Donald Davidson's Semantics: Radical Interpretation, Triangulation, and the Ambiguity Problem

Karen Duffy
College of Saint Benedict/Saint John's University

Follow this and additional works at: https://digitalcommons.csbsju.edu/philosophy_students

Part of the Philosophy Commons

Recommended Citation
https://digitalcommons.csbsju.edu/philosophy_students/1
Donald Davidson’s Semantics: Radical Interpretation, Triangulation, and the Ambiguity Problem

An Honors Thesis

College of Saint Benedict and Saint John’s University

In Partial Fulfillment

of the Requirements for All College Honors

and Distinction

in the Department of Philosophy

by

Karen Duffy

Advised by

Dr. Stephen Wagner

May, 2012
Contents

Introduction 4

Chapter 1
Davidson’s Meaning Theory 8
  1.1 The compositionality Constraint and the Inutility of Meanings-as-Entities 8
  1.2 The Introduction of a Truth Theory 13
  1.3 The Initial Project 15
  1.4 The Extended Project 22

Chapter 2
The Ambiguity Problem and the Triangulation-Based Solution 30
  2.1 Underdetermination 30
  2.2 The Ambiguity Problem 31
  2.3 Davidson’s Triangulation Thesis 33
  2.4 Triangulation as a Response to the Ambiguity Problem 38
  2.5 An Additional Comment on Triangulation 39

Chapter 3
Defending Triangulation 41
  3.1 Talmage’s Objection 41
  3.2 Quine’s “Gavagai” Example 44
  3.3 Lepore and Ludwig’s Objection 46
  3.4 Conclusion 50

References 52
Introduction

Most people use language constantly, and are able to understand and generate expressions with little to no conscious thought. Because of this, people do not often question what it is that their expressions actually mean and how they come to know this – they just know what the expressions mean, and that is all they need to get by in the world. Philosophers, however, being the particularly inquisitive bunch that they are, are not so easily satisfied. Thus over the years, philosophers have inquired into the realm of meaning, formulating various theories of how to specify the meaning of expressions.

A common approach to developing a meaning theory (a theory that specifies the meaning of any expression of a given language)\(^1\) is to suggest that the meanings of expressions are contained in entities called “propositions,” which are connected to their corresponding expressions by some relation of reference. Other philosophers, however, have taken issue with the “meanings as entities” approach, and have reacted against it by devising their own theories that do not employ such a conception of meaning. Among the most prominent of these philosophers is Donald Davidson. In several individual articles, Davidson advances a non-propositionalist account of the meanings of expressions, revolving around his famous suggestion that a properly constrained Tarski-style truth theory for a language can give us the meaning of expressions as well as their truth conditions. Davidson also develops an empirical method by which he argues we could construct such a truth theory. This is the project known as “radical

---

\(^1\) Ernie Lepore and Kirk Ludwig, *Donald Davidson: Meaning, Truth, Language, and Reality*, Oxford: Oxford University Press (2005): 19-20. Lepore and Ludwig distinguish the term “meaning theory” (as defined above) from a “theory of meaning,” which is a theory about the concept of meaning – what meaning is – and does not purport to be able to give the meanings of expressions in a particular language. The present paper is concerned only with a meaning theory, not a theory of meaning.
interpretation.” Taken together, these form a unified theory that purports to be able to give a procedure for specifying the meaning of any expression in any language – a meaning theory.

Davidson’s united meaning theory is far-reaching, innovative, and complex. As such, there are many controversial aspects, each of which could probably generate a book’s worth of discussion. Confronting the question of radical interpretation’s success comprehensively is a task that would require a hefty volume (i.e. Lepore and Ludwig’s *Donald Davidson: Meaning, Truth, Language, and Reality*), and is quite beyond the reach of one paper. However, once confronted with Davidson’s theory as a whole, the natural question for one to ask in response is, “okay, now, does it work?” So, given the fact that I do not desire to write a book, I will approach the question of the success of radical interpretation by looking at one particular sub-issue.

My project is this: I will attempt to contribute to the evaluation of the success of Davidson’s radical interpretation by considering a particular issue that arises during radical interpretation that, if unsolved, could be fatal to its success. This issue is sometimes referred to as “the ambiguity problem.” The issue is that ambiguities that arise during the process of interpretation seem to indicate that the evidence available to the interpreter underdetermines any theory confirmed on the basis of this evidence. If these ambiguities cannot be resolved, radical interpretation will be unsuccessful in producing a meaning theory for a language. Davidson employs his triangulation thesis – the claim that speakers must be or have been in communication with other speakers in order to actually be speakers of a language (make meaningful utterances) – as a solution to these ambiguities. My primary thesis is that the triangulation response provides a successful solution to the ambiguity problem. I will argue for this first by presenting the triangulation view and the initial argument that it solves the problem
at hand, and then by defending it against certain objections that have been leveled against the view by some of Davidson’s readers. As a consequence of arguing for this primary claim (that an appeal to triangulation solves the ambiguity problems), I am automatically arguing in support of a secondary claim: that radical interpretation is a successful. Since the ambiguity problem was a threat to the success of radical interpretation, by arguing that it can be solved, I am adding to the case that radical interpretation is successful in producing a meaning theory for a given language.

Clearly, a significant amount of groundwork needs to be done before I can get to directly arguing for my thesis. The paper will proceed as follows: in order to give any sort of evaluation of Davidson’s theory, we first need to spell his theory out with an adequate degree of precision. Chapter 1 will cover the motivation for Davidson’s theory by describing his dissatisfaction with traditional “meanings as entities” approaches and his process for arriving at the suggestion that a properly constrained truth theory can also work as a meaning theory. I will then attempt to give a clear characterization of how Davidson’s theory actually works in two parts, based on a helpful distinction between Davidson’s “initial” and “extended” projects from Lepore and Ludwig. The first part aims to show how a truth theory can work as a part of a meaning theory. The second part explains Davidson’s “radical interpretation,” which is his procedure for actually constructing a meaning theory based purely on the behavior of speakers.

Once the theory under discussion has been clearly characterized, I will turn in Chapter 2 to the ambiguity problem and the triangulation-based solution. I will first give a brief description of underdetermination, which is necessary in order to understand the weight of the ambiguity problem. I will then spell out the ambiguity problem itself, and explain why the ambiguity problem poses an issue for the success of radical interpretation. Next I will explain what
Davidson’s triangulation thesis is, how he argues for it, and finally argue that triangulation solves the ambiguity problem.

Finally, Chapter 3 seeks to defend triangulation against objections advanced by some critics of the view. I will consider three objections to triangulation: one from Catherine Talmage, one derived from Quine’s “gavagai” example, and one from Lepore and Ludwig. In this chapter, I will present these three objections to triangulation, and offer my own responses to each objection in turn. By doing so, I hope to add support to the claim that triangulation solves the ambiguity problem.
Chapter 1

Davidson’s Meaning Theory

1.1 The Compositionality Constraint and the Inutility of Meanings-as-Entities

Davidson begins his investigation into meaning by thinking about what form a meaning theory must take in order to be adequate. Davidson’s first proposal is that a meaning theory must exhibit meaning to be *compositional* – that is, it must show how the meanings of complex expressions depend on the meanings of their parts. This was not a new idea; as Davidson says in the beginning of his article “Truth and Meaning,” many philosophers of language at the time, as well as some linguists, had come to this conclusion as well.\(^2\) However, since this compositionality constraint is at the very foundation of the theory Davidson develops, it is important to look at why one might suppose a meaning theory needs to exhibit the compositionality of meaning in order to be adequate.

The argument for the compositionality constraint is based on the learnability of language. Meaning must, Davidson argues, be compositional, because if it were not, it would be impossible to learn a language. Natural languages contain an infinite number of nonsynonymous expressions. To know a language is to be able to understand any expression of that language.\(^3\)

---


\(^3\) This statement is probably too strong. Taken at face value, it implies that if one does not know every single word of the English language, one cannot know English. This would likely result in no person actually knowing English, which is not a desirable conclusion (I encounter a number of people everyday who I would say know English). However, this might be resolved by realizing that knowing a language is not an exclusive “either you know it or you don’t” situation. Rather knowing a language might be conceived of as on a continuum, with more and less competent speakers. We might then say that one knows a language when one has reached a certain level of competency – when one is in a position to be able to understand the vast majority of expressions in that language, or something of that sort.
And for Davidson, to understand an expression is to be able to specify what it means.⁴ So, to know a language is to be able to specify what any expression⁵ in a language means. Therefore, if meaning were not compositional, each expression would have to be learned individually in order to know a language. Since there are infinite expressions, assuming each expression takes a finite amount of time to be learned, learning a language would take an infinite amount of time. Therefore, language would be unlearnable, since humans have finite life spans. However, since we know that humans do learn languages, it must be the case that meaning is compositional – languages consist in a finite number of basic expressions and a finite set of rules, by which one is able to understand an infinite number of nonsynonymous expressions.⁶

From this reasoning, Davidson concludes that any proposed meaning theory that does not demonstrate compositionality – does not show how the meanings of complex expressions are based on the meanings of their atomic parts – has failed to explicate something essential to how meaning operates, and is thus inadequate.⁷ We have thus established the compositionality constraint. After establishing the necessity of this constraint on a meaning theory, Davidson’s research question becomes: how exactly does one give such a compositional account of meaning?⁸ What sort of theory will do the trick?

