Marc Joseph

Donald Davidson

Donald Davidson

This page intentionally left blank

Donald Davidson

Marc Joseph

McGill-Queen's University Press
Montreal & Kingston • Ithaca

ISBN: 0-7735-2780-X (hardcover)
ISBN: 0-7735-2781-8 (paperback)

Published simultaneously outside North America
by Acumen Publishing Limited

McGill-Queen's University Press acknowledges the financial support of the Government of Canada through the Book Publishing Development Program (BPIDP) for its activities.

National Library of Canada Cataloguing in Publication Data

Joseph, Marc A., 1962-
Donald Davidson / by Marc A. Joseph.

Includes bibliographical references and index.
ISBN 0-7735-2780-X (bound).—ISBN 0-7735-2781-8 (pbk.)

1. Davidson, Donald, 1917-2003. 2. Meaning (Philosophy) I. Title.

B945.D384J68 2004 121'.68'092 C2004-900790-4

Designed and typeset in Century Schoolbook
by Kate Williams, Swansea.
Printed and bound by The Cromwell Press, Trowbridge.

For Sheila, and the boys

This page intentionally left blank

Contents

This page intentionally left blank

Acknowledgements

I have been fortunate in writing this book to have had the kind assistance of a number of colleagues, friends and students. Before that, my work on Davidson benefited from the support and encouragement of two people who helped to shape my philosophical interests. The first was Professor Sue Larson, who introduced me to Davidson's writings and helped me recognize that the technical issues addressed by the analytic tradition in the philosophy of language have profound implications. The second was the late Bruce Cooper, in whose work I first saw drawn out in detail the connection between Davidson's philosophy of language and action theory, on the one hand, and deep and important issues in moral and political theory, on the other.

I am indebted to the many people who read versions of this book, in part or in whole. They include Jerry Clegg, Hillary Glick, Sheila Alter Joseph, Sue Larson, Elizabeth Potter, John Shand, Maury Silver and Kate Williams as well as the two anonymous reviewers who made many helpful suggestions. I'd also like to thank the students in my philosophy of language and philosophy of mind classes at Mills College for their useful feedback on earlier drafts. Thanks are owed, too, to Professor Akeel Bilgrami for his help in my clarifying some issues that remain unresolved in Davidson's writings. I am also grateful to my publisher, Steven Gerrard, for his patience and assistance in seeing this work through to press. Finally, I owe a debt of gratitude to Professor Davidson, who passed away just as this book was moving through its final prepublication stages. Early on he supplied me with an up-to-date bibliography, which proved to be crucially helpful, and he answered a number of questions that arose as I was completing the book.

This page intentionally left blank

Chapter 1

Introduction: Davidson's philosophical project

Donald Davidson ranks as one of the most influential philosophers of the second half of the twentieth and beginning of the twenty-first century. Davidson was trained in the analytic tradition in philosophy, which traces its origins back to Gottlob Frege and Bertrand Russell and continues through the logical empiricists and W. V. Quine, who was Davidson's teacher when he was a graduate student. A central focus of this tradition is the nature of language, and some of Davidson's most significant and widely cited work is his contribution to methodological and substantive debates about fundamental matters in the philosophy of language. Davidson argues that the most fruitful way to answer the basic question "What is it for our words to mean what they do?" is to investigate *theories of meaning* that model the knowledge an interpreter possesses when she understands a speaker's utterances.[1] His work on theories of meaning connects with problems in the metaphysics of mental concepts, and his arguments for the position he calls *anomalous monism* present one of the live options in contemporary philosophy of psychology; at the same time, Davidson's ideas about language and mind have a bearing on the nature of action, and since his earliest published work Davidson has been one of the seminal figures in contemporary action theory. From the complex ties that link these disparate writings there emerges, especially in Davidson's later work, a critique of traditional ideas about truth, scepticism and relativism, and the relation of subjectivity to objectivity. This critique is highly controversial, for Davidson counsels nothing less than

"relinquishing what remains of [the] empiricism" that characterized Anglo-American philosophy for much of the twentieth century (Davidson 1990d: 68). In this respect, Davidson's work in philosophy departs from the tradition in novel and exciting ways.

