Kathrin Glüer

The Status of Charity I: Conceptual Truth or Aposteriori Necessity?

International Journal of Philosophical Studies, 2006, Vol. 14, Issue 3 : 337-359

ISSN: 0967-2559

URL: http://dx.doi.org.10/1080/09672550600858320

The Status of Charity I.
Conceptual Truth or Aposteriori Necessity?

Kathrin Glüer

> Successful interpretation *necessarily* invests the person interpreted with basic rationality (1991, 211).[1]

1. Introduction

This paper has a companion: Peter Pagin's *The Status of Charity II*. Together, the papers suggest an answer to the question of the epistemic and modal status of Donald Davidson's principle of charity: It is an a posteriori truth of nomological necessity.

According to Davidson, content is determined by the principle of charity. Both for linguistic meaning and the content of our mental states, the principle is his answer to a foundational question. Skillfully playing on the epistemico-metaphysical ambiguity of 'determination', the principle at the same time provides a method for the radical interpreter. He uses as evidence for the semantic theory he is constructing for a radically alien language precisely what is in the determination base of charity as a principle metaphysically determining meaning: observable behavior in observable circumstances, or, more precisely, what in a first step can be determined on the basis of this: attitudes of holding uninterpreted sentences true (at certain times). This double nature of the principle is in perfect accord with the nature of meaning, according to Davidson: Meaning is essentially public, and, thus, an evidence-constituted property.

On the basis of the principle of charity, Davidson has defended not only such controversial doctrines as that belief is of its nature veridical, or that communication requires significant agreement in belief, but also versions of semantic holism and anti-reductivism about the mental. The principle is without doubt at the very heart of Davidson's philosophy of mind and language. The question of its *justification* thus is equally central to any critical assessment of this philosophy.

After sketching the role of charity in radical interpretation (section 2) and as a metaphysical principle of meaning determination (section 3), I shall outline the Davidsonian arguments for charity (section 4). I shall then approach the question of its justification indirectly, that is, by asking another: What is the *status* of the

[1] I shall quote Davidson's own papers by year only. For any other author, I shall quote by last name and year.

principle? More precisely, what is its epistemic status: if true, is it an a priori or an a posteriori truth (section 5)? And what is its modal status: what kind of necessity, if any, does this truth have (section 6)? I shall begin by treating these questions in an exegetical manner, but ultimately I am of course after the truth of these matters. There is ample evidence that Davidson tended to consider charity as a priori, more precisely, as a conceptual necessity. But I am going to argue that charity is an a posteriori truth, if any (section 5). I shall then explore the prospects for rescuing its necessity by defending charity as an a posteriori necessity of the metaphysical kind (section 6). I shall take some pains to show that that actually squares quite well with a lot Davidson says about charity, but conclude that a conclusion of such modal strength cannot be sufficiently supported. If charity has any necessity, it is of the psychologico-nomological kind.

If this is right, however, it tells us something about what *kind* of justification charity should receive (*not* the Davidsonian kind, for instance), but not, whether charity actually *can be* justified. And it is not easy to see whether it can – a completely new tack on the issue seems to be required. It is from here that the companion paper takes off: In his paper, Peter Pagin is going to explore a new line of defense for charity – as a nomological necessity.

2. Charity in radical interpretation

The principle of charity says:

(PC) Assign truth conditions to alien sentences that make native speakers right when plausibly possible (cf. 1973, 137).

Put thus, the principle is a maxim for the radical interpreter, the main figure of Davidsonian philosophy of language. His mission is understanding, more specifically, understanding of a radically foreign language. A language, that is, about which the radical interpreter does not know anything in advance, nor of the people speaking it and their culture. For this language, he is to produce a semantic theory, a theory, that is, that assigns a meaning to every sentence in the language. According to Davidson, Tarskian T-theories can be used as semantic theories for natural languages. A T-theory thus is used as an empirical theory ascribing semantic structure and truth-conditions to the sentences of an object language. The data the radical interpreter has to go on in developing his T-theory are extremely limited in kind: They consist of

nothing but the aliens' observable behavior, linguistic and non-linguistic, and its observable circumstances. On the other hand, of these limited kinds of data unlimited amounts are allowed. But even given all possible data about what the aliens do and say and under what circumstances, radical interpretation remains a formidable task.