Davidson devotes a fair amount of time to addressing a certain approach to giving a meaning theory that was quite popular at the time he was doing his work. The idea of this approach is that “meanings” of expressions are things, or entities, to which expressions somehow...

---

⁴ Lepore and Ludwig, Donald Davidson, 26.
⁵ Or perhaps the vast majority of expressions, if we go with the “competency” interpretation of knowing a language sketched above in Chapter 1, footnote 2.
⁶ This argument is taken both from Davidson’s own presentation in Donald Davidson, “Theories of Meaning and Learnable Languages,” Inquiries into Truth and Interpretation, Oxford: Clarendon Press (1984): 8-9, and a more formalized presentation in Lepore and Ludwig, Donald Davidson, 28.
⁷ Davidson, “Theories of Meaning,” 3.
Davidson strongly objects to this approach, and this objection provides the impetus behind his suggestion that a truth-theory could also serve as a meaning theory, if properly constrained.

Davidson’s main objection to the meanings-as-entities approach is that they lack utility in constructing a compositional meaning theory. To demonstrate this, Davidson asks us to consider the expression “the father of Annette.” Presumably the meaning of this expression is simply the father of Annette. Can using the meanings-as-entities approach lead us to this conclusion? The meanings-as-entities approach would go about answering this question by first assigning some meaning to each of the parts of the expression – in this case, “the father of,” and “Annette.” Suppose for this example that we have assigned Annette as the meaning of “Annette.” Finding the entity to assign to “the father of” is a bit trickier, however, as Davidson notes, “the answer would seem to be that the meaning of ‘the father of’ is such that when this expression is prefixed to a singular term the result refers to the father of the person to whom the singular term refers.” That is to say, the meaning of “the father of,” according to this approach, would be a function that maps people to their fathers.

So, when asked what the expression “the father of Annette” means, the meanings-as-entities approach gives us that the meaning of “the father of Annette” is some concatenation of a function which maps persons to their fathers, and Annette. It may seem like this does the trick, but Lepore and Ludwig make the perceptive observation that the approach does not succeed. What we were looking for was a way to understand the expression “the father of Annette” based

---

9 Lepore and Ludwig, *Donald Davidson*, 39.
11 Ibid., 18.
12 Lepore and Ludwig, *Donald Davidson*, 46.
13 Ibid.
14 Ibid.
on the meanings of its parts (since the theory must be compositional, as above shown), which are
a function that maps people to their fathers, and Annette. This is not enough, however, to
understand that “the father of Annette” means the father of Annette (which is our goal). Based
solely on the meanings of the parts, we are given no instructions that Annette is the person who
ought to be plugged into the father function to yield the father of Annette. Given only the
meanings, we have no reason to plug Annette into the father function over any other person.\(^{15}\)

Davidson now provides an alternative approach to this example.\(^{16}\) Suppose we construct
a mini-theory for dealing with the expression “the father of Annette.” Our mini-theory has two
axioms: (1) “Annette” means Annette, and (2) “the father of,” when prefixed to a singular
referring term \(x\), means the father of \(x\). Now, when asked for the meaning of “the father of
Annette,” we can say by axiom (2) that the expression means the father of “Annette,” and then
by axiom (1) conclude that its meaning is the father of Annette. Davidson’s point here is that no
meaning-entity is or needs to be assigned to “the father of” in this account.\(^{17}\) It is a fine
distinction. Put another way, the difference between the two approaches is this: the first
approach, the meanings-as-entities approach, has “the father of” mean the function that maps
people to their fathers itself (a general statement, without respect to a specified input), while
Davidson’s approach has “the father of” mean the operation of this function with respect to a
specified input. The key difference is that in the former approach, we do not have enough
information to actually carry out the function – since we are referring to the function itself we

\(^{15}\) One might challenge this interpretation (as suggested to me by Dr. Erica Stonestreet) by suggesting that the
division of “the father of Annette” into only two parts (“the father of” and “Annette”) as opposed to three (“the
father,” “of,” and “Annette”) is arbitrary. The suggestion is that perhaps treating “of” as its own part, with its own
independent meaning, is what connects Annette up with the function mapping people to their fathers. However,
now the question becomes what is the meaning of “the father”? It no longer seems to be a function mapping people
to their fathers. I am not sure what the answer to this is, but my inkling is that since treating “of” as its own part
results in the ambiguous part “the father,” this move does more harm than good for the meanings-as-entities view.
\(^{16}\) Davidson, “Truth and Meaning.” 18.
\(^{17}\) Ibid.
lack an input. On the latter approach, we are not referring to the function itself, but rather the meaning of the expression is two rules: one that gives the function and the other that tells us what the input is. This is crucial, because we need a theory that will be compositional, and only the latter (Davidson’s) approach is able to accomplish compositionality.

One might object here that Davidson is claiming the inutility of meanings as entities while making use of them at the same time – did not one of the axioms specify that the meaning of “Annette” was Annette, and is this not a meaning as an entity? However, I think the answer to this lies in the fact that “Annette” is a proper noun – the function of the term is to name a thing (in this case, a person). So it is an important question. Davidson follows up his own exposition of this example with the caveat that the “task was to give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts; it was not in the bargain also to give the meanings of the atomic parts.” It may seem at this point like this is a cop-out. However, the answer will come shortly when we begin to discuss Davidson’s positive proposals in terms of Lepore and Ludwig’s “initial” and “extended” projects. For the time being, Davidson’s only conclusion from this example is that it is possible to construct a theory giving the meaning of a complex expression (i.e. “the father of Annette”) without appealing to meanings as entities for all the parts of the expression.

This will be acceptable when we consider the limited scope of the initial project. Once we turn to the extended project, however, we will see how Davidson develops radical interpretation as a process that details how all the expressions of a language come to mean.

\[18\] Ibid.

\[19\] Ibid.
1.2 The Introduction of a Truth Theory

At this point, Davidson has made his arguments for the compositionality constraint on the form of an adequate meaning theory and against the utility of treating meanings as entities to which expressions refer. We are now ready for Davidson’s positive proposals. What Davidson is ultimately looking for is a theory that will yield theorems of the form

\[(M) \text{‘} s \text{’ in } L \text{ means that } p.\]

Where \(L\) is the language of study (the “object language”), ‘\(s\’ is a structural description of an expression in \(L\), and ‘\(p\’ somehow gives the meaning of ‘\(s\’ in the language of the theory (the “metalanguage”). It is important to recall at this point that for Davidson, to give the meaning of an expression is simply to enable someone to understand the expression. So whatever replaces ‘\(p\’ must be something that is sufficient to enable the understanding of ‘\(s\’ by someone who does not speak \(L\).

The present question is, then, how to go about performing this matching of ‘\(s\’ in \(L\) with the appropriate ‘\(p\’ in the metalanguage? Recall that the appropriate ‘\(p\’ , as we have just said, is one that enables the understanding of ‘\(s\’. What Davidson proposes is this: we should get rid of the “means that” language for the time being, and instead provide a new predicate, \(T\), that will apply to ‘\(s\’ if an only if \(p\.

\[(T) \text{ ‘} s \text{’ in } L \text{ is } T \text{ if and only if } p.\]

We can then determine the appropriate ‘\(p\’ for each ‘\(s’ by asking, if \(p\, is ‘\(s’ T?\) Vice-versa, we can determine the appropriate ‘\(s’ for each ‘\(p’ by asking, if ‘\(s’ is \(T\, then \(p\?\) If the answer to both of these questions is “yes,” we have found an appropriate match, while if the answer to either is “no,” the match is inappropriate.

\[20\text{ Ibid., 23.}\]
What we need now is to determine what predicate ‘is \( T \) actually is. It is immediately clear to Davidson that “the sentences to which the predicate ‘is \( T \)’ applies will be just the true sentences of \( L \).”21 However, since not everyone is fortunate enough to be as intelligent as Donald Davidson, that ‘is \( T \)’ applies to only the true sentences of \( L \) may not be as immediately apparent to the rest of us. So, this claim calls for a bit of exposition. The idea is this: we know by schema (T) that if \( p \), then ‘\( s \)’ is \( T \). So suppose \( p \). Then there is some corresponding ‘\( s \)’ in \( L \) that is \( T \). Can we say anything more specific about this ‘\( s \)’ besides that it exists and that it is \( T \)? Why yes, we can. We know that ‘\( s \)’ must be true. We supposed \( p \), so \( p \) is true. Then since the expression “‘\( s \)’ is \( T \)” is related to ‘\( p \)’ by a biconditional, “‘\( s \)’ is \( T \)” and ‘\( p \)’ must have the same truth-value. So now we have that ‘\( s \)’ is \( T \) if and only if ‘\( s \)’ is true, thus establishing Davidson’s claim.