1.1 From Plato to the philosophy of language

Given this brief overview of his work and interests, it is at first somewhat surprising to learn that Davidson began his career working on Greek philosophy – as he ironically puts it, the "bold purpose" of his Harvard PhD thesis is "to try to explain the philosophic meaning and intention of Plato's *Philebus*"[2] – after having majored in classics and literature in college. However, a closer examination of his writings reveals an underlying programme and pattern (evident, no doubt, only in retrospect) and, indeed, one of the attractions of Davidson's work is the breadth and unity of his interests. Alfred North Whitehead, one of Davidson's undergraduate professors at Harvard, famously described European philosophy as a series of footnotes to Plato, and one might characterize the arc of Davidson's career as a more or less systematic working through of a number of the problems Plato left us: the nature of meaning and its connection to truth; the relation of belief to knowledge; the nature of human action; and the place of the human mind in the world order.

Davidson has explained that he began to direct his energies to the topics we associate with him only after participating in research on the theory of rational choice in the mid 1950s.[3] The theory of rational choice, or decision theory, is a modern, formal investigation of the ancient concept of deliberation, the *locus classicus* for which is Aristotle's *Nicomachean Ethics*. According to Aristotle, human behaviour rises to the level of ethical concern when we can say that an agent *deliberately chooses* to engage in that behaviour. This seems right, for it excludes acts we perform under compulsion or due to ignorance; in the former the choice to act is made for us by whoever or whatever compels us, while in the latter we lack understanding of what we seem to have chosen to do. In both scenarios, not having chosen the action our behaviour is not subject to moral evaluation, and we should not be blamed or praised for it.

In his discussion of deliberate choice (*proairesis*), Aristotle identifies as *deliberation* (*bouleusis*) the stage that precedes our making

a choice. An agent deliberates over which of several courses of action open to her is likely to eventuate in an outcome she values, and on the basis of that deliberation she chooses to pursue a course of action. In thus constructing a theory of deliberation or rational choice, we model the process of an individual's decision-making: how she chooses to realize her goals through actions in which she is able to engage. A moment's reflection reveals the very wide import of such a theory, as Davidson writes, for its goal "is to throw light on how people make decisions in the circumstances of everyday life" (Davidson 1957: 7).

Davidson has made some contributions to the theory of rational choice, but the most important consequence of his work is the way it led him to ask questions about the nature of action, belief and meaning. To see how reflection on problems in decision theory led Davidson in this direction, consider the case of a researcher, Jane, who offers an experimental subject, Jack, the opportunity either to receive $5 (option A), or to choose to gamble and receive $11 if a tossed coin comes up heads and nothing if it comes up tails (option B). The pattern of Jack's choices, given his beliefs about his chances of tossing a head in option B (e.g. he might believe that the coin is weighted one way or another), is of considerable interest to decision theorists. For example, suppose further that Jack has expressed his preference for option B over option A, and he has also expressed a preference for some third option (option C) over option B; will Jack also prefer option C to option A? That is, is the pattern of his preferences transitive?

An agent's choice behaviour is a function of two independent factors: the strength of his beliefs and the strength of his preferences. In our example, whether Jack chooses option A or option B depends on how likely he believes it is that he will win the money in option B, and it also depends on the value he places on receiving different sums of money. Davidson notes that there is a third factor that plays a role in Jack's deliberation; namely, his interpretation of the words Jane speaks in setting up the situation. If Jane is to succeed in teasing out the relative contributions of Jack's beliefs and desires then, as part of her analysis of Jack's pattern of choices, she has to assume that Jack understands her instructions in setting up the choice scenario, and this supposition is non-trivial.

