Carnap Ponders Canberra: Creating a Theory of Meaning Based on Carnap’s Criteria of Cognitive Significance and the Canberra Plan

Andrew Whiteley Magrath
Department of Philosophy, Kent State University, Kent, USA
Email: amagrath@kent.edu

Received April 29th, 2013; revised May 29th, 2013; accepted June 7th, 2013

Although logical positivism has fallen out of favor within many philosophical circles, one might remain sympathetic to the logical positivists’ critique of the meaningfulness of philosophical terms. In an attempt to address this open problem, I will propose an updated theory of meaning by combining the Canberra Plan methodology and Carnap’s four criteria of cognitive significance as explicated in “The Elimination of Metaphysics Through Logical Analysis of Language”. I will argue that of Carnap’s original criteria, the notion of protocol sentences (C2) and verification (C4) remain problematic. I will further argue that protocol sentences can be replaced by step-one of the Canberra Plan. This shift towards Lewis’ more robust conception of O-terms as any known/old terms will still limit the types of terms that are meaningful yet proven to be less problematic than Carnap’s notion of observation terms. Additionally, I will argue for the replacement of verification with a modified version of step-two of the Plan. Due to the emphasis of locating a unique realizer within the Canberra Plan, I will propose two types of meaningful terms: meaningful with a located unique realizer (MLR), and meaningful but an absent unique realizer (MAR).

Keywords: Canberra Plan; Rudolf Carnap; Frank Jackson; David Lewis; Logical Positivism; Logical Empiricism; Theory of Meaning

Introduction

Despite the failures of the logical positivists to bring about a full working theory of meaning, closeted logical positivists (and likely philosophers of other stripes) may nevertheless remain sympathetic to the positivists’ original critique of the meaningfulness of many philosophical terms. If the positivists’ objections remain salient, then a positivist-esque theory of meaning is a worthwhile project. I will attempt to lay the groundwork for such a theory using 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. Although the Canberra Plan shares many traits with logical positivism, the Plan represents a more flexible approach to philosophy and philosophical analysis. Despite engaging many of the same problems that interested the logical positivists, an attempt to create a theory of meaning has not, to my knowledge, been put forward within the Canberra Plan literature. Frank Jackson, in his text From Metaphysics to Ethics (Jackson, 1998) asserts that that the Canberra Plan can contribute to any problem of philosophy. Given these qualities of the Plan and a still relevant critique of meaning as left by the positivists, the solution I will put forward is a fusion of Carnap’s criteria of meaningfulness (outlined in “The Elimination of Metaphysics Through Logical Analysis”) and the Canberra Plan’s two-step process.

Although I am not the first to reexamine logical positivism,1 I believe my approach is nevertheless novel. While many that attempt to address some perceived problem of logical positivism do so within what can broadly be understood as a logical positivistic framework, I am attempting to examine what I take to be problems with Carnap’s criteria of cognitive significance from a standpoint of a Canberra planner. This shift in frameworks allows me to work towards two goals: First, address a problem that I believe exists within Carnap’s original theory of meaning. Second, address what I see as a deficiency within the Canberra Plan—lack of a theory of meaning.

In “The Elimination”, 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 (Carnap, 2003: p. 158).

I will argue that of Carnap’s original criteria of meaningfulness, C2 and C4 are philosophically weak, and will need to be replaced. I further assert that C2 can be directly replaced with step-one of the Canberra Plan, as the two play functionally similar roles. Yet, despite these functional similarities, step-one

1For a noted contemporary example, see Michael Friedman’s Reconsidering Logical Positivism.
of the Canberra Plan is less problematic than C2. I will further replace C4 with a version of the second step of the Canberra Plan.

**The Strengths of C2 & C4**

Of primary importance to the positivists was 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 since nothing is asserted beyond the existence of experience. 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 follow this line of thinking. The positivist’s conception of the protocol sentence is meant to express the empirical data provided by sensory perception of the world in its rarified raw form. The strength of this approach is that one can potentially create sentences that are reduced to epistemically strong sense impressions. Using this approach within a theory of meaning, Carnap is able to pronounce many terms in philosophy as meaningless. Of particular interest to Carnap are metaphysical concepts that are claimed to bypass experience.