This seems to suggest that the predicate ‘is true’ is a strong candidate for ‘is \( T \)” (which should not come a surprise to anyone who has encountered Tarski, as Davidson’s schema (T) bears a striking and intentional resemblance to Tarski’s Convention T). Davidson gives us another reason to suppose that the predicate ‘is \( T \)” is actually the predicate ‘is true.’ A definition of the predicate ‘is true’ taking the form of Tarski’s Convention \( T \) (‘\( s \)’ is true in \( L \) if and only if \( p \)’) pairs each sentence of the object language with its truth conditions. That is, it tells us what it is for any sentence in the object language to be true. And this, Davidson suggests, amounts to understanding a language.22 And since specifying the truth conditions of an expression enable us to understand the expression, they also amount to specifying what the expression means.

---

21 Ibid.
22 Ibid., 24.
So we have finally arrived at Davidson’s first proposal: that “a theory of meaning”\textsuperscript{23} for a language $L$ shows ‘how the meanings of sentences depend upon the meanings of words’ if it contains a (recursive) definition of truth-in-$L$.\textsuperscript{24} Our next step will be to see if this proposal pans out – does a recursive definition of truth-in-$L$ actually succeed in specifying the meanings of complex expressions of a language based on the meanings of their parts?

1.3 The Initial Project

One of Lepore and Ludwig’s most insightful interpretive claims about Davidson’s work on meaning is that Davidson actually ended up being engaged in two projects at once. Davidson set out with the task of simply developing compositional meaning theory. All the theory needed to do was show how the meanings of complex expressions depended on the meanings of their parts. Davidson did not need to confront the issue of how these “parts” came to be meaningful as well in order to be successful at his original task. This is what Lepore and Ludwig term Davidson’s “initial project.”\textsuperscript{25} They suggest, however, that confusion has arisen in interpretation of Davidson because Davidson got ahead of himself. At some point in putting forth his proposal with regard to the initial project, Davidson must have seen how this same proposal could be extended to enable us to specify what all expressions of a language mean. Davidson then began pursuing this project as well, which Lepore and Ludwig term the “extended project”\textsuperscript{26} without making it explicit that his aims had changed (perhaps Davidson himself was not fully conscious of the switch). However, it is clear that at least at some point, all Davidson had in mind was

\footnotesize
\textsuperscript{23} The distinction between “theory of meaning” and “meaning theory” mentioned above was introduced by Lepore and Ludwig, and is not one that Davidson himself made. To be consistent with the terminology of this paper, one should understand this as “meaning theory.”
\textsuperscript{24} Ibid., 23.
\textsuperscript{25} Lepore and Ludwig, Donald Davidson, 22.
\textsuperscript{26} Ibid.
providing the machinery of a compositional meaning theory, as evidenced when he states “the task was to give the meaning of all expressions in a certain infinite set on the basis of the meaning of the parts; it was not in the bargain also to give the meanings of the atomic parts.”

At this point, we have already seen much of Davidson’s proposal with regard to the initial project – it is that a properly constrained truth-theory provides the machinery of a compositional meaning theory. All we have left to do in service of the initial project is to show that this proposal actually works. We will do this by first giving an example of a truth theory of the form in question for a much simplified version of French. Then we will discuss what constraints needs to be applied to this truth theory in order to guarantee that we can infer from the theorems of the truth theory (‘s’ is true in L if and only if p) to theorems of a meaning theory (‘s’ in L means that p).

Lepore and Ludwig give an excellent example of a truth theory for a simple language (which they call Simple English) to show how Davidson’s truth theory/meaning theory proposal works. Since giving this example is so essential to showing Davidson’s success in the initial project, I will largely reproduce their example here. I want it to be clear that this example is essentially not my own creative work. I will, however, make a slight alteration. Lepore and Ludwig’s Simple English is meant to be a different language from English, but it uses English expressions. This can be slightly confusing, since it makes the theorems of the truth theory appear trivial. So, in providing this example, my language will use words in French instead.

We will call the language for which we want to define truth “Simple French.” I will follow Lepore and Ludwig in putting expressions in Simple French in a different typeface, so it is even easier to distinguish between our object language (Simple French) and our metalanguage.

---

28 Lepore and Ludwig, Donald Davidson, 66-74.
(English). Simple French consists of two names, “Claude” and “Michel,” one untensed predicate, “est gentil,” and two logical operators, “pas,” and “et.” Expressions in Simple French consist of finite strings of these symbols. An atomic expression of Simple French consists of a name followed by a predicate: either “Claude est gentil,” or “Michel est gentil.” Complex expression in Simple French are made by connecting atomic expressions according to the following two rules:

1. If $x$ is an expression in Simple French, then its negation, “Pas:$x$” is an expression in Simple French (ex. “Pas:Claude est gentil”)

2. If $x$ and $y$ are expressions in Simple French, then their conjunction, “$x$ et $y$” is an expression in Simple French (ex. “Claude est gentil et Michel est gentil”).

Now we need to stipulate the axioms for our truth theory. The axioms I use here are going to draw on what I know these French words to actually mean in English. This will play an important role in the discussion of what constraint needs to be put on the truth theory for it to work as a meaning theory, which we will come to below. We will have two types of axioms: base axioms, for the referents of names and the truth conditions of atomic expressions, and recursive axioms, for the truth conditions of complex expressions. For the purpose of this example, let “ref($x$)” stand for “the referent of $x$” and “true$_F$” stand for “is true in Simple French.”

**Base Axioms**

(A1) Ref(Claude) is Claude.

(A2) Ref(Michel) is Michel.
(A3) For all names \(x\), “\(x\) est gentil” is true\(_F\) if and only if \(x\) is nice.\(^{29}\)

**Recursive Axioms**

(A4) For all expressions \(y\), its negation “\(\text{Pas:}\ y\)” is true\(_F\) if and only if it is not the case that \(y\) is true\(_F\).

(A5) For all expressions \(y, z\), their conjunction “\(y\ et z\)” is true\(_F\) if and only if \(y\) is true\(_F\) and \(z\) is true\(_F\).

The first thing to point out about these axioms is that the recursive axioms determine the truth conditions of a complex expression based on the truth conditions of its parts. Thus this theory is compositional – a promising sign. We will now use these axioms to give a sample proof of a theorem in our truth theory for Simple French. This theorem will specify, entirely in the metalanguage, under which conditions the expression is true\(_F\). In order to have our truth conditions entirely in the metalanguage, any expressions of Simple French and the predicate “true\(_F\)” cannot appear on the right side of the biconditional. Therefore we must keep applying our axioms recursively until these are eliminated. Once all traces of Simple French and “true\(_F\)” have been eliminated from the right hand side of the biconditional, we know we have reached our theorem, because we can no longer apply any axioms.

Let us take “Claude est gentil et Pas:Michel est gentil” as our expression. The proof of its truth conditions goes as follows:\(^{30}\)

---

\(^{29}\) Since the predicate “est gentil” is untensed in Simple French, its metalanguage counterpart (“is nice”) must also be untensed. So let us stipulate that \(x\) is nice if at some time \(t\) in \(x\)’s life, \(x\) has done or will do something nice.

\(^{30}\) For our proof, we will also need some inference rules in order to apply the axioms. For this proof, we will need to specify two inference rules. First, let an expression be universally quantified if all the variables in the expression are quantified by “for all.” Now our rules: (1) Universal Instantiation: from any expression universally quantified over the variable \(x\), we may infer the expression that replaces \(x\) with any other expression. (2) Substitution: for any expression containing a name in Simple French, we may infer the expression that has replaced the name with its metalanguage referent.
(1) By A5, “Claude est gentil et Pas: Michel est gentil” is true\(_F\) if and only if “Claude est gentil” is true\(_F\) and “Pas: Michel est gentil” is true\(_F\).

(2) By A4, “Claude est gentil et Pas: Michel est gentil” is true\(_F\) if and only if “Claude est gentil” is true\(_F\) and it is not the case that “Michel est gentil” is true\(_F\).

(3) By two applications of A3, “Claude est gentil et Pas: Michel est gentil” is true\(_F\) if and only if “Claude” is nice and it is not the case that “Michel” is nice.

(4) Finally, by A1 and A2, “Claude est gentil et Pas: Michel est gentil” is true\(_F\) if and only if Claude is nice and it is not the case that Michel is nice.

Now that we have the truth conditions of our expression, our question becomes: when can we be sure that we are justified in inferring to the meaning of our expression based on these proven truth conditions? Tarski knew that in order for a truth definition such as the one proposed above to be adequate, the predicate “is true” as specified by the definition must apply to all and only the true sentences of the object language. Tarki insightfully provided a criterion for determining the adequacy of the truth theory: a truth theory for a language has all and only the true sentences of that language as its extension if and only if the theorems of the theory include all sentences of the form “s is true if and only if p” where ‘s’ is replaced with a structural description of an expression in the object language and ‘p’ is replaced with a translation of that expression into the metalanguage. This works because translation preserves meaning, and what an expression means, in relation to the world, determines that expression’s truth value. So object
language and metalanguage expressions that are the same in meaning (are translations of each other) must also be the same in truth value.\textsuperscript{31}

This leads directly into the constraint that must be put on a truth theory for it to give the meanings of expressions as well as their truth conditions. How do we guarantee that, when the truth conditions proof is finished, the right side of the resulting theorem will be a translation of the left? Since in proving the theorem, the only things we draw on are the content of the axioms, the right side will translate the left if the axioms provide metalanguage interpretations of the basic object language expressions – that is, if the axioms give the meanings of the object language expressions in the metalanguage. Since we draw only on the content of the axioms in each step, if the axioms are interpretive, each step of the proof will preserve meaning as well. So having interpretive axioms guarantees that the truth theory will be adequate, and thus that the right side of the theorem provides a translation of the object language expression on the left into the metalanguagae.\textsuperscript{32} This is essential for using our truth theory as a meaning theory because it justifies the inference from a truth theory theorem (‘s’ is true if and only if p) to a meaning theory theorem (‘s’ means that p).\textsuperscript{33}

We have not forgotten that the most important thing we are looking to satisfy with the use of the truth theory is the compositionality constraint. Each step in the proof shows how the truth conditions of the parts of the expression contribute to the truth conditions of the whole. The truth conditions of the parts are determined by the axioms, so if the axioms are interpretive, the process of the proof will also demonstrate how the meaning of the whole expression is built on the meanings of the parts, and how they are combined. So it is in the proof that we see the compositionality.