Davidson's observation runs deeper. Jane is interested in the pattern of her subject's beliefs and values, but her only access to

Jack's attitudes are his words and other actions; she only knows what Jack prefers because he *says* or otherwise *communicates* that he prefers one option over another. Thus Jane, too, must be an interpreter; she cannot begin to construct or test a theory that describes the pattern of Jack's choices unless she already knows enough about his language to interpret his words. If we model this knowledge as a *theory of meaning* or *interpretation*,[4] knowledge of which would suffice for her interpreting Jack's utterances, then we can express the point by saying that the project of constructing a theory of meaning is prior to constructing a theory of rational choice. In other words, *first* she figures out what he means by his words, *then* she analyses the pattern his choices make. This priority is merely apparent, however, for the evidence on which any interpreter bases her theory of meaning for a speaker includes a description of the speaker's attitudes, especially his network of beliefs and desires, and this is given (in part) by rational choice theory. Hence we ought to see Jane as engaged *simultaneously* in two closely related interpretative projects. In light of this, Davidson sets as his goal "a theory where just by noticing what choices a person makes among *sounds* you could figure out what those sounds meant to them, and *at the same time* then figure out what they valued and what they believed" (Davidson 1994c: 210).

1.2 What is and ought to be a theory of meaning?

This goal points to a difference separating Davidson from one of the main traditions in twentieth-century philosophy of language, represented by J. L. Austin, Paul Grice, and P. F. Strawson and, more recently, John Searle, Stephen Shiffer and Brian Loar.[5] These philosophers adopt an *intention-based* approach to semantics, in the sense that they take as fundamental the idea that when a speaker utters a sentence, she intends to produce certain beliefs in her audience by means of that utterance, and what she intends determines what she means.

Grice, for example, identifies speakers' intentions as the vital component in an account of linguistic meaning, and that which distinguishes linguistic meaning from what he calls "natural meaning".[6] By natural meaning, Grice has in mind the relation we express when we say, for example, that certain spots *mean* that a person has measles or that smoke *means* fire; and natural meaning

differs from linguistic meaning in that as we cannot argue from "'Those spots meant measles' to any conclusion to the effect that *somebody or other meant* by those spots" that he had measles (Grice 1957: 39, emphasis added).[7] This is important, because the concept of linguistic meaning finds its home in an account of interpersonal communication. Thus, in contrast, we can argue from *an utterance's* meaning that someone had no wish to make a long speech to the conclusion that *someone* (namely, Pericles, in his funeral oration to the Athenians) meant that he had no wish to make a long speech; and this, in fact, is what is important about the utterance. Grice has revised his original proposal under the weight of many counterexamples, but setting aside most of these details we can roughly define Pericles' meaning (in the sense of *linguistic meaning*) that he had no wish to make a long speech as his intending to cause his audience to believe that he had no wish to make a long speech, and his intending to produce that belief by uttering the ancient Greek translation of my English sentence, "I have no wish to make a long speech."

Davidson has remarked on the influence that Grice has had upon his work, especially the view that words mean what their speakers intend them to mean (Davidson 1990d: 311). It should be evident, however, that Davidson cannot directly exploit Grice's insight to construct a theory of meaning since, according to Davidson, "an analysis of linguistic meaning that assumes prior identification of nonlinguistic purposes or intentions will be radically incomplete" (Davidson 1990d: 315–16). Recall that in our hypothetical situation Jane wants a theory of meaning to help her interpret Jack's words. If she follows the path that Grice's analysis suggests, then Jane will interpret Jack's words by appealing to his intentions and beliefs, such as his intention to cause Jane to believe that he prefers option *B* over option *A* and his belief that he can accomplish this goal by saying, "I prefer option *B* to option *A*". In effect, this strategy depends on Jane's knowing details of Jack's psychology prior to her knowing the meanings of his words: *first* she knows Jack's intentions and other attitudes, and *then* she interprets his utterance by fitting his words into an account of his beliefs and desires. Davidson urges, however, that an interpreter's only access to a speaker's attitudes are through his actions and that, more generally, "interpreting an agent's intentions, his beliefs and his words are parts of a single project, no part of which can be assumed to be complete before the rest is" (Davidson 1984:

127). In our example, Jane only knows that Jack prefers option *B* to option *A* by his *saying* that he prefers option *B* to option *A*, or perhaps by his doing something else that indicates the relative strengths of his desires (such as pointing to a card on which the words, "option *B*", are written), where this indicating, too, stands in need of interpretation.