How does the radical interpreter proceed? According to Davidson, he can determine which sentences the aliens *hold true* under what circumstances. On the basis of their mere behavior, that is, the radical interpreter can identify a certain psychological attitude of his aliens: that of *belief*. As long as the sentences held true are uninterpreted, however, it is completely open *which* beliefs it is that are expressed by means of these sentences. Davidson comments:

> The assumption that such attitudes can be detected does not beg the question of how we endow the attitudes with content, since a relation, such as holding true, between a speaker and an utterance is an extensional relation which can be known to hold without knowing what the sentence means. I call such attitudes non-individuative, for though they are psychological in nature, they do not bestow individual propositional content on the attitudes (1991, 158).

Non-individuative attitudes of holding sentences true, however, are the joint product of two factors: of what the speakers believe and of what the sentences mean (cf. 1973, 134). The problem for the radical interpreter thus is that any theory assigning meanings to the alien sentences on the basis of these data will at the same time be a theory that ascribes beliefs to them. "(...) I conclude that in interpreting utterances from scratch," Davidson says, "we must somehow deliver simultaneously a theory of belief and a theory of meaning" (1974a, 144). This is a problem because, as long as there are no restrictions on what beliefs to ascribe, *any* kind of meaning can be assigned to a sentence held true under such-and-such circumstances. As long as we do not mind eccentric beliefs, the radical interpreter's choice of T-theory would not be restricted at all by his 'data'.

And it is, of course, right here that the principle of charity kicks in. Its function in radical interpretation is twofold: For one, it is to restrict the number of acceptable T-theories by placing substantive constraints on the beliefs thereby ascribed to the alien speakers. Without some such principle, the things the radical interpreter can observe would not even *be* data, that is, evidence for a semantic theory. For the other, the principle effects a *ranking* of possible T-theories such that the best theory is the

correct one.[2] How does the principle do that? The idea is "to solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning" (1973, 137). The idea, in other words is, to assume that the belief expressed by a certain sentence is a belief *shared* by speaker and radical interpreter. Thus, the principle's counsel to make the speaker right; right, that is, by the interpreter's own lights.

More precisely, the principle of charity connects the radical interpreter's T-theory with his evidence in two ways. In later days, Davidson came to call 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 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* (1991, 158, emph. mine).

Correspondence connects a T-theory with data – sentences held true under certain circumstances – by counseling the radical interpreter to take the circumstances under which sentences are *held* true for the circumstances under which they *are* true. For those sentences whose truth-value varies with the circumstances of their utterance, this will work particularly well. Compare Davidson's own example in *Radical Interpretation* (cf. 1973, 135): If our aliens hold the sentence 'Es regnet' true more or less if and only if it is raining in their environment, the principle of correspondence tells the radical interpreter to take that as evidence for the following T-sentence:

(T) 'Es regnet' is true-in-Alien when spoken by *x* at *t* iff it is raining near *x* at *t*.

---

[2] The idea here is *not* to get the number of correct theories down to one; like Quine, Davidson is very skeptical that that is possible for a natural language. Therefore, he reckons with a number of different T-theories equally acceptable. These he conceives of as empirically equivalent. Moreover, he thinks that the resulting indeterminacy of meaning is equally harmless as, and should be understood in analogy to, the fact that we can, for instance, measure temperature in degrees Celsius or degrees Fahrenheit (cf. esp. 1977b 224f; see also 1974a, 154; 1980b, 156). In what follows, we shall not be concerned with the issue of indeterminacy. I shall therefore disregard it and speak of *the* correct interpretation, *the* correct T-theory, and, especially, *the* determination of meaning for a language. But clearly, for Davidson this can only mean determination up to indeterminacy.