\textsuperscript{31} Lepore and Ludwig, \textit{Donald Davidson}, 71.
\textsuperscript{32} Ibid., 117.
\textsuperscript{33} Ibid., 120-121.
We can now put all this together and see Davidson’s vision for providing a solution to the initial project. Let us call a truth theory whose axioms are interpretive an interpretive truth theory. The constraint that needs to be put on a truth theory in order for it to work in service of a meaning theory is simply that the truth theory be interpretive. Then a meaning theory for a language $L$ can be said to consist of:

1. the axioms of an interpretive truth theory
2. the additional axiom that if “‘s’ is true in $L$ if and only if $p$” is a theorem proven using only the axioms of the truth theory, then ‘s’ means that $p$.

If we go back to our Simple French example, we can turn our truth theory into a meaning theory by stipulating that (A1)-(A5) are interpretive and adding the additional axiom:

(A6) If “‘s’ is true$_F$ if and only if $p$” is a theorem proven using only (A1)-(A5), then ‘s’ in Simple French means that $p$.

If we reconsider our Simple French expression “Claude est gentil et Pas:Michel est gentil,” we can take the proof steps (1)-(4) from above, plus the additional premise:

(5) the biconditional in step (4) is a theorem proven using only (A1)-(A5)

and conclude by (A6) and (1)-(5) our first meaning theory theorem:

(6) “Claude est gentil et Pas:Michel est gentil” in Simple French means that Claude is nice and it is not the case that Michel is nice.

Further, steps (1)-(4) demonstrate how the meaning of this expression is made up from the meanings of its parts.

This concludes Davidson’s initial project. I think his proposal that a properly constrained (i.e. interpretive) truth theory can be used as a compositional meaning theory is very successful,
as I hope to have shown above. However, the reader is likely left unsatisfied with the fact that thus far we have simply stipulated that the truth theory is interpretive. While this is an acceptable move given the aims of the initial project, it is disappointing to those who wish to know how one could come to understand a language in its entirety – *including* its basic expressions – given no initial information. Davidson himself probably felt the same way, which is why his work on the extended project quickly eclipsed the work on the initial project. What we have just shown is that if we stipulate that the axioms of a truth theory for a language are interpretive, then the truth theory works as a meaning theory for the language. So the task at hand is to figure out how to construct the axioms of an interpretive truth theory that does not simply assume that the axioms are interpretive. This is the task of Davidson’s extended project is intended to accomplish just this task. In this part of the project, the question Davidson seeks to answer is how one could construct and confirm a truth theory for a completely foreign language in a way that would guarantee it to be interpretive.

### 1.4 The Extended Project

The result of Davidson’s work on the extended project is his development of the procedure termed “radical interpretation.” The thought is that the radical interpreter, who enters a community of speakers of a language entirely foreign to him, can come to understand this foreign language by constructing a truth theory for the language based solely on empirical evidence. The radical interpreter, of course, can have no prior knowledge of what any of the speakers’ utterances mean, and thus cannot be assumed to know anything about the speakers’ beliefs. So the only evidence the radical interpreter has available is the behavior of the speakers – primarily the acts of uttering and giving assent or dissent – and under what environmental
conditions these behaviors occur. Davidson believes that based solely on this evidence, it is possible to confirm a truth theory for the language that is guaranteed to be interpretive.

The way the radical interpreter goes about this is by determining to the best of his ability under what conditions a speaker holds an expression true. The radical interpreter might do this by first observing that the speaker only utters the expression “x” when the set of environmental conditions Y obtain. So, the next time Y obtains, the radical interpreter might ask the speaker, “x?” and see if the speaker assents or dissents to this inquiry. If the speaker assents, the radical interpreter can make a tentative pairing between the expression “x” and the conditions Y – “For a speaker S, ‘x’ is held true if and only if Y.” Lepore and Ludwig term these “hold-true” biconditionals “L-sentences.” If the speaker dissents, however, the radical interpreter knows that the set of conditions Y is either not sufficient for a speaker to hold “x” true, or there is something extra included in Y that negates the truth of “x.” In this case, the radical interpreter should continue to refine the set of conditions to be paired with “x” by further observation and testing until he has seemed to determine the exact set of conditions under which a speaker of the language will assent to “x.”

Once the radical interpreter has several well-confirmed L-sentences in his arsenal, he can start constructing a tentative truth theory. The idea is that the radical interpreter will use the content of the L-sentences he has empirically confirmed to infer target theorems of the truth theory. By taking them all together, the radical interpreter can start to look for patterns in the parts of an utterance and the conditions they are paired with to begin to suggest axioms giving the truth conditions of these parts. For instance, the radical interpreter may observe that whenever an expression x is paired with another expression, the conditions under which a

---

35 Lepore and Ludwig, Donald Davidson, 184.
36 Ibid., 196.
speaker assents to the pair are the absence of the conditions under which the speaker would assent to the expression alone. From this, the radical interpreter might hypothesize that $x$ functions as a negation, and construct an axiom for $x$ as such.

In order for the content of the $L$-sentences to be used as target theorems for the truth theory, two things need to be established about them. First, a biconditional can only be a theorem of a truth theory if the predicate on the right hand side is “is true.” So, the radical interpreter needs to be able to infer that the conditions under which an expression is held true are the same conditions under which the expression is true. Then the radical interpreter can infer a “$T$-Form sentence”\textsuperscript{37} that is the same as the $L$-sentence, except in the $T$-Form sentence, “held true” is replaced with “is true.” Second, the radical interpreter needs to be sure that these $T$-form sentences inferred on the basis of the $L$-sentences are interpretive. Otherwise, any truth theory that generates these $T$-form sentences as theorems is not guaranteed to be an interpretive truth theory, which is necessary. So, the radical interpreter needs to know that the conditions he is pairing expressions with give the meaning of the expression – the conditions capture the content of the expression.

Davidson introduces his infamous “Principle of Charity” to accomplish this double task. What Davidson actually thinks the Principle of Charity includes is somewhat ambiguous, since he formulates differently at different times, sometimes in terms of the truth of a speaker’s beliefs, and sometimes in terms of agreement between a speaker’s beliefs. Lepore and Ludwig offer an in-depth examination of the different formulations of Charity and argue for one in particular, giving reasons why the other interpretations fail.\textsuperscript{38} They further argue that Charity itself is not enough to justify the move from being held true to being true, and introduce a improved version,

\textsuperscript{37} A bit of terminology – Lepore and Ludwig use “$T$-form sentence” to refer to any sentence of the form “‘$s$’ is true if and only if $p$.” The term “$T$-sentence” is reserved for $T$-Form sentences that are interpretive.

\textsuperscript{38} See Ibid., 182-196.
which they call “Grace,” that they believe is sufficient. Since Charity is not the focus of my project, I will not go into the details of this discussion, but rather simply present it in its strongest form. This will include elements of Lepore and Ludwig’s “Grace,” but I will continue to use Davidson’s terminology, and call what is covered here “Charity.”

To see what we need Charity to accomplish, the first step is to observe that hold true attitudes result from two factors: what the speaker believes and what his expressions mean. When a speaker has a belief, he will hold the expression that expresses the content of this belief to be true. This is simply what it is to believe something. The content of the expression is its meaning, so in this case, the meaning of the expression is the content of the belief. A hold-true attitude with regard to an expression must depend, therefore, on both the speaker’s beliefs, and the meaning of the expression. In order to use hold-true attitudes to isolate meaning, then, the radical interpreter needs a way to hold the speaker’s beliefs constant.

The radical interpreter has to, in a way that will be made precise below, assume some of the speaker’s beliefs to be true. If we allow belief to fluctuate, that is, if do not know whether a speaker’s beliefs are true or false, we are not justified in assuming that the hold-true conditions of an expression give the content of the expression. For if the speaker’s belief is false, he is assenting to the expression under conditions that improperly capture the expression’s meaning. For example, if I falsely believe that there is a cup on my desk, and the radical interpreter asks me “there is a cup on your desk?” I will assent. The radical interpreter will then record all the conditions of my assent, which will not include a cup being on the desk. These conditions clearly do not capture the meaning of the expression. Therefore, if it is possible that a speaker’s beliefs are false, the hold-true conditions of the expression expressing this belief cannot be trusted to capture meaning of the expression. Further, and more straightforwardly, if it is

---

39 Ibid., 188.
possible that the speaker’s beliefs are false, we are not justified in inferring a $T$-form sentence based on the $L$-sentence, since the fact that the expression is held true does not guarantee that the expression is true under these conditions.

