# DAVIDSON

*Ernie Lepore & Kirk Ludwig*

Donald Davidson (1917 – 2003) was born in Springfield, Massachusetts, and raised, from 1924, in Staten Island, New York. He was educated both as an undergraduate and graduate at Harvard University. After a stint in the navy during the Second World War, which interrupted his graduate education, he returned to Harvard to complete a dissertation on Plato's *Philebus* in 1949. He became one the most important philosophers of second half of the 20th century.

He made seminal contributions to many topics but is best known for his work in three areas. The first is the theory of meaning, especially his work on radical interpretation and the use of truth theories to illuminate the compositional structure of natural languages. The second is the philosophy of action, in particular, his view that our reasons for what we do both cause and justify them. The third is the philosophy of mind, especially his theory of anomalous monism, which affirms the token identity of mental with physical events, but denies their type-type identity.

Davidson's work is unusually unified for someone making contributions to so many different areas. At the same time, it is hard to access because it is represented primarily in a series of compressed, even cryptic, articles, written over a period of more than 40 years, difficult even by the standards of analytic philosophy. These essays overlap and often presuppose knowledge of each other. Together they form a mosaic out of which emerges a unified and surprisingly elegant overall view of the mind and its relation to the world. It sees our nature as linguistic beings as the key to the possibility of thought, to the objectivity of the world we think about, to the impossibility of massive error about that objective world, to how the mental can be physical without being reducible to it, and to how the mind moves us to action in a world of physical causes.

In the following, we approach Davidson's mature philosophical outlook through its ontogenesis, for insight into influences on it and the context of its development help to illuminate its underpinnings, its historical context, and its

influence. No fully adequate account of Davidson's contributions can be given in a short space. We trace main lines of development and provide an overview of its place in the larger tapestry of 20th century analytic philosophy.

The two main springs of Davidson's philosophical work were his interest in the nature of human agency and in the nature of language. Though initially separate, the projects these two interests gave rise to later became intertwined, in a way characteristic of much of Davidson's work. Both interests were sparked during his early years at Stanford, where he moved in 1951 from his first job at Queens College in New York. Late in life Davidson said in an interview: "It's a characteristic of mine that anything I work on for very long I get interested in. It's a lucky characteristic to have" (*Problems of Rationality*, p. 234; henceforth 'PR'). When he arrived at Stanford, he had no philosophical project. He began a collaboration with Patrick Suppes and J.J.C. McKinsey on decision theory and measurement theory. This is one source of his interest in the philosophy of action, and it played an important role in the development of his project in the theory of meaning. The other source of his interest in the philosophy of action was his dissertation advisee Dan Bennett, who spent a year in England with Elizabeth Anscombe and Stuart Hampshire, and returned to write a dissertation on the philosophy of action. In reading and thinking about the dissertation, Davidson came to believe that it was a mistake to hold that the properties of reason explanations prohibited them from being causal explanations, in contrast to the then orthodoxy in philosophy, in a period in which Wittgenstein's influence was still pervasive. An invitation from the philosopher Mary Mothersill to present a paper at the American Philosophical Association in 1963, led to his extraordinarily influential essay "Actions, Reasons and Causes," in which Davidson argued against the dominant orthodoxy. Taking an agent's reasons for action to be given by what he wants and what he believes he can do to get what he wants, Davidson argued that ordinary action explanation is causal explanation: an agent's reasons for an action both minimally justify the action from the agent's point of view and cause it to come about. Along the way he cleared up some deep seated confusions

about the relation between causes, events, and their descriptions, which was to prove central to his later argument for anomalous monism. Davidson's argument was so successful that his view quickly became orthodoxy.

During this early period, McKinsey, an early contributor to the development of quantified modal logic, invited Davidson to co-author a paper for the Library of Living Philosophers volume on Rudolph Carnap's method of intension and extension in semantics. Carnap was a central figure in the Logical Positivist movement in the 1930s and 1940s and had a great influence on the development of analytic philosophy in the 20th century. When McKinsey died before the project got started, the task fell to Davidson. Working on his contribution, while teaching the philosophy of language, stirred an interest in the theory of meaning and in the problems of compositionality, especially with respect to sentences attributing beliefs and other so-called propositional attitudes.

There were two problems which reflections on the semantics of belief sentences rendered salient. The first was the question of how we understand complex expressions on the basis of their significant parts. A sentence such as 'Galileo believed that the earth moves' is clearly understood on the basis of understanding its significant parts. We do not have to learn such sentences one by one, and we hear and understand sentences of this form we have never encountered before without difficulty. But such sentences present special problems because the embedded sentence 'the earth moves' does not have to be true in order for the attribution itself to be true, and we may obtain a false from a true sentence by, e.g., replacing 'the earth' with a co-denoting term (e.g., 'the third planet from the sun'). Carnap's method, following Frege, involved assigning to expressions both an extension (a referent, set of things it is true of, or a truth value) and an intension (roughly a sense or meaning). In belief reports, it is the intensions of the terms in the complement clause that are, as it were, active, on Carnap's view. Davidson came to think that there were significant difficulties with Carnap's account of belief sentences (and with Fregean accounts more generally). In particular, he came to doubt that the

Frege-Carnap approach to attitude sentences was compatible with the requirement that we understand belief sentences on the basis of grasping a finite number of semantical primitives and rules for their combinations. This requirement renders salient the real difficulties involved in coming to a proper understanding of the compositional structure of natural languages (see "Theories of Meaning and Learnable Languages"). The second connected problem was the question of how to tell when a proposed account of the compositional structure of a natural language sentence was correct. The solution to both problems came together.

