to the concept of homosexuality and its moral
evaluation. Yet, when my friend finally came out to her family, her father embraced her,
felt deep shame for his past words and thoughts, and completely changed his attitude.
Now, he is an advocate for LGBT rights in the community. These actions reflect his real
disposition about how homosexuality is to be morally evaluated. The discrepancy
between his theoretical reaction to homosexuality and his real reaction to the
homosexuality of his daughter are difficult to dismiss, and point to the heart of the
problem Braddon-Mitchell raises.

If thought experiments conducted from the armchair are able to provide
conceptual analyses that can help solve complex problems ranging from ethics to
metaphysics, it is deeply troubling that there can be such discrepancy between the
theoretical dispositions revealed through the armchair analysis of step one and the real
dispositions revealed in actual situations of the sort that may occur as part of step-two. If
we conclude that the theoretical dispositions are even occasionally but significantly out of
line with the real dispositions experienced in step-two, we may wonder about the viability
of the armchair conceptual analysis in step-one of the Plan. To salvage step-one we need
to determine which is the relevant disposition on which to base the analysis of the
concept for which a unique (real) realizer is sought in step-two: the theoretical disposition
or the real disposition. Braddon-Mitchell proposes three solutions to the problem. Yet, each solution carries a heavy price, and he finds all three problematic.\(^5\)

**Braddon-Mitchell's First Solution**

The first solution Braddon-Mitchell proposes is to assert that the theoretical disposition is the relevant disposition for determining what would count as a unique realizer of the concept in all cases. The theoretical judgment is treated as an analytic a priori judgment reflecting our understanding of the concept. Thus, should the decision about what to say in a real case be different than the decision about what to say from the armchair, then this is evidence that the individual in the actual case has changed the subject—it is not the same concept being analyzed.

While this resolves the problem, Braddon-Mitchell points out a new cluster of problems arising if this solution is adopted.\(^6\) We would normally assume that the individual in the real case likely has more information than that same individual in the armchair. For one thing, the real judgment, if it differs from the theoretical judgment, is likely rendered at a later point in time. Presumably, in the time between the armchair analysis and the real world experience, the individual may have learned a great deal more that is relevant to the analysis of the concept. Further, there will inevitably be more details perceived in an actual case than can be imagined in the merely hypothetical case. No matter how much deep and careful thinking went into identifying the theoretical disposition, it seems unreasonable to assume that the individual doing the armchair

---

\(^5\) Ibid. 40.
\(^6\) Ibid. 40-41.
analysis simply has the best grasp on everything relevant to the application of the concept in all cases. To assume that the theoretical disposition is the disposition relevant to conceptual analysis unacceptably turns the armchair into a throne.

To be sure, changes in subject can occur, and there can be instances when it is difficult to ascertain whether a change of subject has occurred. For example, the modern notion of the atom almost certainly reflects a change of subject in comparison to Democritus' conception of the atom. On the other hand, it is harder to ascertain whether the modern conception of the electron is a change of subject or a refinement of the notion of an electron as used in J.J. Thomson's plum pudding model of matter. Even so, I agree with Braddon-Mitchell that the first solution is the least persuasive. Experience at least sometimes seems to play an additive role in our understanding of the range of possible cases relevant to analyzing a concept. Even granting that changes of subjects can occur, it seems strange to posit that all disagreements between theoretical dispositions/judgments and real dispositions/judgments must be a change in subject, and that the theoretical disposition involved in armchair analysis is privileged to be the disposition that accurately analyzes the concept.