Therefore, in order for the radical interpreter to be able to move forward, he will have to make an assumption that some of the speaker’s beliefs are true. This idea at first probably seems abhorrent to any philosopher, as it is so often the case that people’s beliefs are not actually true. It would indeed be foolish to assume all the beliefs of a given speaker are true. However, if we limit the type of beliefs that are assumed to be true to a set that are quite likely to actually be true, then making this assumption will not be so offensive. The set of beliefs suggests itself fairly readily if we consider what the radical interpreter is actually doing. Since the radical interpreter has to pair expressions with the environmental conditions under which they are assented to, this task will be accomplished most easily for expressions that a speaker sometimes holds true and sometimes holds false in response to changes in the conditions of his environment. These expressions are the contents of beliefs that are highly context-sensitive, and usually about the actual conditions prompting them. For example, these are expressions like “the cup is on the table,” “it is snowing right now,” “the lamp is on,” etc.

Barring any significant oddities or abnormalities, it is highly likely that a speaker’s beliefs about these types of matters are true. All things being equal – that is, the speaker is an average, sane human being with fully functioning perceptual apparatus, and nothing is being done to deliberately deceive the speaker – it is reasonable to assume that a speaker will have largely true beliefs of this type. If this is right, then it is not so offensive to apply Charity to these beliefs, and assume for the purpose of interpretation that a speaker’s beliefs of this type are in fact true.

40 Ibid., 182.
Now we can formally characterize Charity as consisting of the following assumptions:\(^{41}\)

(C1) A speaker holds true an expression \(s\) under a set of environmental conditions expressed by the sentence ‘\(p\),’ because and only because he knows that \(x\) means that \(p\), he or she believes that \(p\), and knows that \(s\) is then true if and only if \(p\).\(^{42}\)

(C2) The conditions that \(L\)-sentences identify as the conditions under which a speaker holds an expression true are also the prompting conditions of the belief that the expressions express.\(^{43}\)

(C3) A speaker’s beliefs prompted by conditions in his or her environment, by and large, are about these prompting conditions, and are true.

Does this characterization of Charity accomplish what we want it to? Does it guarantee that the conditions under which a speaker holds an expression true are the conditions under which the expression \(is\) true (allows us to infer \(T\)-Form sentences from \(L\)-sentences), and that these conditions also give the content of the expression (guarantees that the inferred \(T\)-Form sentences will be interpretive)? (C2) tells us that the conditions under which an expression is held true are the prompting conditions of the belief it expresses, and thus by (C3) give the content of the expression. This is enough for interpretiveness. (C3) also tells us that a speaker believes that \(p\) (an environmental prompted belief) if and only if \(p\). Then by (C1), the speaker holds true the expression \(s\), which expresses the content of his or her belief that \(p\), if and only if

\(^{41}\) Lepore and Ludwig emphasize that the assumptions in Charity are \textit{ceteris paribus} assumptions. Of course there are always a manifold of intervening conditions that one could dream up which would negate the validity of these assumptions, for example the speaker is being deceived, the speaker is insane, or whatever else could go wrong. If one were to actually carry out the process of radical interpretation in real life, these would be real concerns that the radical interpreter would have to look out for. However, radical interpretation is presented theoretically, and as such we can assume idealized conditions. So for (C1)-(C3), note that they are made assuming all these other potential factors are held constant.

\(^{42}\) Ibid., 195.

\(^{43}\) Ibid., 194.
Therefore, by (C1) and (C3), the conditions under which the speaker holds an expression of this type true are also the conditions under which the expression is true. So (C1)-(C3) have given us what we needed to move forward.44

The rest of the process of radical interpretation is as follows.45 Justified by Charity, the radical interpreter infers T-sentences (interpretive T-Form sentences) based on the empirically developed L-sentences. He then uses these T-sentences as the target theorems of the truth theory, looking for structural patterns between parts of expressions and paired conditions that will allow the radical interpreter to suggest tentative axioms giving the truth conditions of these parts. Once the radical interpreter has obtained several tentative axioms, he uses these axioms to derive new theorems – T-sentences other than the T-sentences he or she used to construct the tentative theory. For example, if the radical interpreter uses the expressions “the cat is on the mat” and “the dog is on the couch” to construct the tentative truth theory, the new theorems he proves might give the truth conditions of “the cat is on the couch” and “the dog is on the mat.” The radical interpreter takes these new T-sentences, which give the truth conditions of expressions for which he does not yet have an L-sentence, and tests them by checking whether the speakers assent or dissent to the expression in a pattern that would be predicted by the truth conditions given by the hypothesized theory.

If the speakers assent and dissent entirely as predicted, the truth theory has been confirmed, and the work of the radical interpreter is done. In the initially more plausible case

44 I am aware that I have glossed over Charity a bit. The reader may not be sufficiently convinced that the assumptions involved in Charity are truly warranted, or that they give us all we really need to continue with radical interpretation. This reader would certainly not be alone; Davidson’s principle of Charity is extremely controversial, particularly because he uses the truth of Charity to argue for several additional claims in other areas. However, since Charity is essential for radical interpretation to even get off the ground, and the objection I focus on is in regards to a different aspect of radical interpretation, I am not able to confront the issues surrounding Charity here. It should simply be noted that any conclusion I may reach in support of radical interpretation would be conditional on the validity of Charity as well.

45 See Ibid., 196 for a succinct outline of the stages of radical interpretation.
that the speakers assent and dissent not quite as predicted, the radical interpreter records the patterns of assent and dissent to these new expressions as further evidence, uses them to refine the tentative truth conditions, and constructs $L$-sentences for them based on this evidence. He does with several new expressions, and then repeats the above procedure – inferring the corresponding $T$-sentences, and using these $T$-sentences as the truth theory’s target theorems, adds to and refines the tentative axioms. Then the radical interpreter again derives new $T$-sentences from the refined axioms, and again tests with speakers. The radical interpreter repeats this procedure over and over until he has found axioms that yield $T$-sentences that predict the assent and dissent patterns of speakers to a very high degree of accuracy. The truth theory has then been confirmed for the language, and can be used in a meaning theory as specified above.

In this way, the radical interpreter comes to be able to discern the structure of the foreign language, and the meanings of its most basic expressions, theoretically enabling the understanding of any expression in the language, thus accomplishing the goal of the extended project. We have now seen Davidson’s proposal for a meaning theory in full. I find it to be an extremely impressive proposal. However, given the complicated nature of the subject matter it is dealing with, and the breadth of the proposal, it is naturally controversial in many aspects. In what follows, I aim to contribute to the discussion about the success of Davidson’s proposal by looking at one particular threat to the success of radical interpretation.
Chapter 2
The Ambiguity Problem and the Triangulation-Based Solution

This chapter will focus on the issue that arises during radical interpretation that I will be considering – the ambiguity problem. As a foundation for understanding the ambiguity problem, we need to first take a brief look at the phenomenon of underdetermination. I will then proceed to spell out the ambiguity that arises, and why it is an issue for radical interpretation. I will then argue that Davidson’s triangulation can provide a solution to this problem.

2.1 Underdetermination

A set of evidence underdetermines a theory confirmed on the basis of this evidence if multiple, incompatible theories can be confirmed equally well by the evidence.\(^{46}\) Let T\(_1\) and T\(_2\) be two theories accounting for a body of evidence E. We will stipulate that T\(_1\) and T\(_2\) are incompatible. E underdetermines T\(_1\) if T\(_2\) accounts for E equally as well as T\(_1\) does, and vice versa. Note of course that underdetermination is always relative to a specified body of evidence.\(^{47}\) T\(_1\) and T\(_2\) may be underdetermined in relation to E, but when further evidence E’ is incorporated, one theory may be shown to be more adequate than the other. For instance, T\(_1\) may account for E’ while T\(_2\) does not. In this case, although the two theories are underdetermined relative to E, they are not underdetermined relative to E’ – T\(_1\) has been shown to be more adequate than T\(_2\) when E’ is our body of evidence.

\(^{46}\) Ibid., 223.
\(^{47}\) Ibid.
Radical interpretation is a procedure for constructing an interpretive truth theory that accounts for the behavioral assent/dissent patterns exhibited by speakers of a language. If multiple, incompatible truth theories can account for the radical interpreter’s evidence equally well, by the above reasoning, this would suggest that radical interpretation fails. This is ultimately the point ambiguity problem seeks to show – it claims that two incompatible truth theories can be confirmed using the process of radical interpretation and based on the radical interpreter’s evidence. The radical interpreter’s evidence thus underdetermines any truth theory. If this is the case, any truth theory confirmed by radical interpretation cannot be guaranteed to be interpretive, and therefore radical interpretation fails to yield a meaning theory for a language.

2.2 The Ambiguity Problem

The ambiguity problem arises when constructing our $L$-sentences. Recall from Chapter 1 that $L$-sentences correlate an utterance with the environmental conditions under which it is held true. Further, in order for the interpreter to be able to infer a $T$-sentence (a sentence that gives the interpretive truth conditions of an utterance), from an $L$-sentence, we need to be confident that the belief expressed by the utterance is caused by and about these conditions. When discussing Charity, I suggested that this is plausibly the case when considering beliefs about our environment.

