

Critical Notice:
Donald Davidson's *Collected Essays*

Kathrin Glüer

Oxford University Press recently published *Language, Truth, and History*, the fifth and last volume of Donald Davidson's *Collected Essays*. 25 years after the first edition of *Essays on Actions and Events* there finally is easy (and affordable) access to one of the 20th Century's greatest philosophers.

Davidson's work almost entirely consists of essays of rare elegance, style and density. A true master of the form, Davidson seldom exceeded 20 pages. Thematically tightly interwoven, his oeuvre is holistic to a point where understanding only gradually dawns upon the whole. Full of references and allusions, to the classics and to modern analytic philosophy, these papers fully tax both our Bildung and analytic acumen. Yet what emerges is perhaps the most systematic and coherent vision of the human mind and its relation to the world offered by any philosopher since the great system builders of centuries past; a truly magnificent lifetime achievement.

This collection presents precisely what it says: Davidson's collected essays. Except for some corrections of typos and very minor revisions, the papers are reprinted as they were originally published. This is as it should be; what we want are precisely the original papers, not some tampered-with versions smoothing over tensions and changes. Short introductions have been added, those to the two last volumes by Davidson's widow, Marcia Cavell. The apparatus of cross-references, notes and comments telling us where Davidson changed his mind, and how later thought precisely relates to earlier work could have been more detailed. Some very helpful such notes are provided, but work enough is left for the scholar.

What the collection does not contain is any of the lecture series Davidson gave, most notably the *Dewey Lectures*, one of the few occasions where he systematically surveyed some of the central themes of his thought. But together with Davidson's very last lectures on predication, the *Dewey Lectures* have just been published as *Truth and Predication* by Belknap Press.

There is every reason to actually (re-)read at least some of these masterful essays. All of the discussions that Davidson made contributions to have been importantly reshaped by them. Yet, their interest is anything but merely historical. Davidson's contributions are not only of an impressive originality and variety, ranging from general theory of meaning and content

over formal semantics, the theories of truth, explanation and action, to metaphysics and epistemology, they are also often eminently relevant to present interests. Among many other things, Davidson vigorously defended such controversial claims as the veridical nature of belief, anti-conventionalism and holism about meaning, the causal nature of action explanation, anomalous monism, anti-reductivism about the mental, the impossibility of a solitary language, the dependence of thought on language, and a special brand of externalism about content.

Moreover, disparate as they may *prima facie* appear, from a Davidsonian perspective all of these claims are intimately connected: there is a sustained vision behind the individual contributions. According to Davidson, minded creatures are essentially rational animals; their mental states form rational systems and provide reasons-explanations for their actions. The contents of these states, including those of linguistic expressions, depend in systematic and observable ways on the creatures' behavior and the world around them. Rational animals thus are essentially *interpretable animals*; creatures whose minds (and languages) are in principle accessible to other minded creatures.¹ Davidson did think that the circle of rational animals was quite select; in fact, all animals, according to him, are 'out' due to lack of language. But one can find the general perspective congenial without going all the way, and I am not going to take him to task on behalf of any one's cat or dog.

Rather, I would like to spend the remainder of this notice reflecting on some of the ideas that are most basic to the Davidsonian outlook, ideas that, I suspect, quite wrongly appear so outdated to many as to rob his whole work of anything but historical interest. More specifically, I am going to focus on questions of *content determination*. Regarding language, these concern, if you will, 'foundational semantics' (the term is Stalnaker's). That is, the question is *not* what the semantics of any specific kind of expression, proper names, for instance, is. *Nor is* it which kind of formal semantic theory is preferable. One of Davidson's earliest contributions to the philosophy of language was the suggestion to use a Tarski-style theory of truth (T-theory) as a semantic theory for a natural language *L* (see esp. 1984, Essays 2-5). The idea is that such a theory compositionally specifies truth conditions for all the sentences of *L*. But even if you, for instance, prefer model-theoretic semantics and possible worlds truth conditions, that leaves the foundational question open: What is it that makes particular expressions have the semantics they do, that is, what is it that determines their

¹ In their recent book, Lepore and Ludwig (2005) emphasize this interconnected nature of Davidson's work and the central and unifying role of the rational as the interpretable animal. This is also the perspective taken in my own earlier, more modest attempt at a systematic introduction to Davidson's philosophy (Glüer 1993).

meaning?