**Braddon-Mitchell's Second Solution**

Braddon-Mitchell's second proposed solution is to redefine the notion of armchair analysis. Braddon-Mitchell suggests that the armchair analysis must necessarily entail extremely detailed envisioning of cases. This envisioning is to be so detailed that either the theoretical judgment and the real judgment are guaranteed to coincide or the case
envisioned in the armchair is relevantly different from the actual case such that no conflict is created by a disagreement between theoretical and actual judgments. We can see motivation for this solution in Jackson's conception of armchair analysis. Given Jackson's emphasis on trial by possible cases, the armchair is intended to be a place where the planner asks such questions as: under what circumstances would it make sense to say that she had changed her mind rather than changed the subject? Jackson's goal is to create a thorough understanding of a set of conditions discovered from the armchair that allow the planner to have a roadmap ready in case the real disposition diverges from the theoretical disposition. Thus, no matter what the facts on the ground, the planner has already thought ahead and knows the path she will likely take in such a case. Requiring that the thought experiments conducted from the armchair include possible cases in which the theoretical and real dispositions diverge allows the theoretical disposition to remain the preferred disposition, but does not turn the armchair into a throne.

In the context of the earlier example, solution two suggests that the father had not adequately conducted the armchair analysis. The father failed to truly consider all possibilities in sufficient detail. His homophobia was due to his inability to vividly imagine the scenario where his own daughter was a lesbian. Furthermore, he was unable to vividly imagine himself loving her as a homosexual woman. The moment his daughter came out to the family, he understood that possible case with clarity, and saw details he had failed to grasp from the armchair. The discrepancy in his dispositions reflected a failure of adequate armchair analysis rather than a change in the concept picked out by his use of the term 'homosexual'.
Of the three proposals, Braddon-Mitchell seems to favor this second option. Yet, he still finds major faults with it. In many ways, the problem is not solved by this approach. This model relies on the planner not only reasoning exceptionally well, but also being able to imagine a wide range of contingencies that may or may not come to pass. Yet, if the planner fails to imagine a possible case in which theoretical and real dispositions would diverge that later comes to pass, the authority question remains. Was the original armchair analysis mistaken or has there been a change of subject? Braddon-Mitchell indicates that it seems the fundamental question remains unanswered.\(^\text{59}\)

Although it seems to be Braddon-Mitchell’s favorite, I find the second proposal to be the least satisfying. The second solution seems to require a meta-armchair analysis in which we reflect from the armchair on why the theoretical judgment potentially would diverge from a real judgment. Yet, why should we think that such meta-armchair judgments are infallible? By a similar principle, we would then need a meta-meta-armchair to think of why our meta-armchair analysis might potentially fail. A regress threatens. If the regress is vicious, then the proposed solution will not succeed. Although the regress itself may not be vicious, it raises warning signs that should weaken our confidence in this as a potential solution.

\(^{59}\) Ibid. 40-42.
Braddon-Mitchell's Third Solution

Braddon-Mitchell's final solution is to "sever the link between the a priori and the analytic" creating a new notion: the analytic a posteriori. On this solution, the real disposition is the disposition relevant to analysis of the concept, and, as a corollary, armchair analysis is allowed to be fallible. The armchair analysis creates a guide that should be followed, but does not have absolute preference in cases in which the real disposition to apply the concept differs. Instead, it is the real disposition that is discovered only a posteriori that is relevant to the analysis of the concept.

Under the third solution, the homophobic father would be wrong in his armchair analysis. The privileged disposition relating to his concept of homosexuality is reflected in how he actually reacted after learning that his daughter was a lesbian. The experiences of having a beloved daughter as a member of the LGBT community were more metaphysically and epistemologically revealing than his ability to theorize about merely hypothetical cases. On this solution, the real disposition is the disposition relevant to analysis because the real disposition is ultimately stronger and more informed than the theoretical disposition.

The third solution strongly coincides with the methodologies of the natural sciences. It is frequently the case that very convincing a priori analytic mathematical models ultimately fail to coincide with experimental results. In such cases, the information gained through experience is generally prioritized over pure theorizing. In my opinion, the third solution best captures the role accorded experience in the empiricist

---

Ibid. 40.
tradition largely embraced by the natural sciences. I will argue for the third solution in 
greater detail in the next chapter. Here I will examine Braddon-Mitchell's critique of this 
solution.