However, it is not the case that there is only one set of environmental conditions that might be paired with an utterance. There is an ambiguity about what conditions are relevant to an utterance. There are actually at least two types of conditions that we can identify: distal stimuli, which are actual objects and events in the world, and proximal stimuli, which are patterns of stimulation the speaker undergoes in the presence of distal stimuli. Which set of
stimuli constitutes the relevant cause? Since the distal stimuli and the proximal stimuli occur at just the same times, any correlation of an utterance with the distal stimuli is equally a correlation of the utterance with the proximal stimuli. Thus the radical interpreter can actually confirm two sets of \( L \)-sentences from the evidence: one that specifies the hold-true conditions of an utterance as the distal stimuli, and one set that specifies the hold-true conditions of an utterance as the proximal stimuli.

This is an issue, because if we can equally confirm two incompatible sets of \( L \)-sentences, we are going to be left with the underdetermination of theory by the evidence. By applying Charity to each set of \( L \)-sentences, we can obtain two sets of \( T \)-sentences. From these two sets of \( T \)-sentences, we can project to two different truth theories for the speaker’s language. These truth theories will have differing axioms since they are projected from different sets of \( T \)-sentences. Since their axioms differ, the theorems of these truth theories will differ as well. Thus we will be left with two sets of theorems that give different truth conditions for the same utterance. Therefore the two truth theories are incompatible. Since we can equally confirm two incompatible theories, both theories are underdetermined by the evidence.

If this is right, the underdetermination of these two theories would show that radical interpretation is unsuccessful. For since we can equally come to two sets of truth conditions, we cannot be sure that either one or the other set is interpretive. Therefore, we are not justified in inferring meaning theorems from the theorems of either truth theory. Radical interpretation, then, has failed to give us a meaning theory for the language of our speaker.

A successful solution to the ambiguity problem would have to give us some way to determine the relevant stimuli – distal or proximal – in the environmental conditions correlated with an utterance. If we can come up with some way of saying for certain what the relevant
cause is, one of the sets of $L$-sentences will no longer be compatible with the evidence. This would eliminate the underdetermination of theory by evidence, and bolster the case that radical interpretation can be successful in producing a meaning theory for a language.

2.3 Davidson’s Triangulation Thesis

Davidson is well aware of problem of ambiguity between the distal and proximal stimuli. Davidson asserts that if it were possible for a creature to exist without having ever engaged in communication, and we were to observe this creature responding to its environment, no matter how complex these responses were, they could never be taken to demonstrate that the creature is reacting to or thinking about events “a certain distance away [distal stimuli] rather than, say, on its skin [proximal stimuli].”\textsuperscript{48} This is precisely because of the same ambiguity noted above – the creature’s responses could equally be correlated with distal and proximal events. Thus we cannot tell which type of stimulus the creature is responding to. Further, if we were to take the creature’s responses as indicative of the contents of its thoughts, we would have no way to tell what type of stimulus constitutes these contents, based on the observed correlations of response and conditions of response.

By “thoughts,” Davidson means the intentional states, or propositional attitudes, of a creature – beliefs, desires, and so on.\textsuperscript{49} Intentional states, in order to be intentional, must be directed, or about something.\textsuperscript{49} So if we in principle are unable to specify what the contents of the creature’s thoughts are, this would mean that the creature does not have thoughts at all, in the sense that the creature does not have true intentional states. It would be natural to object at this point that while we cannot determine what the creature’s thoughts are about from the third

\textsuperscript{49} Ibid., 165.
person perspective, this does not prevent the creature itself from knowing what its own thoughts are about, from the first person perspective. If this were the case, the creature could be said to have thoughts even if the contents were indeterminate from the standpoint of an observer. However, this would require the creature to be able to have the concept of an object – the stimulus he is responding to – independent of any communication. Davidson thinks this is impossible.\(^{50}\)

Davidson thinks a solitary speaker cannot have the concept of an object independent of all communication for the following reasons. In order to have the concept of an object, one must be able to understand that the concept can be misapplied.\(^{51}\) For example, a creature could not believe it sees a giraffe if it did not know “that some things are correctly identified as giraffes and some things are not.”\(^{52}\) If a creature does not understand this, it cannot be said to have the concept of a giraffe, because it does not understand giraffes to be things that exist independently of its perceptions of them. But if this were so, then the creature’s thoughts would not actually be about actual giraffes at all. The creature’s thoughts that seem to be about giraffes would actually have as their content whatever seems to the creature to be a giraffe. But this could be anything at all. So, if a creature does not have the concept of an object – understanding that the concept can be misapplied – then any of the creature’s thoughts could have anything as its content, which is to say that it would have no determinate content.

How does one acquire this understanding of the potential for misapplication? Davidson believes it is only through communication with another creature. For in order to understand that one’s concepts can be misapplied, one has to have the experience of one’s responses not

---

\(^{50}\) Davidson acknowledges his debt to Wittgenstein in the genesis of many of these views.  
\(^{52}\) Davidson, “Responses,” 698.
matching up with what is actually the case. One cannot encounter this alone, because there is no source of comparison – one would not respond to an event in a certain way if one did not think that it was the correct response. One will continue to do this unless one observes another creature that has responded to the same event in a different way. Only then does understanding of the possibility of error enter the picture. And it is only with an understanding of the possibility of error that an understanding of an inter-subjective world can enter the picture.

This leads us nicely into describing the actual triangle that constitutes “triangulation.” The three points of the triangle are two speakers and an object or event in the world. Each speaker is simultaneously interacting with the world and with the other speaker. Each speaker recognizes what appear to be similarities in the world, and correspondingly responds to the world in similar ways. Each speaker also recognizes similarities in the responses of the other speaker, in correlation to events in the world. In short, “each creature learns to correlate the reactions of the other with changes or objects in the world to which it also reacts.”

How can this triangle be said to determine the contents of thoughts? First off, engaging in this type of triangulation fixes the cause of the responses. Consider the case of a speaker responding to the presence of a table by uttering the word “table.” For this single speaker, both actual tables, as well as patterns of stimulation caused by tables are will be similar events that are correlated to the similar response of the utterance of “table.” Again, this is the ambiguity issue we are ultimately concerned with.

However, say we introduce another speaker into the situation, who also responds to the presence of tables by uttering “table.” Suppose both speakers recognize that their own responses and the responses of the other speaker are the same. This causes the speakers to hypothesize that

---

53 Davidson, “The Emergence of Thought,” 12.
54 Ibid.
they are reacting to the same object. The only way to know this is to try to communicate. If they are successfully in communication about the object they both have in mind, then it can be said that this object is a shared cause of their responses. Whatever this shared cause is, then, must be able to account for the responses of both speakers. This rules out the proximal stimuli as what the speakers are thinking about. Proximal stimuli are by definition not shared and so the proximal stimuli experienced by either speaker cannot account for the responses of both speakers. Since they have communicated, their thoughts have a shared object, and thus the distal stimuli are the only candidates for the cause, because they are the only stimuli that are shared between speakers.

Through the process of triangulating, then, speakers come to a determinate object to which they are responding. And then if we take their responses as indicative of the contents of their thoughts, we are left with the distal stimuli as the determinate contents of their thoughts.

Secondly, by participating in this triangle, a speaker can obtain the concept of an object. This is because by locating the common cause in the world, the two speakers together identify what that cause is as the intersection of their responses. Recognizing the existence of a triangle with the object in question as the intersection of similar responses of two speakers allows one to understand the potential for error – in some situation, the triangle could break down. In that case, one of the speakers is not responding in the appropriate way. Thus, identifying the cause as the intersection of the responses is sufficient for obtaining the concept of an object.

Note that when I say “first off,” and “secondly,” I do not mean these terms in a temporal sense. The two tasks of triangulation are not temporally situated, with one coming before the other. Rather, the reader may have noticed that the two tasks triangulation accomplishes go hand in hand – they occur, and must occur, simultaneously. See Claudine Verheggen, “Triangulating with Davidson,” The Philosophical Quarterly 57, no. 226 (January, 2007): 96-103 for an extended discussion of this simultaneity.

One point of clarification – the phrase “concept of an object” might be a bit ambiguous. On some views of concepts, one can be said to have the concept “table,” for example, if one is simply capable of using the term table to refer to things. This is not the “concept of an object” that I am meaning to use here. For it seems that one can have this type of concept without communication – simply by learning to correlate responses with environmental events. This type of concept is not sufficient for knowing that one’s concepts can be misapplied, however. What is needed to know that one’s concepts can be misapplied is a second-order concept of sorts. This second-order concept is a concept of the first concept “table” – it tells me propositional information about what a table is and thus under what conditions “table” is applied appropriately and not appropriately. And it is this second-order concept that we cannot acquire without communication. This seems clear if we consider a situation, as Davidson often does, in which a child is learning a new word. When a child acquires the ability the ability to use the world “table,” he has the first order concept. But the child might wildly misuse table initially (children often under or over apply terms as they learn language). The child learns the conditions under which “table” can be asserted through being corrected or rewarded by other speakers. Through this process of trial and error, the child slowly acquires the concept of what a table is, and thus learns that the concept can be misapplied.