This question assumes a distinctive guise in Davidson's work. Consider the often quoted beginning of *Radical Interpretation*:

Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? (1984, 125.)

Davidson's answer to the 'what'-question is, of course: a correct T-theory. Its *outputs* are what we know when we understand what another speaker has said, what his words, on that occasion, meant.² But which T-theory is the correct one? According to Davidson, this is ultimately an individual affair; it concerns the individual speaker and his language. Moreover, it is an *empirical* affair. To know what someone has said is to have a piece of empirical knowledge, knowledge justifiable by means of empirical evidence (since understanding is not like chicken sexing). On the assumption that we can know what others say, what they say is, therefore, determined by evidence available to us as interpreters. And the assumption is by no means optional; according to Davidson, language is *essentially public*:

The semantic features of language are public features. What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning (1984, 235).

For Davidson, the foundational question thus is the question of the evidence available for a semantic theory: the 'how'-question of our original quote. For the answer to be informative, it must be possible to state this evidence without already ascribing individual meanings or contents to the speaker.

What we should demand (...) is that the evidence for the theory be in principle publicly accessible (...). The requirement that the evidence be publicly accessible is not due to an atavistic yearning for behavioristic or verificationist foundations, but to the fact that what is to be explained is a social phenomenon. (...) As Ludwig Wittgenstein, not to mention Dewey, G.H Mead, Quine, and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is 'nothing but' observable behavior; but it *does imply that meaning is entirely determined by observable behavior, even readily observable behavior*. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language (2005b, 56, *emph. mine*).

So, this is Davidson's answer to the foundational question: Meaning is entirely determined by

² That does not mean that every competent speaker knows a formal semantic theory, however implicitly; all Davidson commits to is that we can, and often do, know what others say, and that the underlying ability can be modelled by a T-theory.

observable behavior. This amounts to what I would like to call *a weak behaviorism* about meaning.³ There is a convenient ambiguity about this claim; it has an epistemic as well as a metaphysical reading. The epistemic reading is most pertinent in radical interpretation, the situation where nothing but the ultimate evidence is available for interpreting an alien speaker. It is the speaker's observable behavior in its observable circumstances that allows the radical interpreter to determine, in the sense of finding out, what the speaker says. At the same time, observable behavior is what metaphysically determines, or constitutes, what his words mean: Meaning is an *evidence-constituted property*.

How does this work? According to Davidson, the radical interpreter can detect what he calls 'non-individuative attitudes' (2001, 158) such as the attitude of holding true towards uninterpreted sentences. Such attitudes are psychological in nature; holding a sentence true is, after all, having a belief. But since the sentences are uninterpreted, no individual content is assigned to the belief and, thus, no meaning-theoretical question is begged. The idea is, then, that data about when and under which circumstances speakers hold sentences true determine their truth conditions. The idea is not, of course, to simply *equate* holding true with being true; it is only that there is *a determination relation* between the two. But which?