What Davidson seeks, therefore, is akin to David Hume's "science of Man": a unified theory that encompasses the study of thought, language and action (Hume 1978: xv). Is such a "theory of everything" possible? Whatever is actual is possible, hence Davidson would argue that a unified theory is possible; after all, we do manage, in fact, to interpret the words our fellows speak, and *at the same time* we fit those words into our overall picture of their lives. We accomplish these feats, moreover, based on only those resources that Davidson identifies as being available to our hypothetical researcher and experimental subject, including a catalogue of people's utterances and other actions, and the attitudes we can observe in these actions. What Davidson is after, then, is nothing more than making explicit or rationally reconstructing what we all, in effect, already possess in some form.

1.3 Quine and Davidson

The greatest influence on the development of Davidson's philosophy is the work of Quine, his friend and teacher. Quine, in turn, is most deeply influenced by the revolution in modern logic, beginning with Frege, Russell and Kurt Gödel, and by the empiricist tradition running from Locke and Hume through the logical positivists. These two traditions intersect in the person of Rudolf Carnap, who was never Quine's formal teacher, but was his mentor and friend, and the frequent target of his criticism.

By the early 1920s, Carnap came to see the "new logics" that Frege, Wittgenstein and Russell and Whitehead had developed as supplying a key to removing vitiating defects in tradi-tional empiricism. By adopting a symbolic or formal method, these new theories provided that key by indicating the true character of logic and mathematics; in this way, they opened the door to solving "the greatest difficulty" that empiricism had faced – the problem of our knowledge of necessary truths. That true character is that "all the sentences of logic are tautological and devoid of content", and thus the difficulty

is removed in recognizing that logical and "mathematical sentences are neither empirical nor synthetic *a priori* [as Immanuel Kant had thought] but analytic" (Carnap 1930–31: 143). In other words, Carnap (with thanks especially to Frege, Russell and Wittgenstein) solved the problem that dogged Hume and J. S. Mill by arguing that our knowledge of the truths of logic and mathematics is no knowledge at all, except in the trivial sense that we know the rules of the languages we speak.[8]

Ironically, though, Carnap's project shared with Kantianism the *anti*-empirical idea that the empirical scientific enterprise is preceded by an a priori investigation of the framework of science. This affinity is easily overlooked, for while Kant identifies that framework with the structure of the human mind, Carnap instead focuses on the framework implicit in a language system.[9] For Carnap this includes logic and pure mathematics, very general statements about the structure of the physical world (e.g. the statement that space-time does, or does not, obey the laws of Euclidean geometry) and the ontology of the theory. The present point, however, is that Carnap and the other logical empiricists retain a very powerful a priori apparatus, even if it is not the a priori apparatus envisaged by Kant.

Quine turns away from Carnap's latent a priorism and returns to a model of philosophizing more associated with Locke and Hume than Kant. Like Locke and Hume, Quine takes the methods and subject matter of philosophy to be continuous with those of natural science. In a different respect, though, Quine is Carnap's faithful student, for he also takes as his project the constructing of an improved empiricism, unflawed by what Quine famously identifies as its twin dogmas: the thesis that *analytic statements* (e.g. "All bachelors are unmarried") are true in virtue of meaning or linguistic convention, independently of matters of fact; and the principle that every *synthetic statement* can be translated into a report about a discrete range of immediate experience. Both claims depend on the idea that one can separate the *meaning* of a statement from its *informational content*, which Quine subjects to an extended critique in a pair of landmark works bracketing the 1950s.[10]

To see Quine's point, consider how a person learns a theory – a theory being a set of sentences closed under logical or evidential relations – that describes the properties of some particular kind of objects, for examples, sets or molecules. We might suppose that

her learning takes place in two distinct stages. In the first stage she learns the identity of the objects the theory is about (i.e. sets or molecules), and in the second stage she learns what the theory, with its conceptual resources, says about those objects: what truths about those objects it asserts (such as that for any two sets *A* and *B* there is a set *C* to which they both belong, or that if the molecules of a non-ionic compound are linked by polar covalent bonds, then the compound is water soluble).