In November 1954, Davidson presented a paper on Carnap's method of intension and extension at the University of California at Berkeley. In the audience was the great Polish logician Alfred Tarski, who was on the faculty. Afterwards, Tarski gave Davidson a copy of his article "The Semantic Conception of Truth and the Foundations of Semantics," which led Davidson to Tarski's groundbreaking technical paper on truth, "Wahreitsbegriff." Tarski provided an axiomatic definition of a truth predicate for a formal language which was demonstrably extensionally correct, that is, which enabled one to prove for each sentence of the object language (the language for which the truth predicate was defined) a sentence in the metalanguage (the language of the theory) which says under just what conditions it is true, for each of the infinity of sentences of the object language. It was a great breakthrough in the semantics of formal systems.

Davidson's retrospective remarks show clearly how these various threads came together in his program in the theory of meaning:

> …when I understood [Wahreitsbegriff] it really turned me on. Still, I might not really have appreciated it if I hadn't done that stuff on decision theory. I had an appreciation for what it is like to have a serious theory, and I think the other people who were working in philosophy of language didn't have an appreciation for what it was like to have a serious theory. … Tarski, who knew what a serious theory was like alright … didn't have much philosophical interest…. I saw how to put these two things

> together. It came to me as if the heavens had opened and then I started to write a whole bunch of things. (PR, p. 253)

Davidson saw Tarski's work as a providing a way of bypassing many of the traditional problems of the theory of meaning. Properly deployed it would provide an account of the compositional structure of language and a standard for correctness of an account of the logical form of a complex expression, namely, integration into a theory of the language as a whole which locates the role of the words in the relevant expression in a theory of their role in any other grammatical construction in which they can appear.

Tarski provided a criterion of adequacy for a truth definition for a formal language and showed how to construct a definition of a truth predicate that met the adequacy condition. The adequacy condition is Convention T, which requires an adequate truth definition to be formally correct and to have as theorems all sentences of the form (or of a form analogous to) (T),

(T) *s* is T if and only if *p*

where 'is T' is the truth predicate being defined, '*s*' is replaced by a description of an object language sentence in terms of its significant parts, and '*p*' is replaced by a metalanguage sentence (the language of the theory) which translates *s*. (S), for example, is an instance of (T), where '°' is the symbol for concatenation (thus, A°B = the concatenation of A with B).

(S) 'La'°'neige'°'est blanc' is T iff snow is white.

This guarantees that *s* is in the extension of 'is T' if and only if it is true, because if '*p*' is a translation of *s*, then it is true iff *s* is. The definition can be stated in the form of a set of base and recursive axioms that provide 'truth conditions' for every object language sentence. The base axioms apply to primitive expressions. The recursive axioms apply to expressions built up out of others, and ultimately out of the primitive vocabulary. Tarski's aim was to define an extensionally adequate truth predicate for a formal object language, a predicate true of all and only the true sentences of the language. Davidson's interest was in meaning, rather than truth. But he saw a way of exploiting the structure of an axiomatic truth theory in Tarski's style in pursuit of a meaning

theory. For if in (T) ‘*p*’ translates *s*, replacing ‘is T if and only if’ with ‘means that’ yields a true sentence. Furthermore, the canonical proof of the T-sentence (a proof that draws intuitively only on the content of the axioms) will reveal the structure of the sentence relevant to stating meaning giving truth conditions for it, and, thus, exhibit how we can understand the sentence on the basis of its parts and their mode of composition. In this way the truth theory does duty for a compositional meaning theory.

A truth theory for a natural language must be adjusted to accommodate context sensitive expressions, and the adequacy condition must be correspondingly modified. One way of modifying the truth theory is to introduce a truth predicate relativized to features of context, such as the speaker and time of utterance. That the adequacy condition can be extended in a straightforward way can be seen by reformulating the relevant portion of Conventional T as: an adequate truth theory must have as theorems all sentences of the form (T) such that replacing ‘is T if and only if’ with ‘means that’ yields a true sentence. For a context sensitive language, we introduce context sensitive truth and meaning predicates, ‘T(*s*, *t*)’ and ‘means(*s*, *t*) that’, and replace ‘is T’ and ‘means that’ in the above with these predicates.

For natural languages, axioms for predicates (like ‘is tired’) will be relativized as appropriate to features of context. This is required minimally to accommodate tense. In Tarski’s construction, the base axioms were for predicates only, since the language(s) he explored lacked referring terms. In application to natural languages, one may introduce reference axioms in addition for referring terms. Some will be context sensitive, like ‘this’, ‘I’, ‘now’ and ‘there’. Their reference axioms will be relativized to context, e.g., ‘I’ will be said to refer to the speaker of it, ‘now’ to the time of utterance. For demonstratives like ‘this’ and ‘that’, reference axioms will be conditional in form to accommodate the possibility that a speaker fails to refer to an object in using a demonstrative: roughly, for any *x* and any time *t*, if a speaker demonstrates *x* with ‘that’ at *t*, then ‘that’, taken relative to the speaker and *t*, refers to *x*.

The concept of truth Davidson took as basic. Relative to it, the concepts of reference and satisfaction are treated as theoretical concepts. Although the concept of meaning is not expressed by any predicate of a truth theory, the truth theory aims to capture it indirectly, via meeting Convention T, and the concept of meaning is treated as theoretical as well (see centrally "Truth and Meaning" and "In Defense of Convention T").

The final piece in Davidson's program fell into place when W. V. O. Quine, the most influential American philosopher of second half of the 20th century, visited the Center for Advanced Study in the Behavioral Sciences at Stanford as a fellow in the 1958-9 academic year, while on leave from Harvard University. At the time, he was preparing the final version of the manuscript of his magnum opus, *Word and Object*, which Davidson agreed to read. Davidson and Quine had known each other since Davidson’s days as an undergraduate at Harvard. Indeed, Davidson credits Quine with having changed his attitudes toward philosophy when he took a seminar with him on Logical Positivism in which Quine presented his criticisms of Carnap’s central doctrines: “What mattered to me was not so much Quine’s conclusions (I assumed he was right) as the realization that it was possible to be serious about getting things right in philosophy” (“Intellectual Autobiography”, p. 23). But Davidson’s interests in the philosophy of language developed largely after he had left Harvard and it was during the year that Quine was at Stanford that he had his greatest influence on Davidson. “When I finally began to get the central idea,” Davidson later wrote, “I was immensely impressed; it changed my life” (“Intellectual Autobiography”, p. 41).