Coherence tells the radical interpreter to invest his speaker with a plausible degree of internal logical consistency. This is partly due to what Quine calls the "interanimation of sentences" (Quine 1960, 9), that is, the fact that sentences often are held true because other sentences are. Coherence thus connects the radical interpreter's T-theory with data on sentences forming part of the circumstances under which other sentences are held true. These are particularly important for identifying the logical constants and, thus, the purely logical relations between the aliens' beliefs. But Davidson sometimes uses 'logical relations' in a more loose sense including relations of evidential support. Both principles together thus constitute a method for radical interpretation that is, as Davidson puts it, "nothing but epistemology seen in the mirror of meaning" (1975, 169). They counsel the radical interpreter to treat his aliens as good reasoners. Therefore, Davidson likes to characterize them as principles of a *basic rationality*.

Moreover, Davidson for a number of reasons I cannot go into here takes the interdependence of belief and meaning to also include preference or desire (cf. esp. 1980b). Strictly speaking, the radical interpreter therefore has to solve for three unknowns at once. To do that, the principles restricting acceptable T-theories need to link the radical interpreter's evidence not only with his beliefs and meanings, but also with his desires, intentions, and actions. Here, too, the counsel is to make the alien come out a basically rational person, an agent, that is, whose actions are understandable in the light of his beliefs and desires. This is important for our purposes only insofar as it explains why Davidson in many of the passages relevant later on lumps principles of rationality, even axioms of decision theory, and charity together. It also explains why what he says about decision theory is directly relevant for charity. Other than that, I shall mostly ignore elements of charity beyond the principles of coherence and correspondence.

At this point, the usual disclaimers are in order: Charity does not tell the radical interpreter to make his aliens right no matter what. Charity does not exclude the possibility of mistaken belief, and it does not recommend making the aliens right as often as possible, either. Davidson very soon replaced the unfortunate early talk about 'maximizing agreement' with that of 'optimizing agreement': the counsel is to make the alien right "when plausibly possible" (1973, 137). Charity thus actually *recommends* ascribing mistakes in situations where it would be hard to see how our alien could possibly have a true belief. For instance, in a situation where he cannot

possibly perceive the rain near him. Or where he believes the situation to be one in which appearances of rain typically are illusory.

Charity tells the interpreter to pick that T-theory that stands in the relation of "best fit" (1973, 136) to the totality of his data. Thus, charity fulfills its second function, that of determining the *correct* theory. With charity in place, T-theories can be ranked according to how well they fit the data. And the theory with the best fit is the right one.

To sum this up: In radical interpretation, the principle of charity establishes the crucial evidential link between what is readily observable by the radical interpreter and the T-theory it is his task to develop. The principle of charity moreover effects a ranking of possible T-theories in terms of how well they fit the evidence and thus determines the correct theory or theories as those with the best fit. Thus, it allows the radical interpreter to determine what the sentences of Alien mean.

What I am interested in here, is the question of the *status* this principle has. With respect to radical interpretation, we can approach this question by asking: Why just charity? Why not any other principle evidentially linking semantic theory with observation and effecting a ranking of theories? What makes charity valid for radical interpretation? Let's start investigating these questions in the next section.

3. Charity as a metaphysical claim and a presumed argument for it

So far, we have looked at the principle of charity as a maxim for the radical interpreter, a method he is told to employ in order to 'untie' belief and meaning. Clearly, this method is valid only if the radical interpreter arrives at the correct result by employing it. It is valid only, that is, if the T-theory that the principle of charity determines as the correct one ascribes to the alien sentences the meanings, and to the alien beliefs the contents and interrelations, they actually have. Charity is valid only if the aliens actually *are* basically rational, and their beliefs largely consistent and in agreement with those of the radical interpreter. In other words, the radical interpreter can (correctly) determine meaning only if there actually obtains *a prior relation of determination* between observable behavior and meaning in just the way charity predicts. Claiming that the radical interpreter can determine meaning, in the epistemological sense of 'finding out about', thus amounts to making a metaphysical claim about meaning determination for language in general. I shall call this metaphysical claim 'Charity', and for simplicity's sake I shall formulate it thus:

(Charity) Meaning is determined by the principle of charity.

The term 'principle of charity' thus becomes ambiguous between the maxim for the radical interpreter and the prior determination relation the maxim's validity depends on. Asking why charity is valid in radical interpretation thus naturally turns into investigating the status of Charity. Is it true? And if yes, what kind of a truth is it? Let's start by finding out how Davidson himself motivates Charity.