Whether a speaker holds a sentence true is a product of two factors: what the sentence means and what the speaker believes to be the case (cf. 1984, 134). That means that as long as the interpreter is willing to ascribe wild enough beliefs, he can ascribe any meaning whatsoever to the speaker's sentences. In other words, as long as the ascription of beliefs is not restricted in a relevant way, there simply is no evidence-relation between holding true and T-theory. It is here that the famous principle of charity kicks in. In one of Davidson's earliest formulations, it says:

(PC) Assign truth conditions to alien sentences that make native speakers right when plausibly possible

(cf. 1984, 137). More precisely, the principle of charity connects T-theory and evidence in two ways. Later, Davidson called charity's two aspects the 'Principle of Coherence' and the 'Principle of Correspondence'. He explains:

The process of separating meaning and opinion invokes two key principles which must be applicable if a speaker is interpretable: the Principle of Coherence and the Principle of Correspondence. The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the

³ This terminology is in line with that suggested by Pagin in (2000). That Davidson's whole line of thought here is essentially Quinean in spirit should be obvious.

world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: One principle endows the speaker with a modicum of logical truth, the other endows him with a degree of true belief about the world. Successful interpretation necessarily invests the person interpreted with basic rationality (2001, 158).

The principle of charity thus makes observable behavior, the evidence for the holding true of uninterpreted sentences, into evidence for a T-theory by restricting the belief ascriptions that automatically come with interpreting held true sentences.⁴ The principle thus does double duty as a principle of determination: it determines both meaning and belief content on the basis of observable behavior. In this sense, it is the general principle of content determination in Davidson.⁵

These days, foundational questions are rarely discussed. Semantic behaviorism might simply be thought superseded by the externalism dominating the theory of content over the last decade or so. However, dismissing Davidsonian behaviorism this easily would be hasty. As far as I can see, it is not even clear that it actually conflicts with mainstream externalism.

Take Twin Earth. Is the Davidsonian committed to the claim that ‘water’ in the mouth of 1740 Toscar means the same, has the same satisfaction conditions as in the mouth of 1740 Oscar? It might seem that he is: if Oscar were to radically interpret Toscar, wouldn’t all the Davidsonian evidence tell him to interpret Toscar’s ‘water’ as satisfied by water? Well, would it? In recent discussions of natural kind term externalism, the role of certain kinds of semantic intentions has been emphasized; natural kind terms refer in the externalist way they do, it is claimed, because speakers intend them to apply to instances of the same substance as that of certain samples. Now, if speakers in general had such intentions they would be detectable even by 1740 Oscar. Toscar would, for instance, hold the following sentence true:

(W) If something isn’t of the same substance as *that* [demonstrating a sample] it isn’t water.

Now, even though the surrounding circumstances of such utterances do in fact determine whether the demonstrated samples and water are of the same substance, that is *not* observable

⁴ The idea here is not to get the number of correct theories down to one; like Quine, Davidson reckons with a number of different, but equally acceptable T-theories. These he conceives of as empirically equivalent. Moreover, he thinks that the resulting *indeterminacy* is equally harmless as the fact that we can, for instance, measure weight in grams or in ounces. Cf. 2004, 156.

⁵ It might be worthwhile to point out that Davidsonian meaning holism has its source in the metaphysico-epistemic double nature of the principle of charity: charity determines meaning by establishing a confirmation relation between behavior and content. As any confirmation relation, this relation is holistic: a T-theory is correct if it stands in an overall relation of ‘best fit’ to the totality of the data (cf. 1984, 136). The interpretation of any sentence thus depends on that of all other sentences. However, since the relation is one of best fit, there can be *different determination bases* determining *the same meaning* for a particular sentence. In other words: content determination by charity is a many-one relation (a kind of supervenience relation, if you like), and it therefore does not have the extremely counter-intuitive consequences that Ned Block once dubbed ‘the old hat’: it is not the case, that is, that every difference in belief engenders a difference in meaning (a point neatly worked out in

for 1740 Oscar. But whether or not there is a relevant switch in environment *is* observable. Toscar's observable behavior in its observable circumstances therefore does *not* provide decisive Davidsonian evidence for interpreting his 'water' as satisfied by water. And what's more, Toscar's holding true of (W) does provide Davidsonian evidence for *abstaining* from that very interpretation. What Oscar, on the basis of all the Davidsonian evidence, justifiedly could come up with is something like: For Toscar, 'x is water' is satisfied iff x is of the same substance as *that* (or: as the demonstrated samples). And now, satisfaction conditions being the extensional objects they are, whether Toscar's 'water' has the same or different satisfaction conditions than Oscar's seems to depend on the environment in just the way desired by the externalist. (And analogously for the concepts they express.)