The methodological centrepiece of *Word and Object* is the project of radical translation. The radical translator approaches the task of understanding another speaker without any prior knowledge of the speaker’s meanings or attitudes. He restricts himself to the speaker’s dispositions to verbal behaviour in response to stimulus in constructing a translation manual for him, and thus isolates the empirical content of a theory of translation. Translation manuals alike in empirical content were judged to capture all the meaning facts which

there were. In his “Epistemology Naturalized” (in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969), Quine explains the ground for this conclusion as follows:

> The sort of meaning that is basic to translation, and to the learning of one’s own language, is necessarily empirical meaning and nothing more. ... Language is socially inculcated and controlled; the inculcation and control turn strictly on keying of sentences to shared stimulation. Internal factors may vary ad libitum without prejudice to communication as long as the keying of language to external stimuli is undisturbed. Surely one has no choice but to be an empiricist so far as one’s theory of linguistic meaning is concerned. (p. 81)

This conception of the ground of meaning facts had an enormous impact on Davidson. Davidson “thought it was terrific” and reported: “I sort of slowly put what I thought was good in Quine with what I had found in Tarski. And that’s where my general approach to the subject came from” (PR, p. 258).

The earlier work on decision theory played an important role in this synthesis. Davidson brought two morals from the study of decision theory to bear on the theory of meaning. The first was that “putting formal conditions on simple concepts and their relations to one another, a powerful structure could be defined,” and the second was that the formal theory itself “says nothing about the world” and that it is interpreted by its application to data to which it is applied ("Intellectual Autobiography," p. 32). Tarski’s work provided the essential framework for developing a formal theory. Quine’s resolutely third person approach to meaning provided Davidson with an important restriction on the evidence in relation to which the formal theory was to be interpreted.

Tarski provided the framework for a formal theory of truth for a language. If we take truth as basic and use axioms which employ metalanguage terms that interpret object language expressions in giving their truth or satisfaction conditions (e.g., for any $x$, 'rot' in German is true of $x$ iff $x$ is red), then the truth theory illuminates in the proof of T-sentences the compositional structure of

language. We connect meaning with its basis in speakers' behaviour and interactions with their environment and other speakers by treating a formal truth theory as an empirical theory whose empirical content is located in how it would be confirmed for a speaker or speech community. Illumination of its theoretical concepts is sought not in reductive analyses, but rather in showing how evidence can be marshalled in support of a theory of interpretation for a speaker. In this way we make clear holistically, in Davidson’s words, "what it is for words to mean what they do."

Here we have three reorientations of the philosophical project of illuminating meaning. First, there is the introduction of the truth theory as the vehicle for the meaning theory, which represents the attempt to get out of the resources of the theory of reference all that we want by way of a compositional meaning theory. This discounts the utility of the traditional ontology of senses, intensions, properties, relations, and propositions in the theory of meaning. Second, there is the eschewing of the traditional project of providing a reductive analysis of ‘is meaningful’ in favour of a looser and more holistic form of conceptual illumination as represented by the application of the theory as a whole to the evidence as a whole. Third, there is the restriction of the evidence in terms of which the theory is to be interpreted to what is available from the third person point of view absent any assumptions about the (detailed) psychology or meanings of the speaker.

This project in the theory of meaning became intertwined with the project in the philosophy of action in two ways. The first was through the application of the methodology of uncovering the logical form of action sentences, which resulted in Davidson’s important contributions to the logic of adverbial modification ("The Logical Form of Action Sentences"), to which we will return below, and to the logic of singular causal statements ("Causal Relations"). The second was through the application of the body of theory developed in understanding human agency to the problem of interpretation. To understand this, we must consider in more detail Davidson’s account of the project of radical interpretation and the relevant portions of his work in the philosophy of

action, and this will lead us then into the developments of Davidson's work in the philosophy of mind and in epistemology.

The project of radical interpretation is similar to Quine's project of radical translation. In each case, the evidence ultimately available consists in the speaker's responses to his environment which reveal his dispositions to verbal behaviour. But whereas the radical translator aims to produce a translation manual for the speaker's language, the radical interpreter seeks to produce a theory of interpretation that says not what expressions and sentences are the *same in meaning* but what expressions and sentences *mean*. And whereas the radical translator keys his translations to responses to stimulus patterns, the radical interpreter keys his interpretation to the speaker's responses to distal events rather than patterns of stimulation at the sensory surfaces.

Central to the radical interpreter's project is the confirmation of a Tarski-style axiomatic truth theory for the speaker's language. But, though this is central, in Davidson's account, to the enterprise of interpretation, it is not all that the interpreter aims to do. He must also use the truth theory to interpret speaker utterances, and he must fill in the picture of the speaker as a rational agent responding to his environment and others. Speaking is an activity embedded in a form of life appropriate for rational agents. As Davidson puts it at one point, "[a]ny attempt to understand verbal communication must view it in its natural setting as part of a larger enterprise" (PR, p. 151). This means that understanding what people mean by what they say must be fit into and made coherent with a larger theory of them as rational, linguistic beings.