Braddon-Mitchell fears that, if this analytic a posteriori solution is correct, then an 
advanced neuroscience could better provide an account of our real linguistic and 
conceptual dispositions, eliminating the need for armchair analysis altogether. The 
meanings of theoretical concepts could, under this advanced science, be revealed to a 
third party when they are unavailable to me by introspection alone. Presumably, 
Braddon-Mitchell takes this outcome to be a reductio of this proposed solution. Yet, 
Braddon-Mitchell fails to explain why this scenario would be a problem.

As step-one is completed, the Canberra planner is willing to turn the end product 
of her conceptual analysis over to the scientist to potentially locate a unique realizer of 
the concept in the world. This, it seems, is done for obvious reasons: science has proven 
itself good at solving some location problems. If there will come a time when science is 
also capable of doing conceptual analysis better than philosophy, then let this analysis be 
done by the a posteriori methods used in the natural sciences. Braddon-Mitchell's fears 
might be somewhat assuaged if he reflected on the fact that such a situation would not 
come about without the help of philosophers. It seems that scientists wishing to develop 
this kind of brain scanning approach to conceptual analysis would need to work closely 
with philosophers who would identify relevant possible cases from which data could be

---

61 Ibid. 40-41.
collected. Only with the direct input from philosophers could an advanced neuroscience eventually be able to provide a largely a posteriori analysis of new concepts.

Just as the location problem is handed off to scientists, so too should the conceptual analysis stage be, *if and only if science has learned to do it better*. Braddon-Mitchell should not fear the replacement of philosophy with science. Yet, between now and the emergence of the advanced neuroscience that would realize Braddon-Mitchell fears, there will be many difficult philosophical problems to solve. Should Braddon-Mitchell's fears be realized, and it comes to pass that all analysis can be performed a posteriori by an advance neuroscience, philosophy will have completed its goals. By helping to create an ideal *science*, philosophers will have helped create an ideal *philosophy*.

**Where Does This Leave the Plan?**

Braddon-Mitchell has exposed a potential flaw in the Canberra Plan. If the theoretical judgment cannot be reconciled with the real judgment, then it seems that the Canberra Plan cannot be considered a viable philosophical method. I have demonstrated why I reject two of the three solutions proposed by Braddon-Mitchell. I have additionally demonstrated why Braddon-Mitchell's concerns about his third solution are unfounded. In the next chapter, I will argue for a variant of Braddon-Mitchell's third solution. I will then take the resulting Canberra Plan methodology and combine it with elements of Carnap's criteria of meaning.
CHAPTER VI
CARNAP, CANBERRA, AND MEANING

My goal in this thesis is to preserve as much of Carnap's theory of meaning as possible by modifying Carnap's four criteria in light of the later work of Canberra planners. In Chapter 2, I presented Carnap's original four criteria of meaning for a term, a, in a sentence, S(a):

(C1) The empirical criteria for 'a' are known.

(C2) It has been stipulated from what protocol sentences 'S(a)' is deducible.

(C3) The truth-conditions for 'S(a)' are fixed.

(C4) The method of verification of 'S(a)' is known.

In Chapter 3, I argued that C2 and C4 are problematic in ways that contributed to the fall of logical positivism. In Chapter 4, I outlined the Canberra Plan and in Chapter 5 I presented a significant problem for the Plan as identified by Braddon-Mitchell. In this chapter, I will show how the Canberra two-step can effectively take the place of C2 and C4 in Carnap's criteria of meaning.

A Partial Defense of Braddon-Mitchell's Third Solution

Before addressing the inclusion of the Canberra Plan into Carnap's criteria, I wish to first remind the reader of the problem associated with the Plan from the previous chapter. If Braddon-Mitchell's problem goes unaddressed, it does not seem worthwhile
to shoehorn one broken philosophy into another. In order for my solution to salvage Carnap's criteria, Braddon-Mitchell's concerns must be addressed.