This model works to the extent that we have good analogies for introducing the objects of the theory. We might tell the person learning the theory, for example, that the objects spoken about by set theory are like groupings of physical objects, or that a molecule is an object the size of which compares to an amoeba as the size of an amoeba compares to a mastodon. These analogies cannot bear much weight, however, and her grasp of what sets or molecules are – alternatively, her understanding of the words that occur in set or molecular theory – awaits her learning the truths about sets or truths about molecules expressed by the statements of the theory. As Quine says, "our coming to understand what the objects are *is* for the most part just our mastery of what the theory says *about* them. We do not learn first what to talk about and then what to say about it" (Quine 1960: 16). Quine concludes that we cannot separate the meaning of a term, what one would find in a dictionary entry for that word, from information that bears on that term, or the sort of information one would find in an encyclopedia entry under that term; there is no isolatable meaning that attaches to the word "molecule" distinct from the truths of molecular theory.

The point is not merely that the meanings of someone's words are constituted by their context in the language or theory to which they belong. It is, rather, that that context is a *seamless web*. Quine famously illustrates this *holistic* conception with the image of human knowledge as

> a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. ... But ... no particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (Quine 1961: 42–3)

Because the epistemological bearing that experience has upon any sentence, which Quine identifies as the empirical content of a sentence, is mediated by the theory or language to which it belongs, the meaning of a sentence is distributed across the network of sentences that constitute the language. Quine thus draws from the image of human knowledge as "a man-made fabric" or "field of force" the implication that a sentence means what it does as a nodal point in a network of epistemologically and semantically interrelated sentences, and likewise for terms and other expressions. We have to surrender, therefore, the dogmas that the truth of some sentences ("All bachelors are unmarried") depends all on meaning and not on how things stand in the world and that the meaning of other sentences ("*That's* a molecule") can be identified with a determinate range of experiences, since the content of each statement is dispersed through the theory as a whole.

Davidson describes this observation, and the methods Quine founds on it, as "having saved philosophy of language as a serious subject by showing how it could be pursued without what there cannot be: determinate meanings" (Davidson 2001a: 145). Much more than Quine, however, Davidson uses this observation as a fulcrum with which to move contemporary philosophy of language and mind away from the empirical tradition. Davidson, as we shall see in subsequent chapters, undertakes a radical critique of the notions of meaning and mind, and arrives at a position that stands apart from that of the tradition in which he had his philosophical training.

The reader will find in the following chapters a sympathetic reading of Davidson's philosophy, not because he has the right or final answers to all the outstanding questions that define the contemporary scene in philosophy, but because, first, he offers what few other philosophers do today, namely, an over-arching theory of persons as rational animals. Secondly, this interpretation is undertaken in the spirit of the principle of charity – about which the reader will hear a great deal in what follows – in that the best way to understand a difficult thinker's work is to see it as making the best overall sense we can.

To this end, I begin Chapters 2 and 3 with an account of the assumptions and structure of Davidson's philosophy of language. This involves taking the reader through Davidson's compositionalism and extensionalism and his commitment to adopting a

Tarski-style theory of truth as the model for theories of meaning. This discussion is at times technical, and the reader with a less formal background may wish to proceed from §2.1 directly to Chapter 4, which she may do without too much loss of continuity (although with loss of formal detail). For those wishing to work through the formal details, I emphasize, at the end of Chapter 3, the philosophical importance of the concept of truth for Davidson, and thus begin to give the reader a sense of the bigger picture that embeds Davidson's philosophy of language.

Chapters 4 and 5 turn from elucidating the framework of Davidson's philosophy of language to showing how that framework is to be applied, at least in principle. There are two points I especially emphasize in these chapters. The first is the importance of Ramsey's writings on decision theory for Davidson. Davidson often comes back to Ramsey in his published writings, but that influence receives inadequate attention in other treatments of Davidson's philosophy intended for audiences of non-specialists. The other point is that Davidson's account of meaning rewrites the traditional picture, and this sets the stage for the metaphysical conclusions Davidson comes to, which we shall discover in Chapter 9.

Chapter 6 marks a transition. It introduces the concept of an event, which figures prominently in Davidson's work on mind and action theory. A great deal more could be said about the topic, but since most non-specialists will not have come to the present book with a particular interest in this issue, Chapter 6 serves mainly as a "service chapter" for other parts of the book.