The nexus between the project of interpreting another's language and of interpreting his attitudes lies in identifying, as an intermediate stage in interpretation, attitudes toward the truth of sentences. Such attitudes, Davidson assumes, can be identified ultimately on the basis of purely behavioural evidence. In the first phase of his work on radical interpretation, hold true attitudes toward sentences took centre stage. A hold true attitude toward a sentence *s* is a belief that *s* is true. A speaker holds true a sentence *s* on the basis of two things, first, what he believes the sentence means, and, second,

what he believes to be so. If a sentence *s* means that *p*, and a speaker believes that *p*, then (at least generally or typically, Davidson assumes) the speaker will hold true *s*. If we can identify the belief on the basis of which a speaker holds true a sentence *s*, then we can say what it means (on that occasion). If we can say what it means, we can identify the content of the belief on the basis of which he holds it. The trick is to figure out from observation of the relations between the speaker and his environment how to break into the circle. There is no way to do this without the aid of an additional principle governing the relation of a speaker's attitudes toward his environment. Since there must be (on Davidson's view) a way of doing it, whatever principles are needed are justified by their necessity for interpretation – "the alternative being that the interpreter finds the speaker unintelligible" (PR, p. 157).

Davidson invokes the Principle of Charity, following Quine. Later, Davidson distinguished two distinct principles which he had subsumed under this general heading, the Principle of Correspondence, and the Principle of Coherence. The Principle of Correspondence is the most pertinent to our current problem. The Principle of Correspondence requires "the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances" (*Subjective, Intersubjective, Objective*, p. 211). This represents a clarification and refinement of the way that Davidson originally states the Principle of Charity. In its application to the problem of breaking into the circle of belief and meaning, the Principle of Charity aimed to fix one factor, namely, belief, by holding that it is a constitutive principle of interpretation that a speaker's beliefs about his environment are largely correct. Thus, by correlating a speaker's hold true attitudes with what prompts them, i.e., with what prompts the beliefs on the basis of which they are held, one can make tentative assignments of meaning to the sentences, and, hence, of contents to the underlying beliefs. From correlations of the form

*S* holds true *s* at *t* iff *p*

we infer tentatively, where 'L' designates S's language, that

For any speaker *x*, and time *t*, *s* is true in L at *t* for *x* iff *p*

should be a target theorem for a truth theory for L which meets the appropriate analogue of Tarski's Convention T. The Principle of Correspondence plays the same role, but it responds to two different worries about this way of saying how the Principle of Charity is applied. First, there is the need to take into account how belief is affected by both history and perspective. Second, there is the question of what to count as what prompts a belief among the many identifiable causes of it. The Principle of Correspondence solves both problems by asking the interpreter to assign beliefs in the light of what he would take to be prompting his beliefs were he in the position, with the history, of the person being interpreted. This represents a movement toward what the philosopher Richard Grandy has called The Principle of Humanity ("Reference, Meaning and Belief," *Journal of Philosophy* (1973): 439-452).

The Principle of Coherence requires us to find the other, so far as possible in the light of his behaviour and interactions with others and his environment, to be a rational agent. This means that we must find his attitudes largely arranged in patterns that make for both theoretical and practical rationality, and which connect them in coherent ways with his intentions as revealed in his choice behaviour. This is the point of connection between Davidson's work in the philosophy of action and in his project of understanding language and linguistic communication through reflection on the project of the radical interpreter, for the detailed exploration of the structure of the norms which govern attribution of propositional attitudes on the basis of behaviour, the converse of which is the study of the explanation of behaviour as intentional action, is the project of the philosophy of action.

This represents the core of Davidson's philosophical program, which was conceived and developed in the 1950s through the early 1970s. It puts us in a position to review the major developments of his views out of this foundation.

We begin with a fuller sketch of Davidson's work in the philosophy of action and some of its central developments.

"Actions, Reasons and Causes" championed the view that ordinary action explanation is causal explanation of a special sort, namely, a sort which cites causes which also show minimally what was to be said for the action from the agent's perspective. An action explanation, according to Davidson, is successful when it indicates what he called a primary reason for the agent's action. A primary reason consists of a belief and pro attitude (a term that covers any propositional conative or motivational state, such as desires, wants, urges, and the like). The pro attitude is direct toward actions of some type, and the belief is to the effect that the action the agent performed was of that type. This applies both to actions done for their own sake and for the sake of further things. Often explanations of actions do not cite a full primary reason, but they nevertheless explain because they give enough information in the context for the primary reason to be inferred. If we explain someone's buying flowers by saying it is his wedding anniversary, it is easy to fill in the rest of the story about his beliefs and desires with respect to his action, given common knowledge. A primary reason shows something about what was to be said for the action from the agent's point of view, for it tells us that the agent thought it was of a type which he wanted to perform. From the primary reason we can construct a practical syllogism showing what was to be said for it:

> Action A is of type F
> Actions of type F are desirable (insofar as they are of type F)
> \-------------------------------------------------------
> Action A is desirable (insofar as it is of type F)

The first premise is supplied by the belief and the second by the desire. It is natural to take the primary reason cited in explanation of an action also to be what causes it. While there were a number of objections to this, a central one was the thought that the relation could not be causal because there was a logical connection between the reasons for an action and the action itself because the concept of an action is the concept of a piece of behaviour that can be made

intelligible by reasons. If one is blown off a cliff by the wind, it is something that happens to one but not something one does. If one jumps off, in contrast, it is something one does and one does it because of reasons one had. But the objection rests on a fundamental mistake. First, that one has a reason to do something does not entail that one will, and any given action may have been motivated by a great variety of different primary reasons. So there is in fact no entailment from the specific primary reason cited to explain any action and the occurrence of the action, or vice versa. Second, and more fundamentally, the causal relation holds between events, but logical relations hold between descriptions of events. The fact that two events were picked out by way of descriptions which were logically related would not in itself show that they did not stand in the causal relation, for any two events standing in the causal relation admit of descriptions which are logically connected. Thus, for example, 'The cause of B caused B' may be true despite the fact that from 'Something is the cause of B' it follows that 'B exists'.