Many empiricists have historically been uncomfortable with concepts that are claimed to bypass experience. While attempts have been made to demonstrate such concepts are inconsistent, overly speculative, and/or false, the positivists’ unique approach was to assert that terms describing such concepts are meaningless. From Carnap’s criteria of meaning it is clear that for such terms to be meaningful, they must either be grounded in sense impressions—i.e. have empirical content—or must be derivable from analytic truths, else the term is meaningless. Carnap asserts that there may be psychological reasons why a concept is believed to reach around experience, but, upon close inspection, such concepts often ultimately rely on empirical content. With the illusion gone, and the term’s empirical nature uncovered, Carnap argues that the concept the term described is best left to the sciences to understand. If, on the other hand, the term is derivable from analytic notions, then the term says nothing about the synthetic world owing to the nature of the positivist’s conceptions of the synthetic a posteriori and the analytic a priori. In either case, for the term to be meaningful, it cannot truly reach around experience.

I remain sympathetic to Carnap’s overall project and what I take to be the driving forces behind C2 and C4. I agree with Carnap that there should be a division of labor between science and philosophy, which seems to be a motivating factor in C4. While I will ultimately find the positivists’ particular conception of verification problematic, I will nevertheless attempt to preserve Carnap’s balance between science and philosophy. I also remain sympathetic to the attempt to ground terms in an epistemically strong vocabulary set. While I remain unconvinced that these terms need be fully reduced to the austere language of sense impressions, I nevertheless agree with Carnap that meaning is closely related to the ability to reduce terms to a familiar, tested, well-defined vocabulary set. My goal moving forward is to preserve the relationships that C2 and C4 establish (terms and familiar vocabulary sets, science and philosophy), but avoid what I consider to be problems with Carnap’s original execution.

**The Troubles with C2 & C4**

The epistemic strength of the sense data protocol comes from its austerity. Yet, it may be the case that this sense data language is so austere that it hinders its ability to describe. 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 exported a great deal of philosophy’s explanatory project to the natural science. The use of sense data greatly compromises this project. It is often argued that 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. 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 eventually 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.

Protocol sentences involving direct observation terms—that is to say sense data directly perceived by the senses with no other structures or features inferred by the intellect. Although more robust, direct observation protocol sentences are more susceptible to skepticism than are protocol sentences involving sense data. Nevertheless, the convincing attacks on sense data protocol sentences caused many positivists to adopt direct observation protocol sentences. Unfortunately, this move is also problematic.

One of Carnap’s projects was an attempt to replace all theory terms (T-terms) with observation terms (O-terms) through the process of Ramsification. 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 Φ. 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;\(^2\) 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:

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

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 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 different direct observation translations of the term ‘pretendon’:

- Human Detector: \( \forall x (\text{Pretendon}(x) \leftrightarrow \text{Blue light}(x)) \);
- Martian Detector: \( \forall x (\text{Pretendon}(x) \leftrightarrow \text{Chime}(x)) \).

Eliminating the \( T\)-term ‘pretendon’ using observation terms from the human detector, results in utilizing the \( O\)-term ‘blue light’. Yet, the Martian detector replaces the \( T\)-term ‘pretendon’ with ‘chime’. \( \textit{Ceteris paribus} \) the two theories are identical; it therefore seems that the term ‘pretendon’ should be translated into the same expression in the language of \( O\)-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. Yet, a human scientist would have no trouble understanding what a Martian pretendon was because functionally the human pretendon is identical. It is unlikely that the human (or Martian) scientist would believe that their respective pretendons were different based solely on a chime or light. It therefore seems that the role of direct observation inherent to \( C2 \) is too limiting, resulting in cases were intuitively identical objects must be considered to be of different kinds. It seems that \( C2 \) therefore fails.

\( C4 \) can be understood as following from \( C3 \). Where \( C3 \) establishes that there is fixed truth conditions for ‘\( S(a) \)’, \( C4 \) emphasizes how one would determine whether those conditions obtain. Since the positivists asserted that all propositions were supported by the types of observations typically made—particularly in the sciences. It may also strike one as intuitively strange to assert that there is no 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 \( x \)s is unknown, no amount of observations can verify a universal claim.\(^3\)

For these reasons, I contend that \( C2 \) and \( C4 \) must be revised.

### The Canberra Two-Step

The goal of the Canberra Plan is to analyze a group of related concepts which, when taken together, form what is referred to by practitioners of the Plan (referred to as 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 analysis of the folk theory is accomplished, broadly speaking, by the “Canberra two-step” (Nolan, 2009: p. 269).