Of course, these rough considerations do not even come close to clinching matters here. Nor do I mean to suggest that the Davidsonian really should take the sketched route. What I want to insist on, however, is that it is completely premature to regard Davidsonian behaviorism as superseded by externalism. Moreover, I want to suggest, even if it ultimately proves incompatible with externalism, it is by no means clear whose the last word will be.

Consider social externalism. After long and inconclusive debate concerning externalism's compatibility with privileged access,⁶ things have recently taken a new turn here. A much discussed question right now is whether externalist concepts can capture a subject's cognitive perspective. Can externalist concepts account for the subjective rationality of a subject's reasoning and actions? Social externalism leads to ascriptions of incompletely understood concepts. Bert, who thinks he has arthritis in his thigh, is Burge's classical example (cf. Burge 1979). But subjects reasoning with incompletely understood concepts will make all sorts of rather absurd mistakes; they will draw clearly, some would even say conceptually, false inferences and miss equally clearly valid ones. Externalism thus threatens to turn what from the subject's own perspective seemed to be easily explicable conclusions and actions into deep irrationalities.⁷ Even if not all of them recognize their trouble as the headlong clash with the principle of charity it in fact is, externalists are becoming more and more aware of its seriousness. If such rampant irrationality cannot be kept at bay, almost everyone involved seems to agree, that is very bad news for externalism indeed.

All this is not to say, of course, that there aren't any problems or open questions connected

Pagin 1997).

⁶ A debate Davidson himself has contributed to; in his later thought, he gives his views on content determination an explicitly externalist causal-historical twist (we shall come to that presently) and thus needs to deal with these issues himself. Cf. 2001, esp. Essays 1 and 2.

⁷ The point is neatly developed in Wikforss 2006; for an overview, see Brown 2004.

with Davidsonian behaviorism. In fact, some seriously open questions concern precisely the principle of charity. One of these is the question of the status the principle has. What kind of a truth is it (supposed to be)? And how is it to be justified? Let's approach these questions through the answers Davidson gave to them.

What kind of a truth did Davidson think charity is? There are many passages suggesting that he thought of charity as a conceptual necessity knowable *a priori*.⁸ Here are just two of them, the first actually from a paper that did not make it into the collection:

It should be emphasized that these maxims of interpretation are not mere pieces of useful or friendly advice; rather they are intended to externalize and formulate (no doubt very crudely) essential aspects of the common concepts of thought, affect, reasoning and action. What could not be arrived at by these methods is not thought, talk, or action (1985, 92).

What makes interpretation possible, then, is the fact that we can dismiss *a priori* the chance of massive error (1984, 168f, *emph. mine*).

But Davidson never tells us how to integrate such an *a priori* truth (be it ever so synthetic) into our, and certainly his own, broadly Quinean epistemology. An idea close at hand would seem to be to employ some weakened or relativized notion of the *a priori* utilizing the idea that even in a Quinean framework there are degrees of apriority. Exegetically, this is questionable; not only is there no positive evidence for Davidson's employing such a weak notion, there even is some evidence that he continues to think of the *a priori* in a very traditional way as not defeasible by empirical evidence, for instance when he says: "I am profoundly skeptical about the possibility of significant experimental tests of theories of rationality" (1985, 88).

But even if charity were *a priori*, how to justify it? A thought common in the literature seems to be the following:

The Principle of Charity is justified by the assumption that the position of the radical interpreter is the most fundamental position from which to investigate meaning and related matters, and it is needed to make sense of how the interpreter can see, on the basis of his evidence, another as a speaker (Ludwig 2003, 17).