In Chapter 7 I turn to Davidson's action theory, which is a topic on which Davidson's influence has been enormous. I compare Davidson's theory both to Aristotle's account of the practical syllogism and to more recent work influenced by Wittgenstein and represented by Peter Winch. In this chapter, too, I discuss a charge commonly made against Davidson's theory of action and mind: that it leaves the category of the mental as a mere "epiphenomenon" of the physical. I argue that a defence against this charge can be sustained if one bears in mind the systematic structure of Davidson's philosophy.

Chapter 8 takes up the topic of Davidson's philosophy of mind. I set that philosophy in a context that should be familiar to many readers, by comparing Davidson's account of mind with Cartesian dualism; later, I show that Davidson's theory achieves one of the

paramount results that any philosophy of mind should achieve, that of showing how we as minded beings (to use John McDowell's apt phrase) belong to the physical world and yet retain our autonomy. Finally, in Chapter 9, I set out the conclusions of Davidson's picture of mind and meaning for traditional views of subjectivity and objectivity, and, more generally, the relation between minded beings and the physical and mental world they occupy.

Chapter 2

Meaning and truth I

In his *Philosophical Investigations*, Wittgenstein compares a natural language, for example, English or German, to an ancient city. Our everyday speech, he says, is like the ancient town centre with its "maze of little streets and squares, of old and new houses, and of houses with additions from various periods", while more recently added idioms (e.g. a specialized scientific vocabulary), like newly constructed suburbs, are regular and predictable in their structure (Wittgenstein 1958: §18). Part of Wittgenstein's point is to stress the complex interrelatedness of different parts of a language, but the image also appeals to him for its implication that mapping a language's structure is no easier than mapping the geography of an ancient city. The merit of this comparison should be evident to anyone who has navigated the back streets of London or Boston.

Davidson is sympathetic to this analogy between finding one's way around a city and within a language, but unlike Wittgenstein (and like Frege) he is persuaded that a language must be amenable to systematic semantic analysis. In this chapter I begin to present Davidson's philosophy of language by examining those formal constraints he takes to be needed if one is going to find one's way within a language. These constraints have the effect of identifying the structure an adequate theory of meaning may take, and in Chapter 3 I focus on Davidson's appropriation of Alfred Tarski's work in the semantics of formal languages as supplying the leading candidate for this structure.

2.1 A first constraint on a theory of meaning

A leading idea of Davidson's philosophy of language is that – subject to a number of qualifications – a compositional theory of truth can serve the purposes of a theory of meaning. This thesis, obviously, stands in need of elucidation, and in the remainder of this section I explain the meaning and motivation of the thesis that a theory of meaning ought to be *compositional* in structure; then, in §2.2 and §2.3, I explore the notion of a *theory of truth* and begin to examine its suitability to play the role of a theory of meaning.

The idea that a theory of meaning ought to be compositional in structure expresses Davidson's commitment to language's being amenable to systematic analysis; and this commitment, in turn, is a legacy of Plato's insight that sentences are semantically complex, which Frege went on to clarify in an exact fashion 22 centuries later. For Plato, the problem of the nature of language arises in the context of the "paradox of false propositions", a problem he inherits from the Eleatic philosophical tradition.[1] Here is the paradox. Consider that when I say that the man Theaetetus sits, I express or mean (let us assume) the fact that Theaetetus sits; but what do I express when I say that he flies? Since Theaetetus does *not* fly, there is no fact that I mean; I have meant something (namely, the fact that Theaetetus flies) that does not exist. But, following Parmenides, we ought to say that what is not, in no way is. We conclude, therefore, that when I say that Theaetetus flies I have not said anything. Yet my words are *not* nonsense; in uttering them I do *seem* to be saying something.