We have now gone through the details of triangulation in what has hopefully been an adequately clear manner. In summary, the argument from triangulation is this:

(T1) A creature’s thoughts (propositional attitudes) have determinate objects only if it engages as a speaker in triangulation with another speaker in communication about a common object of thought.

---

57 See Ibid., as well as Davidson, “The Emergence of Thought, and Davidson, “The Second Person.”
58 Adapted from Lepore and Ludwig, Donald Davidson, 406-407.
(T2) Nothing can have propositional attitudes unless there can be determinate contents of these attitudes.

(T3) Therefore, nothing can have thoughts unless it engages as a speaker in triangulation with another speaker.

The argument for premise (T1) is that “thoughts” can only have determinate content if the creature has the concept of an object, and we can discern the relevant object (discerning between the distal and proximal stimuli). Triangulation is required for both of these tasks. Therefore triangulation is required to give “thoughts” determinate content.

2.4 Triangulation as a Response to the Ambiguity Problem

Triangulation serves as a response to the ambiguity problem by placing a constraint on what a creature can be taken to have in mind when it communicates (utters). In order for interpretation of a creature’s sounds to work at all, it is obviously necessary that the creature is interpretable – its sounds must be meaningful. This is to say that it is necessary that the creature has thoughts – beliefs, desires, etc., the contents of which its utterances express. Thus by engaging in radical interpretation, we are forced to assume that our subjects of interpretation are creatures who have thoughts (if they did not, there would be no content in their expressions to interpret). If what has just been said about triangulation is correct, it is then also necessary that we treat our subjects of interpretation as creatures who have triangulated, since engaging in triangulation is necessary for the determination of thought.

As necessarily treating our subjects as creatures who have thoughts, we are acknowledging that there is a determinate content of their thoughts, and thus a determinate content of their utterances that express these thoughts. As necessarily treating our subjects as

---

59 Ibid., 406.
creatures who have triangulated, the possibilities for the content of their thoughts and utterances are constrained to the distal stimuli, since this is what triangulation determines as the content of thoughts.

Therefore, if triangulation is indeed necessary for thought, it eliminates the ambiguity problem and the resultant underdetermination. For by constraining the possible content of our subject’s thoughts and utterances to the distal stimuli, the set of $L$-sentences that correlates utterances with their proximal stimuli is no longer confirmed by the evidence. As a result, we are left with only one set of $L$-sentences that is compatible with the evidence, and are thus no longer at risk for underdetermination due to the proximal/distal contrast.

2.5 An Additional Comment on Triangulation

Up to this point, I have taken Davidson’s proposals about meaning to be these: knowing what an expression means is the same as understanding that expression, and since knowing the truth conditions of an expression is sufficient for understanding the expression, truth conditions can be said to “give the meaning” of an expression. With triangulation now in the picture, another piece of Davidson’s ideas about meaning and thought can come more into focus.

I think Davidson develops his notion of what meaning is in his later, triangulation-related work by suggesting that an expression means what it is that a speaker intends to be understood by his use of the expression. If we combine this notion with Davidson’s triangulation thesis – that the meanings of expressions become determinate through triangulating – this implies that one cannot mean something that, in principle, cannot come to be understood by another speaker. One might also note this has the further implication that there can be no private languages.

---

60 Davidson, “The Second Person,” 112.
61 Again, Davidson acknowledges Wittgenstein as the source of or inspiration behind many of these ideas.
This aspect of Davidson’s view, or what I take to be an aspect of his view at least, will play a role in how I respond to some of the objections or problems for triangulation that I will introduce in the next chapter.
Chapter 3
Defending Triangulation

This chapter ultimately seeks to add to the argument that triangulation solves the ambiguity problem by defending triangulation against some objections that have been raised or could be raised against it. If it can be shown, for instance, that triangulation is not actually necessary for thought, or that it fails to accomplish what it purports to, triangulation would not be a viable line of response, and thus it would fail to solve the ambiguity problem. There exists a multitude of objections to triangulation, and so it would be impossible for me to confront every single objection in this paper. The objections I consider are related in that they all mainly seek to show that triangulation is not sufficient to determine what expressions mean. They will include an objection from Cathrine Talmage, a related objection one might make that is derived from Quine’s “gavagai” case, and an objection from Lepore and Ludwig.

3.1 Talmage’s Objection

Talmage’s claim is that triangulating does not actually fix the contents of thoughts, and thus the meanings of the utterances that express these thoughts, as it purports to do. If triangulating cannot actually give thoughts determinate content, then the argument (T1)-(T3) from section 2.3 is unsound because premise (T1) is false. We will have then failed to establish that triangulating is necessary for thought, which is essential to triangulation serving as a solution to our ambiguity problem for radical interpretation.
Talmage’s argument is this:\(^{62}\) She asks us to consider a person uttering the word “table.” We are asked to suppose that this person is able to conceive of the stimulus of his utterance of “table” as an object in the world, and Talmage even is willing to grant that the person has acquired this ability by interacting linguistically with another person on a previous occasion. Granting all this, it is possible, Talmage argues, that the person conceives of the stimulus of his utterance of “table” as an object, but not \textit{as a table}. He could, for instance, conceive of the stimulus as “a horizontal board with four supporting legs.”\(^{63}\) If this is so, it is clear that this person’s utterances of “table” have a meaning – their meaning is the content of the person’s thought, which is a horizontal board with four legs. Further, it would be incorrect to say that the person’s utterances of “table” mean table, since the person would not apply the word to tables with more or less than four legs, for instance.

Now suppose there is a second person, let him be called B, who means table when he utters “table.” The issue is that B could triangulate with the first person, call him A, in the presence of a four legged table, and come to the incorrect interpretation of A’s utterances of “table.” In triangulating, A and B would locate the common cause of their similar utterances of “table” as the present four-legged table. This triangulation would thus posit the contents of both A’s and B’s thoughts as the same. But the contents of A’s and B’s thoughts are \textit{not} the same – the content of A’s thought is a four-legged horizontal board, while the content of B’s thought is a table. This shows, according to Talmage, that triangulation fails to \textit{actually} determine the contents of thoughts, and thus also the meanings of utterances, since the way it purports to “determine” them can determine them wrongly.


\(^{63}\) Ibid.
What Talmage says would be true, I think, if triangulation was meant to be a unitary phenomenon. However, I think this is a mistaken understanding of how triangulation actually occurs. My understanding of triangulation is that it is truly a process – speakers are meant to correlate their own responses with repeated occurrences of an event in the environment and with the repeated responses of another speaker. Without the event occurring multiple times, one cannot possibly justify a correlation. All that one could justify would be the statement that at one point in time, event E occurred in conjunction with response R. What we need in order to correlate event E and response R is multiple observations of E occurring in conjunction with R.

If this is right, Talmage’s objection carries no weight. For then it could not be the case that full-fledged triangulation would result in determining the contents of A’s and B’s thoughts as the same, when they are not. While initially it may appear that A and B mean the same thing by “table” as a triangle begins to emerge so long as they only jointly encounter four-legged tables, as soon as A and B are both in the presence of a non-four-legged table, they will realize that their responses are no longer similar – A does not respond by uttering “table” while B does. The emerging triangle thus breaks down, and it is clear that the meanings of their respective utterances of “table,” and thus the contents of their thoughts, are different.

Triangulation, I think, must be conceived of as this “honing in” on the common cause, because Talmage is quite right that triangulation fails to determine a definitive common cause as a unitary phenomenon – as one interaction. I cannot imagine that Davidson would have made this mistake. However, if Davidson did intend triangulation to be a unitary phenomenon, my response still works. In this case, I would frame my response not as explaining how Talmage’s objection misrepresents the actual process of triangulation, but rather as a reformulation of triangulation in order to overcome the issue Talmage points out. No matter what, Talmage’s
objection either becomes irrelevant, or able to be accommodated by strengthening the position triangulation represents.

3.2 Quine’s “Gavagai” Example

In a famous thought experiment, Quine asks us to imagine an interpreter trying to understand what a foreign speaker means by the expression “gavagai.” The interpreter has observed that the speaker assents to “gavagai?” if and only if the speaker is aware of the presence of a rabbit. So it seems natural to suggest that “gavagai” means rabbit. Quine, however, makes the claim that this is unwarranted. The speaker could mean something like “undetached rabbit parts” by “gavagai” and we would have no way to tell, since undetached rabbit parts are present just in case a rabbit is present.

While Quine does not employ this thought experiment as an objection to triangulation, the ambiguity between rabbit and undetached rabbit parts could be seen as a special, trickier case of the situation Talmage suggests. My response to Talmage hinged on the fact that the differences between meaning horizontal board with four legs by “table” and meaning table by “table” would become manifest in the usage of these terms. If confronted with a three-legged table, one speaker would assent to “table?” while the other would not. However, in the gavagai case, the difference in meaning would apparently never become apparent in the usage of the term. Whereas a three-legged table serves as a test case for whether a speaker means table or horizontal board with four legs by “table,” it seems that there is no test case that would reveal whether a speaker means rabbit or undetached rabbit parts by “gavagai.”