Davidson does claim that charity is the one and only valid method for the radical interpreter. But does he really motivate its necessity by considerations of radical interpretation? Frankly, I do not see how he could. For the simple reason that the supposed argument from radical interpretation would seem to have the structure of a simple fallacy: that of taking a merely sufficient condition for a necessary one. If radical interpretation shows anything with regard

⁸ At one point, Davidson even seems to suggest that charity is an example of the *synthetic a priori*: see 1980, 221.

to charity, it would be this:

(RI) If charity determines meaning, speakers are radically interpretable.

But this would equally hold for any other principle determining meaning by evidentially relating observable behavior and semantic theory. By its very (double-) nature, any such metaphysical principle would provide the radical interpreter with a method for finding out what a speaker means. The mere idea that language is radically interpretable thus does nothing to privilege charity over any other such principle. The question is why we should think that the specific result predicted by charity actually is the *correct* one, and this needs to be motivated independently.

That, I think, is precisely what Davidson tries to do. Moreover, the motivation for charity is one of the themes where a certain shift in emphasis between the earlier and the later Davidson becomes quite salient. In a nutshell, Davidson, early or late, tries to motivate charity from *the very nature of belief*. But earlier, the emphasis is on the claim that beliefs are such that they necessarily form largely consistent and true clusters. *That's why*, according to Davidson, charity is the one and only valid method for radical interpretation:

What justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement. (...) If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything (1984, 137).

Much later, when Davidson once more rehearses his argument for charity, he stresses that the argument actually has two parts, one corresponding to each of the parts of charity itself:

The first part has to do with coherence. Thoughts with a propositional content have logical properties; they entail and are entailed by other thoughts. Our actual reasonings or fixed attitudes don't always reflect these logical relations. But since it is the logical relations of a thought that partly identify it as the thought it is, thoughts can't be totally incoherent (...). The principle of charity expresses this by saying: unless there is some coherence in a mind, there are no thoughts (...). The second part of the argument has to do with the empirical content of perceptions, and of the observation sentences that express them. We learn how to apply our earliest observation sentences from others in the conspicuous (to us) presence of mutually sensed objects, events, and features of the world. It is this that anchors language and belief to the world, and guarantees that what we mean in using these sentences is usually true. (...) The principle of charity recognizes the way in which we must learn perceptual sentences (1999, 343).

This later version of the argument, especially its second part, is clearly more externalist and historico-genetic in character than the earlier. It involves the idea of *triangulation* around which a good deal of Davidson's latest work centers. Originally introduced in *Rational Animals* (2001, Essay 7), Davidson came to use triangulation to argue for a number of rather

strong and controversial claims: that no solitary creature can have thoughts and that thought requires language (cf. esp. 2001, Essays 8, 9 and 14), that epistemological foundationalism is wrong and epistemological skepticism groundless (cf. esp. 2001, Essay 14).

However, the most basic claim here is that without triangulation, thoughts would not have objective empirical content. If we endorse a certain kind of ‘perceptual externalism’, i.e., the idea that basic perceptual beliefs are ‘about’ their typical causes, Davidson argues, we need to solve for certain ‘ambiguities of the concept of cause’. A plausible externalism needs to combine a perceptual with a social element, it needs to bring a second creature into the picture, a creature sufficiently like the first. For

the cause is doubly indeterminate: with respect to width, and with respect to distance. The first ambiguity concerns how much of the total cause of a belief is relevant to content. The brief answer is that it is the part or aspect of the total cause that typically causes relevantly similar responses. What makes the responses relevantly similar in turn is the fact that others find those responses similar (...). The second problem has to do with the ambiguity of the relevant stimulus, whether it is proximal (at the skin, say) or distal. What makes the distal stimulus the relevant determiner of content is again its social character; it is the cause that is shared (2001: 130).

So, thoughts would not have objective empirical contents if these, in the most basic perceptual cases, were not determined by shared causes. Hence the slogan “communication begins where causes converge” (2001, 151). Davidson explains: “your utterance means what mine does if belief in its truth is systematically caused by the same events and objects” (ibid.), and this is what, in the most basic cases, motivates charity’s counsel to attribute beliefs that are true when held true (when plausibly possible).