Although similar to Braddon-Mitchell's solution three, my solution stresses a slightly different aspect of the real and theoretical dispositions. I believe the problem can be solved by focusing more on the internal relationship between the real and the theoretical dispositions and the conditions that obtain in the world. My goal is as follows: maintain that the real disposition revealed in actual experience trumps the armchair analysis; yet not conclude that analysis could be done away with entirely. I will, therefore, embrace a highly modified version of Braddon-Mitchell's third solution.

If we accept that armchair analysis of concepts is fallible, it may be tempting to proclaim that we do not need armchair analysis at all. However, it seems to me that the use of thought experiments remains one of the most important tools. Indeed, in many situations, the use of thought experiments is the only way (physically or ethically) that the thinker can attempt to solve the problem at hand. Some experiments are simply impossible to conduct. Putnam's Twin Earth is highly unlikely to exist. Even if it does exist, it would likely never be discovered or visited. Yet, the importance of Twin Earth is not in its physical presence, but in the role it plays in various types of thought experiments. The theoretical dispositions such thought experiments can uncover are fascinating and help to solve problems when the actual experience is unavailable.

Additionally, some experiments would be unethical to conduct. A planner may, for example, be considering the folk term 'torture'. It would be wrong to subject others to actual experiences in order to arrive at an a posteriori analysis of torture based on real
dispositions. Thus, the role of armchair analysis must be preserved in some manner even though concerns about the reliability of this sort of analysis cannot be ignored.

A nearly identical problem exists within physics as the theoretical-experimental divide. Theorists, akin to armchair philosophers, often work in a priori conditions to construct, formulate, and produce theories of the universe. The theorist has certain tools at her disposal: previous data, theories, models, etc. which can be drawn from to form newer theories. In this way, the older models, theories, and data act as platitudes or provide the content of *O-terms*. Once a new theory is developed, experimentalists work diligently to test the theory. The experimentalists testing of theories are akin to identification of real dispositions and/or the second stage of the two-step: locating the entity in the world.\(^\text{62}\) Science solves its version of the theoretical-real disposition problem by differing to the experimental result in most cases. Yet, unlike what Braddon-Mitchell fears for armchair analysis, theoretical physics is by no means viewed as useless. Theory is deeply meaningful and respected, but it is experimentation that ultimately sinks or floats scientific theory. Within science, both purely theoretical and experimental methodologies stand in the same relation to the external world they hope to describe. The theoretical informs the experimental and the experimental informs the theoretical. The theoretical and experimental aspects of science are two sorts of data gathering endeavors that both contribute to the total body of knowledge. Similarly, both

---

\(^{62}\) There is a sense in which this is a naïve conception of science, I realize. Certainly it is not always the case that progress is first made by theorists and then experimentalists go out into the world and find the fruits of the theorists' labors. Frequently new data obtained by experimentalists proves a basis for theorizing.
theoretical and real dispositions should be viewed as standing in equal relation to concept to be analyzed.

C2*: C2 & Step-One

I first propose the replacement of C2 with what I shall call C2*.

(C2*) The folk theory containing the term has undergone step-one of the Canberra Plan: a descriptive sentence consisting of *O-terms* or platitudes (depending upon whether the Lewis or Jackson approach to analysis has been used) has been generated for the folk theory.

The purpose of C2 was to give sentences involving the theoretical terms as much epistemic weight as possible. Converting sentences involving these terms into protocol sentences was hoped to ground the meaning of the terms in firm epistemological bedrock. For reasons outlined in Chapter 3, the use of sense data in protocol sentences is overly restrictive. Additionally, due to the inherently private nature of sense data, sense data seems to exclude the potential for public protocol sentences. Therefore, other positivists asserted that protocol sentences were to involve only direct observations of external states of affairs. Direct observation is less epistemically certain than sense data, but still carries a great deal of weight. Still, as I demonstrated in Chapter 3, grounding protocol sentences in direct observation created a new constellation of problems.