The fist 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 nature of terms that make up the descriptive sentence obtained by step-one has created an internal divide within the Canberra Plan. The first tradition, forward by Lewis, is similar to Carnap’s treatment of theory terms (\( T\)-terms). Through Ramsification, Carnap was able to replace all \( T\)-terms in a sentence with observational terms (\( O\)-terms). Where Carnap specifically attempted to ground his \( O\)-terms, observation-terms, in the realm of sense data/direct observation, Lewis asserts that the \( O\)-terms need only be: “\(^4\)Any other term, one of our original terms an old term we already [understand] before the new

---

\(^2\)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.

\(^3\)I will ignore Hempel’s move towards confirmation as opposed to verification. Suffice to say, I find Hempel’s attempts in “Studies in the Logic of Confirmation” unsatisfying.
theory $T$ with its new $T$-terms…” (Lewis, 1983: p. 79) Under Lewis’ conception, $O$-terms can take the form of any previously understood notion that can be represented by a predicate. Lewis hopes to formalize the relationships between $T$-terms and $O$-terms into axioms and logical systems of first order theories. The other Canberra Plan tradition, as forwarded by Jackson, generates a descriptive sentence containing only platitudes. Unlike Lewis, Jackson focuses on concepts. Shifting the focus allows Jackson to avoid the formal process of Ramseyfication and instead work with natural language statements 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 (although within Jackson’s philosophy, he will often carry out Ramseyfication).

The newly generated sentence describing the analyzed concept can be now handed off to the appropriate department to attempt to identify a unique realizer—step-two of the Plan. The nature of how a term is or might be realized often depends on the nature of the concept that is analyzed. For example, the notion of a theoretical particle in physics would be located differently than the aesthetic notion of beauty. The nature of location is dependent upon the nature of the concept analyzed and the experts attempting location.

**C2*: C2 & Step-One**

To generate a Carnap-Canberra theory of meaning, 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 old terms ($O$-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. Yet, 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 errored 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. If the positivists critique is true—that there are meaningless terms seemingly being meaningfully discussed in familiar philosophical theories—then merely grounding terms in such familiar theories is insufficient. What is required is a controlling set of restrictions: not as limiting as protocol sentences, but more binding than on open-ended appeal to merely familiar notions. I propose that the descriptive sentence generated by step-one of the Canberra Plan is capable of providing the needed controlling role.

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 (Nolan, 2009: p. 269). 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*. Although the positivists would likely agree with such a physicalistic system, this may be too much for any with non-physicalist intuitions.

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. Yet, if physicalism is not inherent to the Plan, then the added requirement of empirical content will help to exclude overly vague/ambiguous terms and notions of the sort that worried the positivists.

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 description and location. 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, while still restrictive, this vocabulary is nevertheless more flexible and adaptable than the positivists’ original conceptions of sense data and/or direct observation.

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

It may be tempting to simply replace C4 with the location step of the Canberra Plan. Yet, 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 possibility that a unique realizer of the descriptive sentence would not/could not be located.

The Canberra planners typically 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.

A comprehensive defense of physicalism of this sort is beyond the scope of this project.
There are many reasons why a folk theory may have a descriptive sentence whose unique realizer (if any) cannot (or should not) be located. Location may be unethical, too difficult 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 platitude denotes 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. Yet, the preference for realizability: meaningful, with a located unique realizer (MLR), and meaningful but absent a unique realizer (MAR). Terms meet criteria C1, C2*, and C3, but cannot or should not have a unique realizer located (due to ethical, fiscal, etc. considerations), 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, C2*, and 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. 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 also 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. 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. 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.

With these accommodations, the new criteria of meaning are still able to effectively winnow out many of the terms that troubled 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).

A Positivist-esque Theory of Meaning: Carnap & Canberra

By revising Carnap’s original criteria to reflect Canberra Plan methodology, I have attempted to solve problems associated with the positivists’ theory of meaning and provide a theory of meaning for the Canberra Plan. Although there is a great deal more to be done, I nevertheless believe I have accomplished what I set out to do: this new criteria seems to take seriously the positivists’ initial charge of rampant meaninglessness, yet poses a solution that, while sufficiently strong, nevertheless avoids some of the pitfalls associated with the positivists’ initial efforts.

Acknowledgements

I would like to thank Deborah Smith and David Odell-Scott for their continued support and guidance. I would also like to thank Kwang-Sae Lee who I hope now spends his time taking walks with Kant and having beers with Heidegger—hopefully in that order.

REFERENCES


*Sorry, Virginia.