One idea found in the literature is that radical interpretation shows that the principle of charity is the principle of meaning determination. Here is what one commentator (out of many) says about it:

> 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, emph. mine. See also Lepore and Ludwig 2005, 204ff).

There certainly are passages in Davidson that seem to support this reading. Recall, for instance, the following passage from Davidson already partly quoted above:

> Successful interpretation necessarily invests the person interpreted with basic rationality. It follows from the nature of correct interpretation that an interpersonal standard of consistency and correspondence to the facts applies to both the speaker and the speaker's interpreter, to their utterances and to their beliefs (1991, 211, emph. mine).

Or this one:

> [T]he *only, and therefore unimpeachable, method available to the interpreter* automatically puts the speaker's beliefs in accord with the standards of logic of the interpreter, and hence credits the speaker with the plain truths of logic. Needless to say, there are degrees of logical and other consistency, and perfect consistency is not to be expected. What needs emphasis is only the *methodological necessity* for finding consistency enough (1983, 150, emph. mine).

Davidson clearly does claim that charity is the one and only valid method for the radical interpreter. In the same breath, he seems to claim not only truth, but *necessity* for the claim that charity determines meaning. But does he really motivate this claim 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 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 to charity, it would be this:

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

That, however, would merely be a sufficient condition for radical interpretability. Moreover, this would equally hold for any *other* principle evidentially relating observable behavior and semantic theory and resulting in any kind of ranking of theories such that the best are good enough. Radical interpretation thus does nothing to privilege charity over any other such principle.

No, that charity provides the radical interpreter with a method does not show that charity actually is valid. That needs to be motivated independently. But if it is necessarily true that meaning is determined by charity, it simply follows that nothing can be correct interpretation that does not arrive at the result predicted by charity. *That* is why Davidson says things like those quoted above that might have led his interpreters into misconstruing the significance of radical interpretation. What radical interpretation in fact does, is merely the following: It lends some plausibility to the claim that conceiving of meaning and belief as determined by charity allows us to satisfy a criterion that, according to Davidson, any theory of these concepts must meet: "[I]t must show how it is possible for one person (…) to come to understand another" (1985a, 88). We shall come back to the significance of radical interpretation for Davidsonian thinking later. For now, we shall turn to the arguments Davidson really gives for Charity.

4. Davidson's argument for charity

Here is how Davidson argues for Charity in *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 (1973, 137).

Charity derives, that is, from the very *nature of belief*. Davidson claims that *beliefs are such that they necessarily form largely consistent and true clusters*. *That* is why we have no reason to count anyone as having beliefs unless we can construct his beliefs as doing just that: form largely consistent and true clusters. Here is why:

> Beliefs are identified and described only within a dense pattern of beliefs. I can believe a cloud is passing before the sun, but only because I believe there is a sun, that clouds are made of water vapour, that water can exist in liquid and gaseous form; and so on, without end. No particular list of further beliefs is required to give substance to my belief that a cloud is passing before the sun; but

> *some appropriate set of related beliefs must be there*. If I suppose that you believe a cloud is passing before the sun, I suppose you have the right sort of pattern of beliefs to support that one belief, and these beliefs I assume you to have must, to do their supporting work, be enough like my beliefs to justify the description of your belief as a belief that a cloud is passing before the sun (1977a, 200, emph. mine).

Hence, charity is the only valid method for radical interpretation:

> If I am right in attributing the belief to you, then you must have a pattern of beliefs much like mine. *No wonder, then, I can interpret your words correctly only by interpreting so as to put us largely in agreement* (ibid., emph. mine. See also 1975, 168).

Much later, in his Schilpp-volume, Davidson once more rehearses his argument for Charity. And while earlier most weight seemed to be on the point that beliefs necessarily come in 'logical clusters', Davidson now 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 is not the place to trace all the subtle differences between the early and the late versions of Davidson's argument for Charity. The later version is undoubtedly more externalist, and more historico-genetic than the earlier, a shift that I personally find less congenial. But these details are less important here than getting clear about exactly *what kind* of an argument this is (supposed to be).