Protocol sentences, while philosophically interesting and potentially useful, are likely too restrictive to provide a plausible criterion of meaning. Replacing protocol sentences with a demand for a Canberra step-one descriptive sentence can effectively
alleviate the problems associated with C2. The positivists' slant on protocol sentences effectively asked the question, from what direct observations/sense data did the use of this term arise? In principle this is a very good question. The positivists erred in practice. We should allow ourselves a great deal more flexibility than a protocol sentence could provide. We could conceptualize a C2-like question: from what familiar notions did this theory arise? This C2-like formulation captures the important aspect of C2, but is now too broad. What is required is a controlling set of restrictions: not as limiting as protocol sentences, but more binding than on open-ended appeal to familiar notions. I propose that the descriptive sentence generated by step-one of the Canberra Plan will accomplish this goal.

Step-one of the Canberra Plan generates a descriptive sentence from either $O$-terms (understood in the Lewis conception) or platitudes. Daniel Nolan has argued (and many planners would agree) that, since one of the goals of step-one is a sentence capable of having a suitable realizer located by relevant experts, it seems that a planner can only seriously forward descriptive sentences capable of physical realization. If it is the case that the only claims that can pass from step-one to step-two are inherently physicalistic, then the only types of claims that can satisfy C2* will also be inherently physicalistic. This will put a significant restriction on the type of terms that can satisfy C2*.

---

64 The Canberra planners diverge from the positivists' meaning of 'physicalism'. The positivists took the central claim of physicalism to be about language (all empirical language could be reduced and/or translated into the language of physics); planners take physicalism as a claim about the objects, properties, and relations. The central claim of physicalism, as used by Canberra planners, is a commitment to there being no objects, properties, and relations over and above physical objects: the only objects that exist are those that are described by the natural sciences.
the positivists would likely agree with such a physicalistic system, this may be too much for any with non-physicalist intuitions.\textsuperscript{65}

If, on the other hand the Plan is not an inherently physicalistic then C1 plays an important role in the shaping of the descriptive sentence. C1 requires that there be empirical criteria for the application of the theoretical term. If physicalism is built into the Plan, then this will likely not be a problem as physicalism is a more demanding principle than the need for empirical criteria, and if physicalism is not inherent to the Plan, then the added requirement of empirical content will help to exclude overly vague terms and notions. This allows C1 to play a more active role than it originally could in the overall theory of meaning. C2, by demanding a very specific type of empirical criteria (sense data and/or direct observation) excluded a great deal of other empirical concepts such as indirect observation. In fact, one may wonder why C1 was required at all given the nature of the positivists' demand for protocol sentences. C2 seemingly crowded C1 out of being relevant. My solution allows C1 to play a more robust role in the theory of meaning.

Recall from Chapter 3 the example of a theoretical baryon. Recall also that creating the descriptive sentence for this proposed baryon on Carnap's view would involve gathering direct observations and/or sense data. In contrast, on Lewis' view, one only needs to gather the old terms that serve to analyze the concept of a baryon. Thus, Lewis could describe the new baryon using the understood term 'quark', but the demands of C2 places Carnap at a disadvantage. Under C2, Carnap's conception cannot make the

\textsuperscript{65} A comprehensive defense of physicalism of this sort is beyond the scope of this thesis.
same elegant move of which Lewis' conception is capable. On Carnap's view, one cannot
describe the baryon using the language of quarks, because Carnap would be equally
dissatisfied with the notion of a quark. Thus, the term 'quark' would have to be replaced
in analysis using terms and predicates denoting the quark's defining properties: color,
flavor, spin, and mass. Yet, given the nature of microphysics, even these predicates
cannot be taken to denote observable phenomenon on Carnap's view. It seems that in
order to satisfy Carnap's strict assertion that the \textit{O-terms} must involve observations, the
quark property terms would have to be defined \textit{purely} by way of observation of
instruments: trails in bubble chambers (or some analog). Yet, Lewis has no problem
with the baryon example as his conception allows for any well-defined term to be an \textit{O-}
term.