This point is relevant to our thinking about the relation between the reasons we have and what we do. For if events stand in the causal relation in virtue of contingent laws that connect events of those types, the fact that there are some logical connections between actions and reasons (at least to the extent that no behaviour can be an action unless it is has its reasons) suggests that in understanding why reasons are causes of our actions we may need to shift from the vocabulary of reasons to the vocabulary of physics. This presages Davidson's later argument for anomalous monism, which we will turn to shortly.

Actions are for Davidson specifically bodily movements – where this is construed broadly to include any bodily changes (see "Agency"). This includes even mental actions, given Davidson's thesis that every mental event is a physical event, and the assumption that mental events are identical with physical changes in the body. Events are particulars, changes in objects, and, hence, can be described in various ways. It follows that reasons are reasons for an action under a description, for reasons focus on desirable features of actions. Many of our descriptions of actions are in terms of their effects. Thus, Booth's

killing Lincoln describes a bodily movement of Booth's in terms of an effect of it, Lincoln's death. In work on the ancient problem of weakness of the will ("How is Weakness of the Will Possible?"), Davidson argued for a model of practical reasoning which is like probabilistic reasoning in the sense that it involves assembling reasons for something which those reasons do not entail. That the sky is red at night may make it probable that the weather will be fair in the morning. That the barometer is dropping may make it probable that the weather will not be fair. But it cannot be that it is both probable that the weather will be fair and probable that it will not be. So these probability statements must be understood as relativized to their supporting evidence. Similarly there may be reasons in favour of and against an action and the practical judgment that an action is good must be conditioned by the reasons for it. In probabilistic reasoning, we follow the rule of total evidence: the best judgment is the one that is made on the basis of all of one's evidence. Similarly, there is in practical deliberation a parallel principle of rationality, the principle of continence: the better course of action is the one supported by all of our relevant reasons. When we come to a decision as a result of practical deliberation, even if it is based on all our relevant reasons, it is still a conditional judgment. But when we act, we thereby express an unconditional judgment that that action is best. If we are rational, the unconditional judgment is based on our all things considered conditional judgment about what it is best to do. We act incontinently, or display weakness of the will, when we choose an act and so judge unconditionally that it is best when our all things considered judgment favours another. The appearance of its impossibility, Davidson thinks, arises from two plausible principles: first, that what we do shows what we most want, so that if we want to do A more than to do B, and both are open to us, then we do A intentionally rather than B if we do either, and, second, that if someone judges doing A to be better than doing B, he wants to do A more than B. These principles apparently rule out doing something intentionally other than what one judges best all things considered. But once we see that there is a distinction to be drawn between the unconditional judgement that

something is best expressed in action and the all things considered judgment that something is best which is the result of practical deliberation, the appearance of inconsistency is removed.

In later work Davidson associated the unconditional judgment with intention, and argued against intentions being reducible to other sorts of states, either beliefs or desires, or to any special kind of action. On this view, an intention is a sort of pro attitude, but differs from desires and other pro attitudes that are the input to practical deliberation in being directly tied to commitments to act. Reasons for belief and reasons for intending are quite different, but reasons for acting and intending are (or almost always are) the same. Intentions can be formed prior to actions or simultaneous with them. When we act with an intention or intentionally, the intention brings about the action.

An important and fruitful part of Davidson's work on action theory is his analysis of the logical form of action sentences. Action sentences present a problem for semantics because of their capacity to take an endless variety of adverbial phrase, which themselves can be endlessly complex. The basic question from the standpoint of a compositional meaning theory is what the rules are that enable us to figure out from the semantical primitives what such sentences mean. A sentence like 'Brutus stabbed Caesar violently with a knife' implies 'Brutus stabbed Caesar', 'Brutus stabbed Caesar with a knife', and 'Brutus stabbed Caesar violently', and these appear to be formally valid inferences. The event analysis that Davidson introduced treats action verbs as introducing an implicit existential quantifier over events, and the adverbs as contributing predicates of it. Thus, our sample sentence above would be understood (roughly) as follows:

> There is an event *e* such that *e* is a stabbing by Brutus of Caesar and *e* was with a knife and *e* was violent.

This analysis has proved enormously fruitful in linguistics and has been elaborated with great sophistication to handle a wide range of syntactical and semantic phenomena.

Davidson's influential argument for anomalous monism ("Mental Events") was prefigured in his work in the philosophy of action as well.

Anomalous monism is the combination of two theses. The first is that each particular mental event is identical with a particular physical event. The second is that there are no strict psychophysical laws. The second implies that there are no type-type identities between mental and physical events, that is, that no mental event type is a physical event type. This is in striking contrast to the traditional view according to which if the mental is nothing over and above the physical it is because mental event types are just physical event types. Davidson's position is a form of non-reductive materialism: materialism (hence monism of substances rather than dualism) because it holds there are only material bodies, and non-reductive because it holds that mental types are not reducible to physical types. The argument has three main premises:

1. The principle of the nomological character of causality: If two events stand in the causal relation, then there are descriptions of them under which they are subsumed by a strict law.
2. The principle of causal interaction: every mental event stands in the causal relation with some physical event which is not also a mental event.
3. The anomalousness of the mental: there are no strict psychophysical laws.

A strict law is one that "there is no improving in point of precision and comprehensiveness" (*Essays on Actions and Events*, p. 223). The laws of physics, for example, aim to be strict laws, laws which form a closed comprehensive system for its domain that are as precise as is possible. If we assume that an event is physical iff it is subsumed by a strict physical law, and that there are just mental and physical events, then it follows, given 1-3, that every mental event is identical to some physical event though no type of mental event is identical to any type of physical event. For, as every mental event causally interacts with some physical event (2), every mental event is linked with a physical event by a strict law (1). But the law cannot be a psychophysical law (3). It must, given our other assumptions, therefore be a

physical law. Thus, every mental event has a description under which it is subsumed by a strict physical law and is, hence, physical.