Needless to say, the triangulation arguments are far more complicated than I can develop here.⁹ What is clear is that Davidson claims that *actual* interaction with a second creature quite like oneself is *a necessary condition* on having thoughts with empirical content. Since the condition is only necessary, no naturalistic reduction is intended. What is less clear is the intended modal strength of the claim: what kind of necessity are we dealing with here? Other worries concern regresses and circles; for instance, it seems regressive to account for the relevant similarity of a creature’s responses in terms of others finding them similar (as in the quote above). But the maybe most perplexing question is: why require *actual* (as opposed to merely possible) triangulation?

One argument Davidson gives relies on the premise that to have any beliefs at all, a creature needs to have the concept of belief. And that concept can only be acquired in a social setting (cf. esp. 2001, Essays 7 and 9). I, for one, have great difficulties with both of these claims; the

second strikes me as (ill-supported) empirical speculation, the first as empirically false (given recent evidence from developmental psychology and autism).

Worse yet, any kind of historico-genetic restriction imposed on concept possession would seem to undercut the very idea that triangulation motivates charity; acquisition history is not part of the evidence on the basis of which charity determines content. For instance, charity would seem to predict thoughts for Swampman, Davidson's lightning created replica (for details, see 2001, Essay 2), or any other creature that it is *possible* to triangulate with. Something here has to give.

But that there are open questions and tensions threatening its foundations does not diminish the grandeur of the Davidsonian structure. What's more, as with any philosophy, open questions only prove that it is still standing strong - to be admired and attacked

⁹ For a critical reconstruction see Glüer 2005.

References

- Brown, J., 2004, *Anti-Individualism and Knowledge*, Cambridge, MA: MIT Press.
- Burge, T., 1979, 'Individualism and the mental', *Midwest Studies in Philosophy* 4: 73–121.
- Davidson, D., 1980, *Essays on Actions and Events*, Oxford: Clarendon Press, 2nd edn. 2001.
- Davidson, D., 1984, *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 2nd edn. 2001.
- , 1985, 'A new basis for decision theory', *Theory and Decision* 18: 87–98.
- , 1999, 'Reply to Andrew Cutrofello', in Hahn (1999).
- , 2001, *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press.
- , 2004, *Problems of Rationality*, Oxford: Clarendon.
- , 2005a, *Truth, Language and History*, Oxford: Clarendon Press.
- , 2005b, *Truth and Predication*, Cambridge, MA: Belknap Press.
- Glüer, K., 1993, *Donald Davidson zur Einführung*, Hamburg: Junius.
- Glüer, K., 2005, 'Triangulation', forthcoming in *The Oxford Handbook of Philosophy of Language*, ed. by B. Smith and E. Lepore, Oxford: Oxford University Press.
- Hahn, L. E. (ed.), 1999, *The Philosophy of Donald Davidson*, The Library of Living Philosophers Volume XXVII, Chicago and La Salle, Ill.: Open Court.
- Lepore, E. and Ludwig, K., 2005, *Donald Davidson. Meaning, Truth, Language, and Reality*, Oxford: Oxford University Press.
- Ludwig, K. (ed.), 2003, *Donald Davidson*, Cambridge: Cambridge University Press.
- Pagin, P., 1997, 'Is compositionality compatible with holism?', *Mind and Language* 12: 11–33.
- , 2000, 'Publicness and indeterminacy', in P. Kotatko and A. Orenstein (eds.), *Knowledge, Language and Logic: Questions for Quine*, Dordrecht: Kluwer.
- Wikforss, A., 2006, 'Content externalism and Fregean sense', forthcoming in *Externalism, Internalism, and the Theory of Interpretation*, ed. by T. Marvan, Cambridge: Cambridge Scholar Press.