One of the goals of Carnap's criteria of meaning was to eliminate vagueness.
Carnap was particularly concerned with terms that appeared to be meaningful but were
actually being defined by equally vacuous terms. The Canberra step-one also forbids such
a defining of theoretical terms with ill-defined terms. Lewis is clear that the \textit{O-terms}
must be well-defined and understood. When one adds the caveat of C1 (the terms must be
empirical), the restriction on the types of terms or predicates used in step-one becomes
fairly exclusive. \textit{O-terms} (or Jacksonian platitudes) have the added benefit that, should
there be a philosophy that still wishes to attempt to work at the direct observation/sense
data level, such terms could be used as the \textit{O-terms} (or function as platitudes). Carnap's

---

66 The properties of quarks are named with a certain strange flair; the reader should not
take 'color', 'flavor', and 'spin' to be identical with ordinary uses of the same terms.
67 Although bubble and cloud chambers have fallen out of usage given advancements in
electronic detection, these two devices remain the most commonly known.
C2 could still be extracted from C2* (if one so desired), but the rigidness of protocol sentences is not required.

Despite the restrictive nature of C2*, if this were the only alteration to Carnap's criteria of meaning, then it would likely be insufficient. Step-one of the Canberra Plan likely still allows for the analysis of a great many concepts that fail to apply to anything that actually exists. Recall that the inclusion of protocol sentences in the original C2 caused C2 to, in a sense, engage in double duty. Protocol sentences sought to ground terms in the world either through sense data or direct observation. To borrow from Canberra terminology, the inclusion of protocol sentences in the original C2 both helped to define terms and to locate their realizers given that, whenever one had the relevant sense data (or made the direct observations), then the referent of the term was present. A strength of the Canberra Plan is the division of labor between step-one and step-two: the armchair and the experimental. With the failure of C2, it seems that the analysis phase and location phase are better split. My inclusion of C2* formally separates analysis and location. C2* represents the analyzing of the concept picked out by the terms, where C4* will fill the role of the location phase.

The new C2* seems to follow in the general guiding principle set forward by C2. C2* still grounds the analysis of a concept in a trusted vocabulary. Yet, that vocabulary is more flexible and adaptable than the positivists' original conceptions of sense data and direct observation. This loosening of the reigns allows for more terms to count as meaningful, but still limits most of the vagueness that Carnap's original criteria sought to
filter out. Although C2* forgoes the locating ability of C2, my updated C4* picks up the slack.

**C4*: C4 & Step-Two**

On the surface, it may appear as if I could simply drop in the second step of the Canberra Plan into C4 in much the same way that I simply replaced C2 with step-one of the Plan. As I shall show, this is not the case. What is needed is a more substantial understanding than a simple substitution allows. As I shall show, the second step of the Canberra Plan is contained in what I will propose as C4*, but the second step by itself is not sufficient—particularly given the nature of the Canberra Plan and the possibility that a unique realizer of the descriptive sentence would not be located.

There are many reasons why a folk theory may have a descriptive sentence whose unique realizer (if any) that cannot (or should not) be located: it may be unethical, too difficult to locate given modern technology, too expensive, physically impossible, etc. In addition, the sentence generated in C2* may involve terms that simply fail to denote existing objects. Location may not be possible if the *O-terms* or platitudes denote objects that do not exist in the actual world. Any attempt to provide criteria for meaning must address these concerns.

Descriptive sentences produced during step-one (now incorporated into C2*) seem to have meaning whether or not they are actually uniquely realized. It seems too much to demand that all terms in the sentence generated at C2* *must* themselves be analyzed (ala step-one) and then have a located unique realizer (ala step-two) located in
the world in order to be meaningful. One cannot discard as meaningless a term like 'torture' simply because one does not have the stomach to attempt to locate the referent of the term. The fruits of step-one must therefore have meaning even without successful location in step-two. Yet, the preference for realization cannot and should not be dismissed.