To put it in a nutshell, Davidson argues for charity as the principle of meaning determination by arguing that *charity is the principle determining belief content*. This, again, is first and foremost a metaphysical claim. Moreover, the claim is not just that charity in fact determines belief content, but that this is *essential* for belief. Here is a passage making that very clear and generalizing it to propositional attitudes, reasoning and action:

> 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 (1985a, 92).

Well, *if* belief content is essentially such that it is determined by the principle of charity, then it is small wonder, indeed, that charity also determines correct interpretations. But have we been given good reasons to believe this? And *what kind* of reasons have we been given? I shall start with the latter question.

5. The epistemic status of charity

Asking for the kinds of reasons Davidson gives us for believing in Charity is asking for its *epistemic* status (according to him). Over the years, Davidson has made a pretty confusing variety of pronouncements on this matter; here are a number of them, but no doubt there are more around:

> What makes interpretation possible, then, is the fact that we can dismiss *a priori* the chance of massive error (1975, 168f).

> Crediting people with a large degree of consistency cannot be counted mere charity: It is *unavoidable* if we are to be in a position to accuse them meaningfully of error and some degree of irrationality. Global confusion, like universal mistake, is *unthinkable* (…) (1970, 221f).

> Just as the satisfaction of the conditions for measuring length or mass may be viewed as *constitutive* of the range of application of the sciences that employ these measures, so the satisfaction of conditions of consistency and rational coherence may be viewed as *constitutive* of the range of applications of such concepts as those of belief, desire, intention and action (1974b, 237).

> I suggest that the existence of lawlike statements in physical science depends upon the existence of *constitutive (or synthetic a priori)* laws like those of the measurement of length within the same conceptual domain. Just as we cannot intelligibly assign length to any object unless a comprehensive theory holds of objects of that sort, we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires intentions, and decisions (1970, 221).

> I have been engaged in *a conceptual enterprise* aimed at revealing the dependencies among our basic propositional attitudes (2005, 73f).

Putting all this together, it is hard to avoid thinking that Charity is supposed to be *a conceptual, though synthetic truth, belief in which is justified a priori.* This should at least be a little surprising – coming from a guy that quite as happily calls himself "Quine's faithful student" (1983, 144) on the analytic/synthetic distinction. To be

sure, the one thing Davidson actually never calls Charity is analytic, but what would Quine have had to say about the synthetic a priori? Seriously, at the time Quine rode his famous attacks on the analytic/synthetic distinction, the only available conceptions of the analytic, the necessary, and the a priori made these at least coextensive. To be Quine's faithful student on the analytic/synthetic distinction while countenancing the synthetic a priori thus would seem to require you to 'pay extra' for your use of those terms. That is, you really would need to explain in quite some detail what exactly it is you are claiming. Accordingly, Davidson used to be much more careful in his statements; in the Seventies he would not go any farther than this in commenting on the status of the principles of decision theory:

> It may seem that I want to insist that decision theory (…) is necessarily true, or perhaps analytic, or that it states part of what we mean by saying someone prefers one alternative to another. But in fact I want to say none of these things, if only because I understand none of them. My point is sceptical, and relative (1976, 272f).

But later, he does not seem to feel these scruples anymore.[3] However, he never really pays up for this; he never really tells us much about how to integrate the a priori, not to speak of the synthetic a priori, into our broadly Quinean outlook on epistemology.

One idea here would be employing some kind of relativized or weak notion of the a priori, thus utilizing the fact that Quinean epistemology does not level all epistemic differences between our beliefs and allows for degrees of apriority. However fashionable that idea might be today, Davidson does *not* seem to avail himself of such a weakened notion of the a priori. My impression is rather that, at least early on, he tries to deal with the situation, not by weakening the a priori itself, but by instead *weakening the degree to which we justifiedly can believe something to be a priori.*

In the early paper just quoted from, Davidson not only professes not to understand the notions of analyticity and necessity but also uses this lack of understanding as a motivation for qualifying his point as *merely skeptical*. What exactly is he skeptical about?

---

[3] Evnine nevertheless thinks that Davidson should be construed as committed to the analyticity of Charity (cf. Evnine 1991, 111ff). Well aware of the passages from *Hempel on Explaining Action* (1976) just quoted, he takes Davidson to later go back not only on the necessity and conceptuality of Charity, but thereby also on its analyticity. Evnine in effect suggests to see Charity as an *implicit definition* of theoretical terms in the theory of rationality, a suggestion the substance of which is due to David Lewis (cf. Lewis 1974). That leaves open the possibility of the terms being implicitly defined having empty extensions (compare 'ether'), a possibility precluded by Davidson's understanding of the apriority of Charity (see below).