Plato unravels this apparent problem in three steps. First, he distinguishes the *meaningfulness* of a statement from its *being true*; when I say that Theaetetus flies, I say something meaningful, but false. This separation of significance from truth is the *sine qua non* of the philosophy of language.[2] Secondly, Plato contrasts the complexity of a sentence with the simplicity of a *semantic primitive*, a meaningful expression, no proper part of which is itself meaningful. In our example the words "Theaetetus" and "flies" are semantic primitives since each is a meaningful expression – which together compose the semantically *complex* sentence "Theaetetus flies" – but no ingredient of either (e.g. the letters "t", "h", etc. that compose the name "Theaetetus") is itself meaningful.[3] Finally, Plato completes his analysis by identifying the

meaning of semantic primitives with assorted objects: for example, the noun "Theaetetus" with the man Theaetetus and the verb "flies" with the property (or Form) of flying. This neatly solves the problem of how a sentence can be both false and meaningful, since it guarantees that an expression is meaningful as long as its parts are; when I say that Theaetetus flies I have spoken meaningfully, since Theaetetus and flying exist, regardless of whether, as a matter of fact, Theaetetus flies. Plato's solution, however, engenders new questions.[4]

Frege clarifies Plato's insights about complex linguistic expressions, and he shows the way to extending those insights in his groundbreaking works on logic and the philosophy of logic. In particular, Frege lays bare the structure of language when he shows in an exact fashion how the sentences of a language result from applying, repeatedly perhaps, some fixed set of rules of composition to symbols drawn from a finite set of primitive symbols.[5] For example, the Arabic numerals are compositional in this sense: starting from a base set containing the simple numerals "0", "1", "2", "3", "4", "5", "6", "7", "8" and "9", we can construct any numeral by applying the rule that affixing a simple numeral to the left of a numeral is a numeral. Thus "639" is a numeral, because it results from applying this rule of composition twice, first stringing together the simple numerals "3" and "9" to generate "39", and then affixing "6" to this result to generate "639".

In the same way any sentence of a compositional language *L* (considered from a syntactic perspective apart from its meaning[6]) can be viewed as having been constructed from simpler expressions according to some fixed set of rules. For example, given the rule for disjunction ("or"),

> If *A* and *B* are sentences of *L*, then their disjunction (*A* or *B*) is a sentence of *L*,

we can compose sentences of unbounded complexity. Given the sentences "Theaetetus sits" and "Socrates stands" (which we suppose to have been constructed from the words "Theaetetus", "sits", "Socrates" and "stands" using other rules), we can construct the sentence "Theaetetus sits or Socrates stands" by applying our rule; and we can go on to construct "Theaetetus sits or Socrates stands, or Plato reclines" by applying the rule again. To spell out the compositional syntax of a language *L*, therefore, one needs

only to specify a finite list of words and other syntactic primitives (prefixes and suffixes, case endings, etc.), along with a set of formation rules for combining those basic expressions into complex ones, according to the basic plan laid out by Frege.

With this syntactic framework in place, we can go on to characterize the compositional semantics of a language. Having fully defined the syntax of L, we transform that description into a definition of its semantics by specifying a meaning for every primitive expression and an interpretation of each formation rule, thereby showing how the meaning of a sentence is a function of the meanings of its parts.[7] For example, we superimpose upon our earlier syntactic rule the following interpretation:

> If either A is true or B is true, then their disjunction (A or B) is true.

To understand a disjunctive sentence, therefore, is a matter of grasping the sentence's semantic structure (e.g. seeing that it *is* a disjunction) and knowing the meanings of the disjuncts. In our example, if I know the meaning of "Theaetetus sits" and the meaning of "Socrates stands", and I understand the structure of the sentence, then I know the meaning of their disjunction.

Most philosophers agree that a constraint on a systematic theory of meaning for a language L is that it treat L as having a compositional semantic structure.[8] This idea, expressed as the *principle of compositionality*, finds support in the observation (usually credited to Noam Chomsky[9]) that competent users of a language can generate and understand indefinitely many sentences that they have never previously encountered. (For example, you had never heard or read the preceding sentence, and yet you were able to understand it.) To see what is at issue, consider the case of a young child first learning to speak. By the time he is 18 months old, the child will have heard some finite number of sentences. Over the course of the next year or so, he will begin to mimic and repeat his elders' utterances, and before he is 3 years old he will be freely generating grammatical sentences that are not only novel, but also apt to the situations he encounters. How does he manage this *limitless* feat, given the *limited* stimulus to which he has been exposed?

Compositionalists see in this phenomenon a first constraint on a theory of meaning. Unless we can regard the meaning of each