The crucial assumption is the third, the anomalousness of the mental. Davidson's argument for this is cryptic and difficult. The central thought is that for two descriptions to be suitable for appearance in a law they must be suited for one another. Suppose (following Nelson Goodman) we define 'grue' to apply to anything which is green and observed before midnight, or which is not observed before midnight and is blue. Then consider the two generalizations from observation: all emeralds are green; all emeralds are grue. Both fit all our evidence, but we only think the first is confirmed by its instances and projectible to the future and unobserved instances. But this is *not* just because 'grue' is a predicate unsuited for appearance in a strict law. For, if we define 'emerire' to apply to anything which is an emerald if observed before midnight and otherwise a sapphire, we can see that 'all emerires are grue' is as well supported and projectible as either of 'all emeralds are green' or 'all sapphires are blue'. The question then arises when predicate pairs are suitable for appearing in laws, and, moreover, in strict laws. Many ordinary rough laws, such as that windows break if you throw bricks at them, we know to admit of exceptions, and they are often called *ceteris paribus* (other things being equal) laws. To turn these into strict laws, to make them more precise, we must often shift to the vocabulary of physics. The concepts of physics are a family of concepts governed by a set of constitutive principles, principles that tell us roughly speaking what constitutes something's falling under a basic physical kind, so that they fix the subject matter of the physical. Davidson holds that strict laws must draw their concepts from a family of concepts governed by constitutive principles. Once this assumption is in place, the argument depends only on one further claim, namely, that the constitutive principles governing the application of psychological concepts are fundamentally different from those governing the application of physical concepts. They are so because the application of psychological attitudes to an agent is governed by the requirement that he be found by and large rational in his thought and behavior.

This is directly tied to agents being capable of acting and so of having their behavior explained in a way that shows it rational to some degree from the point of view of the agent. While this does not mean that agents must be perfectly rational, it does mean that understanding them as agents requires finding in them a pattern of attitudes which make sense of them as doing things for reasons, and so the attribution of any given attitude (belief, desire, or intention) requires fitting it into a pattern of others that exhibits them has having logical relations and roles in relation to which behavior falls into an intelligible pattern. These principles can have no sway, however, over attributions of fundamental physical concepts because those concepts do not apply to objects or states in virtue of any content that they have. Thus, the physical and the mental make up two distinct families of concepts governed by different sets of constitutive principles. Given Davidson's assumption that strict laws must draw on predicates from the same family of concepts, it follows that there cannot be any strict psychophysical laws.

We return now to the project of radical interpretation and the morals to be drawn from reflection on it. If we take the radical interpreter's position to be conceptually basic in the sense of being the standpoint from which to articulate the basic structure of the concepts of a theory of a speaker, then what the interpreter has to assume about his subject matter can be taken to be constitutive of it and not merely to be expressing some aspect of the epistemic limitations of the radical interpreter's position. Davidson's fundamental assumption is that the radical interpreter's position is basic in this sense. This leads directly to a number of important theses about the mind-world relation and our epistemic position with respect to our own minds, the minds of others, and the external world. The central observation is that in order for a radical interpreter to bring to bear the evidence he has on the theory structured by the concepts of meaning, truth, and agency, he must see his subject as (a) knowing what he thinks, (b) as knowing what he means, and (c) as knowing what is going on around him in the world. Why? Suppose we can, as Davidson assumes, identify a speaker's hold true attitudes, i.e., his beliefs about which of his sentences are true, on the basis

of behavioural evidence. Suppose that we can determine which hold true attitudes are prompted by conditions in the environment and what conditions prompt them. Hold true attitudes are the product of how the speaker thinks the world is and what he thinks his sentences mean. As noted above, it seems that the only way to break into this circle is to assume that the speaker is right about his environment. Provided that we assume further that he knows what his sentences mean and what he believes, he will hold true sentences which are about conditions in the environment that prompt those hold true attitudes. If we identify correctly those prompting conditions, then we can read off what the sentence means and at the same time what belief it expresses.

But this all rests on the assumptions that the speaker is right by and large about the world, his own thoughts, and what he means. And if Davidson's fundamental assumption is correct, it follows that it is constitutive of what it is to be a speaker that one is mostly right about the external world, one's own thoughts, and what one's words mean.

The importance of this conclusion can hardly be overemphasized. If it is right, then we have a transcendental guarantee of knowledge of our own minds, the minds of others (for they must be accessible through the interpreter's procedure), and the external world. We secure this without having to explain how it is that we justify our beliefs on the basis of evidence, for knowledge in each of these domains emerges as a fundamental condition on having the capacity to speak and think at all. The traditional philosophical conundrums about how knowledge of the world and other minds is possible is resolved not so much by meeting the challenges head on but by sidestepping them, by refusing the challenge as posed in favour of a roundabout guarantee of what was wanted but could not be got through reflection on our supposed impoverished evidential base in sensory experience. Given this transcendental guarantee, we are in a position then to evaluate individual beliefs by how well they cohere with our overall picture of the world, which is guaranteed to be largely right. To the extent to which a belief is not in line with most of what we believe, especially about the most basic things made available through perceptual experience, to

that extent we have reason to think it not likely to be true. Conversely, to the extent to which a belief coheres well with most of what we belief, appropriately weighted by its connections with perceptual experience, we have reason to think it likely to be true. This does not mean that experience is an epistemic intermediary between believers and the world. It is a causal intermediary only. Its importance has to do with its being a causal intermediary between conditions in the environment and thoughts prompted by them which are, in virtue of that, about them.

This highlights another consequence of Davidson's position. Since radical interpretation reads into the contents of someone's environment directed beliefs the conditions that prompt them, and this expresses a constitutive feature of belief, belief is essentially relational, even general beliefs, because what concepts a speaker has, on this view, depends on what conditions in the world are systematically prompting his beliefs. Thus, Davidson is committed to a form of what is called externalism about thought content, the view that among the conditions that determine what the content of a thought is are conditions which are external to the thinker.