> I am sceptical that *we have a clear idea what would, or should, show that decision theory is false* (…). In this respect, decision theory is like the theory of measurement for length or mass (…). The theory in each case is so powerful and simple, and so constitutive of concepts assumed by further satisfactory theory (physical or linguistic) that we must strain to fit our findings, or our interpretations, to preserve the theory. If length is not transitive, what does it mean to use a number to measure a length at all? We could find or invent an answer, but unless or until we do we must strive to interpret 'longer than' so that is comes out transitive. Similarly for 'preferred to' (1976, 272f).

On a traditional reading of apriority, something is an priori truth if the justification for believing it cannot be defeated by empirical evidence. And what Davidson expresses skepticism about, is precisely this: that Charity can be defeated by empirical evidence. Thus, Davidson does *not* give up on requiring empirical non-defeasibility for apriority; rather, he says that there is very strong reason to doubt that Charity is *not* a priori in precisely this sense. Later on, he often formulates his claims less cautiously, as for instance when he replies to Smart:

> Smart asks 'whether people might not actually be approximately rational and consistent in their patterns of belief and desire'. In my view this *cannot be a factual* question: if a creature has propositional attitudes then that creature is approximately rational (1985b, 245).

But not always does he thus throw caution to the wind, not even in connection with bringing out the supposedly conceptual nature of Charity's truth:

> I am *profoundly skeptical about the possibility of significant experimental tests of theories of rationality*. This does not mean that such theories, or the considerations that lie behind them, have no empirical application. On the contrary, I think of such theories as attempts to illuminate an essential aspect of the concepts of belief, desire, intention and meaning. One criterion a theory of these concepts must meet is this: it must show how it is possible for one person ('the experimenter') to come to understand another ('the subject') (1985a, 88, emph. mine).

What is the motivation for this skepticism? The example Davidson generalizes from is that of the *transitivity of preference*, a part of charity that comes in through the decision theoretic connection hinted at above. When actually involved with experimental work on decision theory, Davidson had to realize how difficult it is to come up with experiments that could provide clear counterevidence to the transitivity of preference. According to him, the difficulty is to come up with experiments such that it would *not* be "at least as plausible to take them as testing how good one or another criterion of preference is, on the assumption that decision theory is true" (1976, 272). Let's grant that. My question is whether it really can be generalized as smoothly to other parts of charity as Davidson assumes.

What would a counterexample to Charity, that is, to the claim that meaning is determined by charity, have to be like? I don't think that we can get anywhere here by considering purely speculative matters like people with radically outlandish, incomensurable concepts. A much better idea is to consider an actual example again. The example I have in mind concerns a group of persons where an established practice of communication prevents us from doubting that they have language and beliefs but where it nevertheless is not immediately clear that Charity holds. And even though I ultimately do *not* think that the example in fact is a counterexample, considering it will prove instructive.

What I have in mind are certain persons suffering from autism. According to a popular and well-confirmed psychological theory, persons with autism suffer from (a degree of) 'mindblindness'. That is, what has come to be called their 'theory of mind'-capacity is (more or less) impaired, the capacity, that is, to ascribe mental states, especially beliefs, to others and thereby to predict and explain their behavior. Such a deficit results in a variety of empathy-related difficulties. When it comes to language, those persons with autism that have language (and there are persons with autism and university degrees) typically have significant difficulties with pragmatics, but much less so with semantics. Persons with autism tend to be literal interpreters.[4]

Now, as I said, it should be clear that there is not a shimmer of a reasonable doubt that a significant group of those with autism have both language and beliefs. And that is precisely what makes them such perfect candidates for being counterexamples to Charity. For here, it would simply not seem an option to take the Davidsonian way out. There is no question here that the 'experimental setup' uses the wrong criterion for detecting language and belief. There is as good evidence for that as there is for any other speaker: a longstanding and, pragmatic oddities aside, smooth practice of linguistic communication. To try and fault this criterion would not only be pure dogmatism, it would by the same token cast equally strong doubt on there being any speakers at all.