To address the concern, I suggest two categories of meaningfulness: meaningful and with a located unique realizer (MLR), and meaningful but absent a unique realizer (MAR). MAR terms meet criteria C1-C3, but cannot or should not have a unique realizer located (for considerations of an ethical, fiscal, etc. nature), even though a method by which a potential unique realizer could be located is known. An example of an MAR term may be the descriptive sentence of analyzing the term 'torture'. It is understood how one could finish the Canberra two-step and complete the location phase, but to do so would be barbaric. An MAR term should not be confused with a meaningless term. To be classified as an MAR term, we have to know what it would take to locate a suitable realizer, but (for whatever reason) the location phase has not commenced or has not yet yielded any results. A meaningless term, on the other hand, is a term in which one or more of the other three criteria have not been met.

The other category, MLR, meets all the requirements of C1-C3 and has a unique realizer located in the world. For example, 'neutrino' is a term that is MLR because a descriptive sentence was generated in line with the added requirements of C1 and C3, and a particle was located in the world that realizes that sentence. An MAR term can become an MLR term if location of a unique realizer occurs. For example, under this
construction, 'neutrino' was an MAR term until the successful location of a particle that realized the descriptive sentence, at which point 'neutrino' became an MLR term. MLR terms have greater epistemic weight than MAR terms because the realizers of MLR terms have been successfully located. This distinction helps to recapture some of the epistemic notions lost when C2 was abandoned in favor of C2*.

This distinction allows for potentially non-denoting terms to remain meaningful. For example, 'Santa Claus' is an MAR term because the method of location is known. We could exhaustively search the North Pole for a person meeting the descriptive sentence generated by analyzing the term 'Santa Claus'. The Santa Claus theory is (presumably) false, but not meaningless: it meets the four criteria of meaning. There is empirical content to the Santa Claus folk theory (C1). The concept picked out by the term, 'Santa Claus', is capable of undergoing a Canberra step-one analysis (C2). The truth-conditions are fixed (C3). Lastly, the method of potential location of a unique realizer is known. C4* allows for non-denoting but meaningful but non-denoting terms to exist, as we would hope.

The provision for MAR terms in C4* allows the new criteria to be more flexible than Carnap's original criteria of meaning. Since MLR terms carry the greatest epistemic weight, the goal of analysis should be the identification of MLR terms. MAR terms can then be seen as a necessary sub-case of terms whose referents are too dangerous, unethical, costly, etc. to locate or that lack referents entirely because they are terms of a false folk theory. In an ideal scenario, all terms would be MLR terms, but the world is

---

68 Sorry, Virginia.
not ideal. Therefore, certain accommodations must be made to allow for these limitations. Yet, even with these accommodations, the new criteria of meaning are able to effectively winnow out the vague terms as desired by the positivists. Terms that are neither MAR terms nor MLR terms have failed some previous criterion of meaning. Thus, such terms are meaningless (just as it would be under Carnap's original construction). Therefore, I propose the following as the new C4: (C4*) The descriptive sentence generated in C2* is either classifiable as meaningful but absent a unique realizer (MAR) or meaningful and with a located realizer present (MLR), but not both.

A New Criteria of Meaning: Carnap & Canberra

I have attempted to solve the problems associated with Carnap's criteria of meaning by revising the criteria to reflect Canberra Plan methodology. While this may create new philosophical problems, it has helped to solve some of the old problems. If these criteria were to be adopted, there is still much work to be done. This is due in part to the Canberra Plan being in its infancy. Yet, I believe I have accomplished what I set out to do: the new criteria seem to persevere many of the goals and results of Carnap's original work, but also avoid some of the same pitfalls.
SELECTED BIBLIOGRAPHY