The result is a profoundly anti-Cartesian theory of mind and a profoundly anti-empiricist theory of knowledge. It is anti-Cartesian in rejecting the first person point of view as methodologically fundamental and in consequence in rejecting the view that the contents of the mind are what they are and are available to their subject independently of his embedding in the world. It is anti-empiricist through rejecting the traditional empiricist theory of knowledge and content according to which sensory experience is our ultimate evidence for the nature of the world and the source of our ideas, i.e., our concepts. The latter view, in its 20$^{th}$ century embodiment in the doctrines of the Logical Positivists, held that the meaning of a sentence was to be sought in the sensory conditions under which it could be confirmed or disconfirmed. The objectivity of thought is secured by its contents being determined by the distal objects in our environment which would form the basis for an interpretation of us. This secures us against idealism and relativism at the same time: against idealism by

making the mind depend on the world rather than vice versa and against relativism by guaranteeing a common intersubjective world that thought is about. Knowledge of the world and others is not seen as based on knowledge of oneself, but all three are seen as essential to the possibility of any of the others and are grounded in our nature as linguistic beings.

Davidson's mature philosophy, though it rejects it, developed out of the empiricist tradition. Seeing how this happened sheds light on its place in philosophy in the 20$^{th}$ century. Carnap was the largest influence on Quine, as Quine was on Davidson. Two central doctrines shaped Carnap's views. The first was that there was a sharp division between the analytic and synthetic, that is, between sentences (putatively) true in virtue of meaning (such as 'All bachelors are unmarried') and those true in virtue of matters of contingent fact (such as 'There are more bachelors in New York than Montana'). The second was that the content of contingently true sentences was to be sought in their method of verification in sensory experience. Quine got his fundamentally empiricist outlook from Carnap. But he rejected both the analytic/synthetic distinction and the assignment of empirical meaning to sentences one by one, as opposed to in whole theories. The rejection of the analytic/synthetic distinction for Quine was tantamount to the rejection of the distinction between the a priori and the a posteriori. Thus, the traditional view of philosophy as an a priori discipline and its pretension to provide a foundation for the pursuit of science goes by the way.

Philosophy becomes distinguished from the disciplinary sciences only in dealing with more general categories. Philosophical theorizing becomes subject then to how well it fits in with the rest of our empirical theorizing about the world. It is in this light that Quine's approach to the theory of meaning and language is to be understood. Quine was sceptical of the analytic-synthetic distinction because he was sceptical about how well grounded the concept of meaning it relied on was. He set out to replace it with something more scientifically respectable in *Word and Object*. Quine's fundamental starting point is the observation that "language is a social art" and it is this that

motivates the stance of the radical translator. For, it follows from the essentially social character of language, Quine argues, that evidence for its acquisition and deployment must be intersubjective, and, hence, recoverable from overt behaviour. In a conservative extension of the traditional empiricism theory of meaning, which keyed content to sensory experience, Quine then keyed sameness of meaning to sameness of response to patterns of physical stimulus of the sensory surfaces. This extrudes the traditional subjective basis of content in the empiricist tradition, sensory experience, to the sensory surfaces and renders it in principle intersubjectively available. A key change in Davidson's approach is that he takes it one step further by pushing the basis of shared meaning out to distal events and objects in the environment. In so doing, he rejects the last vestige of empiricism in Quine's philosophy – the third dogma of empiricism.

Davidson was a synoptic and original thinker and he dealt with large themes, though this is easily obscured by his concentration of the essay form and his compressed style and intensely analytical approach. Whether or not it is ultimately judged to be successful, his development of a unified response to the largest problems of the philosophical tradition, of mind, world and self, which we have tried to bring out in the above, and the way in which he brought into his work so many different strands of influence and combined them, is an enormously impressive achievement. The strands of his own influence have been multifarious, and they are still developing. It is too early to tell what Davidson's place in the history of philosophy will be, the final place of his work in the development of historical patterns of thought. But it is hard to imagine what the landscape of contemporary philosophy would have been like without him.

## BIBLIOGRAPHY

An extensive bibliography of primary and secondary material, compiled by Davidson himself, is contained in Hahn (ed.), 1999.