In trying to see how one might deal with this issue, the question that comes to mind is: do “undetached rabbit parts” and “rabbit” really mean something different? Of course the words

---

are different, but that does not stop the expressions from meaning the same thing. If they do not mean the same thing, it seems like in principle there ought to be some possible test case that would reveal this difference. Since what an expression means ultimately is, for Davidson, what a speaker can intend to be understood by it, meaning has to be public. So if the meanings of these two expressions are different, it seems to me that this difference must, in principle, be publically accessible.

If there is no possible case that shows these expressions to differ in meaning, I would assert that they are the same in meaning. For if there is no possible way that speaker could come to realize a difference in the usage of the two expressions, another speaker cannot intend something different to be understood by each expression.

Back to the original question – do “undetached rabbit parts” and “rabbit” mean something different? This is, I think, a matter of intuition. But in all cases, I think the result is harmless for triangulation. It seems that “undetached rabbit parts” could also apply to a thing made up of all rabbit parts, but not in the typical organization that we find in a rabbit. For instance, one ear might be attached to the stomach, and one foot coming out the back, with the tail on the neck, and so on. It seems to me that this object is certainly undetached rabbit parts. The question is whether this is also a rabbit. Some might have the intuition that this thing is not a rabbit, while others might have the intuition that while it is certainly an atypical rabbit, and likely a rabbit in which something has gone horribly, horribly wrong, it is a rabbit nonetheless.

As I said above, I find neither case problematic. In the first case, the intuition that this odd rabbit-thing is not a rabbit gives us our test case that would reveal the difference in meaning between “rabbit” and “undetached rabbit parts.” In this case, triangulation in principle would reveal the correct interpretation of “gavagai” depending on whether the speaker applies

---

“gavagai” to the weird rabbit-thing or not. In the second case, since there is no test case that would show a difference in use between “rabbit” and “undetached rabbit parts,” I conclude based on my comments above that the two expressions have the same meaning. Thus in this case interpreting “gavagai” as rabbit or as undetached rabbit parts amounts to the same thing. The words may be different, but the meaning, or expressed content, remains the same.

3.3 Lepore and Ludwig’s Objection

Finally, we come to Lepore and Ludwig’s objection. They offer us the following thought experiment:66 Lepore and Ludwig ask us to imagine a being who is equally capable of perceiving events in the environment and events that occur on the sensory surfaces of a speaker, A. This being can shift its attention effortlessly between these two types of events. Thus, if A were to engage in triangulation with this being, the being would have a choice about what to treat as the common cause – the distal event or the proximal event. There is this choice because, in this case, both the distal event and the proximal event are shared by the two triangulators.

The result of this, Lepore and Ludwig argue, is that if the contents of thought are determined by triangulation, we would then be forced to conclude that the content of A’s thoughts changes whenever the other being shifts his attention from proximal to distal, or vice versa. This is absurd – how could the content of A’s thoughts be changing without A’s knowledge? It seems much more reasonable to say, contra the result of triangulation, that A’s expressions have a constant meaning, no matter what the alien chooses to attend to at any given time. If we hold it to be true that the content of A’s thoughts is constant, then triangulation has

66 Lepore and Ludwig, Donald Davidson, 412.
determined the contents of A’s thoughts wrongly. Triangulation thus cannot be sufficient for determining thoughts.

I believe this objection can be resolved. First, though, let us be clear about what would have to happen for the objection to have weight. Suppose that A and the “special being,” (I will call it the “alien” from now on) are in the presence of a table and each respond by uttering “table.” If the alien chose to pay attention only to actual objects, we would obviously have no problem. If the alien chose to pay attention only to A’s sensory surfaces, then it would locate the common cause of the utterances as A’s undergoing a certain set of stimulations, say X, and thus triangulation would supposedly determine A’s undergoing stimulation set X as the content of the alien’s expression “table.” That is to say, when the alien says “table,” it means that A is undergoing stimulation set X.

If this is truly what the alien means, it would be easy to come up with a test case to discover this. All we would need to do is devise a situation in which the alien was in the presence of a table, but A was not undergoing stimulation set X, and see whether the alien assents to “table?” or not. If the alien means by “table” that A is undergoing stimulation set X, it will have to dissent in this test case. If we take it that A still means table by “table,” then this case is the same as the Talmage case. As in my response to Talmage’s objection, continued triangulation can in principle make the difference in meaning between the alien and A become salient.

So, as we probably understood from the thought experiment itself, in order to make the alien problematic, we have to suppose that the alien is (deliberately?) inconsistent in its use of “table.” Again, there are two types of inconsistency the alien might engage in. The first type is that the alien does not direct its attention in a way such that it will ever employ the word “table”
in a consistent fashion. This means that sometimes the alien will assent to “table?” in the
presence of a table, but sometimes it will not (in this case, it would be paying attention to the
sensory stimulations of A, who would happen to not be currently receiving table-stimulations).
Sometimes the alien will assent to “table?” when A is undergoing stimulation set X, and
sometimes not (in this case, A would be in the presence of a table, and thus undergoing
stimulation set X, while the alien would not be in the presence of a table, but choosing to pay
attention to actual objects and not A’s stimulations). The same idea holds for any other speaker
the alien might encounter.

My first claim is that it is impossible to triangulate with the alien if it engages in this first
type of inconsistency, and thus the alien’s utterance of “table” in this case is not meaningful.
Since triangulation requires, on the part of both speakers, recognition of a pattern in the
responses of the other speaker, if no such pattern is present in the alien’s utterances of “table,” no
triangulation will take place, and its utterance will not be given meaning.

This claim can be argued for in another way as well. As mentioned above, Davidson
advances the view that meaning itself consists in what a speaker intends to be understood by the
use of an expression. So the question is: can the alien reasonably intend something to be
understood by its utterances of “table”? I think the answer has to be “no.” What the alien would
intend to be understood by “table” would change case by case. In ambiguous cases where there
is both a speaker and a table present, there is no way for the speaker to know whether the alien
intends table or that the speaker is undergoing a certain set of stimulations to be understood by
its utterance of “table.” Usually this is fixed by triangulation. However, since the alien’s
utterances of “table” in this case are distinctly unpatterned, there is no way a speaker could ever
come to know what it is the alien is intending to be understood – the speaker could never form a
correlation between the alien’s utterance of “table,” and some conditions of its truth. Thus the alien’s utterances of “table” are completely useless in communication, which is another way of saying, I think, that they are simply not meaningful.

I now come to the second type of inconsistency the alien might engage in. This second type of inconsistency is that the alien shifts its attention so that it will always assent to “table?” in the presence of a table. This would involve always paying attention to tables when A (or some other potential speaker) is not undergoing stimulation set X, but having a choice of what to pay attention to when a table is present and A is being so stimulated. It seems to me that we can solve the second case by introducing a third speaker, B, to engage in triangulation with A and the alien. While many example cases of triangulation include only two participants, there is actually no limit on the number of speakers who might triangulate at a certain time. Since A and B cannot pay attention to each other’s sensory surfaces, the only candidate for a common cause in this case is the actual table.

Again, we might look at this case from the perspective of what the alien could reasonably intend to be understood by its utterance in this situation as well. The alien knows when it is switching its attention, but A and B do not. So even if the alien happens to be paying attention to A’s sensory stimulations, since there is in principle no way for A to know this, the alien cannot intend for A to understand that it means “you are undergoing stimulation set X” by its utterance of “table,” since the utterance is equally correlated with the presence of tables. One might even say it is more strongly correlated with the presence of tables, since in this case, the alien will assent to “table?” even when A is not being stimulated.
3.4 Conclusion

With the above response I have offered to Lepore and Ludwig, I conclude my work for the present paper. My intent was to argue that triangulation provides a solution to the ambiguity problem. As I mentioned previously, since there are more objections to triangulation out there, I cannot conclude that I have definitively shown that triangulation is a solution. However, I believe that I have contributed to the case. By doing so, I have also provided an argument for the success of radical interpretation. Since the ambiguity problem was a threat to this success, by arguing that it can be solved, I am also making an argument that radical interpretation can be successful.

Davidson’s ideas about things like language, thought, meaning, truth, reality, and the connections between these things, are intricate and dense. I am certain there is more to Davidson’s views on these matters than I was able to reproduce here. However, I do hope that I have been able to shed some light on Davidson’s semantics with the exposition I have offered of his views, and with my discussion of the ambiguity problem and triangulation. The major proposals Davidson makes that were covered in this paper might be summarized as follows. First, a truth conditions are sufficient for understanding, and thus can be said to give what an expression means. Second, an interpretive truth theory can also work as a meaning theory because it enables the understanding of any expression in a given language, and it does this in such a way as to exhibit the compositionality of meaning. Third, based on purely behavioral evidence, an interpreter can construct and confirm a truth theory for a language that is guaranteed to be interpretive. Fourth, triangulation, or communication with another speaker about similar responses to similar events, is necessary for a creature to have the concept of an object and to make the contents of its thoughts determinate. Triangulation is thus a precondition for creatures
to be speakers – for their sounds to have content, or to be meaningful. Fifth, meaning must therefore be in principle public – if a speaker cannot in principle come to understand an expression, the expression cannot be meaningful. And finally, sixth, meaning consists in what a speaker intends to be understood by an expression.
References