## Primary Works


- 1957, *Decision-Making: An Experimental Approach*, with P. Suppes, Stanford: Stanford University Press, reprinted, 1977, Chicago: University of Chicago Press, Midway Reprint Series.
- 1963, 'Actions, Reasons and Causes', *Journal of Philosophy*, 60, reprinted in Davidson, 2001a.
- 1967a, 'Causal Relations', *Journal of Philosophy*, 64, reprinted in Davidson, 2001a.
- 1967b, 'The Logical Form of Action Sentences', in Nicholas Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press, reprinted in Davidson, 2001a.
- 1967c, 'Truth and Meaning', *Synthese*, 17, reprinted in Davidson, 2001b.
- 1968, 'On Saying That', *Synthese*, 19, reprinted in Davidson, 2001b.
- 1969a, 'The Individuation of Events', in Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel*, Dordrecht: D. Reidel, reprinted in Davidson, 2001a.
- 1969b, 'True to the Facts', *Journal of Philosophy*, 66, reprinted in Davidson, 2001b.
- 1970a, 'How is Weakness of the Will Possible?', in Joel Feinberg (ed.), *Moral Concepts*, Oxford: Oxford University Press, reprinted in Davidson 2001a.
- 1970b, 'Mental Events', in Lawrence Foster and J. W. Swanson (eds.), *Experience and Theory*, London: Duckworth, reprinted in Davidson, 2001a.
- 1971, 'Agency', in Robert Binkley, Richard Bronaugh, and Ausonia Marras (eds.), *Agent, Action, and Reason*, Toronto: University of Toronto Press, reprinted in Davidson, 2001a.
- 1973, 'Radical Interpretation', *Dialectica*, 27, reprinted in Davidson, 2001b.
- 1974, 'On the Very Idea of a Conceptual Scheme', *Proceedings and Addresses of the American Philosophical Association*, 47, reprinted in Davidson, 2001b.
- 1975, 'Thought and Talk', in S. Guttenplan (ed.), *Mind and Language*, Oxford: Oxford University Press, reprinted in Davidson 2001b
- 1977, 'The Method of Truth in Metaphysics', in P.A. French, T.E.Uehling Jr., and H.K.Wettstein (eds.), *Midwest Studies in Philosophy 2: Studies in the Philosophy of Language*, Morris: University of Minnesota Press, reprinted in Davidson, 2001b.
- 1978a, 'Intending', in Yirmiahu Yovel (ed.), *Philosophy of History and Action*, Dordrecht: D. Reidel, reprinted in Davidson 2001a.
- 1978b, 'What Metaphors Mean', *Critical Inquiry*, 5, reprinted in Davidson 2001b.
- 1979a, 'Moods and Performances', in A. Margalit (ed.), *Meaning and Use*, Dordrecht: D. Reidel, reprinted in Davidson, 2001b.
- 1979b, 'Quotation', *Theory and Decision*, 11, reprinted in Davidson, 2001b.
- 1982a, 'Rational Animals', *Dialectica*, 36, reprinted in Davidson, 2001c.

- 1982b, ‘Two Paradoxes of Irrationality’, in R. Wollheim and J. Hopkins (eds.) *Philosophical Essays on Freud* , Cambridge: Cambridge University Press, 289-305, reprinted in Davidson, 2004
- 1983, ‘A Coherence Theory of Truth and Knowledge’, in D. Henrich (ed.), *Kant oder Hegel?*, Stuttgart: Klett-Cotta, reprinted in LePore, 1986, and Davidson, 2001c.
- 1984, ‘First-Person Authority’, *Dialectica*, 38, reprinted in Davidson, 2001b.
- 1985a, ‘Adverbs of Action’, in Vermazen and Hintikka (eds.), 1985, reprinted in Davidson, 2001a.
- 1985b, ‘Reply to Quine on Events’, in LePore and McLaughlin, (eds), 1985, reprinted in Davidson, 2001a.
- 1986, ‘A Nice Derangement of Epitaphs’, in LePore (ed.), 1986, reprinted in Davidson, 2005a.
- 1987, ‘Knowing One's Own Mind’, in *Proceedings and Addresses of the American Philosophical Association* 61: 441-58, reprinted in Davidson, 2001c.
- 1990a, ‘The Structure and Content of Truth’ (The Dewey Lectures 1989), *Journal of Philosophy* 87: 279-328.
- 1990b, *Plato's ‘Philebus’*, New York: Garland Publishing.
- 1991, ‘Three Varieties of Knowledge’, in A. Phillips Griffiths (ed.), *A.J.Ayer Memorial Essays: Royal Institute of Philosophy Supplement* 30, Cambridge: Cambridge University Press, reprinted in Davidson, 2001c.
- 1993, ‘Thinking Causes’ in John Heil and Alfred Mele (eds.), *Mental Causation*, Oxford: Clarendon Press, reprinted in Davidson 2005a.
- 1995, ‘Laws and Causes,’ *Dialectica* 49, reprinted in Davidson 2005a.
- 1999, ‘Intellectual Autobiography’, in Hahn (ed.), 1999.
- 2001a, *Essays on Actions and Events*, Oxford: Clarendon Press, 2nd edn,
- 2001b, *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 2nd edn.
- 2001c, *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press.
- 2004, *Problems of Rationality*, with introduction by Marcia Cavell and interview with Ernest LePore, Oxford: Clarendon Press.
- 2005a, *Truth, Language and History*, with Introduction by Marcia Cavell, Oxford: Clarendon Press.
- 2005b, *Truth and Predication*, Cambridge, Mass.: Belknap Press.


## Secondary Works


- Dasenbrock, Reed Way (ed.), 1993, *Literary Theory After Davidson*, University Park: Pennsylvania University Press.
- Hahn, Lewis Edwin (ed.), 1999, *The Philosophy of Donald Davidson, Library of Living Philosophers* XXVII, Chicago: Open Court.
- Kotatko, Petr, Peter Pagin and Gabriel Segal (eds.), 2001, *Interpreting Davidson*, Stanford: CSLI Publications.
- Lepore, Ernest (ed.), 1986, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell.
- Lepore, Ernest and Brian McLaughlin (eds.), 1985, *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell.

- Lepore, Ernest and Kirk Ludwig, 2005, *Donald Davidson: Meaning, Truth, Language and Reality*, Oxford: Clarendon Press.
- Lepore, Ernest and Kirk Ludwig, 2007, *Donald Davidson's Truth-Theoretic Semantics*, Oxford: Clarendon Press.
- Ludwig, Kirk (ed.), 2003, *Donald Davidson*, New York: Cambridge University Press.
- Malpas, J. E., 1992, *Donald Davidson and the Mirror of Meaning*, Cambridge: Cambridge University Press.
- Preyer, Gerhard, Frank Siebelt and Alexander Ulfig (eds.), 1994, *Language, Mind and Epistemology*, Dordrecht: Kluwer.
- Ramberg, Bjorn, 1989, *Donald Davidson's Philosophy of Language: An Introduction*, Oxford: Basil Blackwell.
- Stoecker, Ralf (ed.), 1993, *Reflecting Davidson*, Berlin: W. de Gruyter.
- Zeglen, Ursula M. (ed.), 1991, *Donald Davidson: Truth, meaning and knowledge*, London: Routledge.