Ultimately, there is no reason to think that speakers with autism actually are counterexamples to Charity.[5] This is not immediately obvious, however. There are

[4] Cf. Glüer and Pagin 2003 where we argue that there is good empirical reason for thinking that there is a group of speakers with autism who lack the concept of belief, and, thus, the capacity to think second order thoughts. This, we argue, *does* make them counterexamples to higher order thought theories of meaning like the Gricean, but *not* to the claim that meaning is determined by charity.
[5] Cf. Glüer and Pagin 2003.

speakers with autism that lack the concept of belief and, thus, the capacity to think higher order thoughts (thoughts about beliefs). And that *does* make it impossible for them to intentionally employ charity as a *method* of interpretation: They cannot be radical interpreters. But that does not mean that what they mean by what they say, and what they understand others as saying, is not *in accordance with charity*. Charity (in both of its readings) makes no claims about the *mechanisms* of actual interpretation; rather, the principle serves to determine the correct semantic theory, no matter what the actual mechanisms of interpretation are. Charity thus is true as long as all actual (correct) interpretation, no matter how, arrives at the result the principle predicts as the correct one. To see that there is no *special* reason to doubt that it holds for speakers with autism, we only need to reflect on the fact that we do interpret these speakers just as we interpret any other speaker. (At least as far as semantics is concerned, but that *is* what we are concerned with.) No wild or strangely unmotivated beliefs need to be ascribed in taking their words to mean what ours do.[6] Thus, if Charity holds for us, it holds for the speaker with autism. Moreover, this presumably is true for *any* other case where a prior practice of unproblematic communication by language ensures that the example would have to be accepted as a counterexample.

Nevertheless, this does *not* mean that there cannot be counterexamples. What it means is only that a counterexample of the kind the possibility of which I have been considering would not merely show that Charity does not hold for all creatures with language. Rather, it would show that it does not hold for us, either. And that still seems perfectly conceivable to me, even though the possibility of course might be merely epistemic. Suppose I perform a methodologically correct radical interpretation on you. And suppose that the result conflicts with my usual scheme of interpretation

[6] There is some evidence that children with autism have a stronger tendency to neologism than other children. But since they also tend to be literal users, it is rather easy to determine what they mean by observing their application of such terms in observable circumstances. Since they are literal users, that is, the application of charity in such a radical interpretation-like situation is, if anything, actually more straightforward than it is for less literal users.

This generalizes: Radical interpretation of a speaker (or group of speakers) with autism by means of charity is, if anything, easier and more straightforward than for other speakers. According to Davidson, meaning is determined on the basis of observable behavior (in observable circumstances) in two steps: First, attitudes of holding true are determined on the basis of behavior, and then, meanings are determined by charity on the basis of attitudes of holding true. It is the first step that is, other things being equal, easier with regard to a speaker with autism (or any other literal speaker): attitudes of holding sentences true are more easily detectable here. Provided there is no additional (e.g. cognitive or perceptual) impairment, there is no reason to think that the determination base (of holding true attitudes) on which charity then operates is, in the case of the speaker(s) with autism, in any significant way different from that it operates on for the rest of us.

for you. Wouldn't I, then, have reason, empirical reason, to think that the fault is charity's? Wouldn't it be dogmatic to insist on charity in this situation? Especially as insisting on charity would amount to embracing the claim that there are no such things as meanings and beliefs as we are used to conceive of them?[7]

To sum these considerations up: It does not seem plausible to generalize from consideration of the transitivity of preference to the empirical non-defeasibility of Charity. We can readily say what counterexamples to Charity have to be like in order to block the escape route Davidson claims to be always open. The empirical evidence to the effect that we do communicate by language with large numbers of people, everyday and everywhere, is too good to be overrun by any fancy claims in the *theory* of meaning or belief. We do think, or at least I do, that the beliefs of all these people in fact roughly accord with Charity. But I don't think Davidson has given us good reason to deny that this, in the end, is an *empirical hypothesis*, that is, a hypothesis defeasible by empirical evidence. For on one conception of the a priori, and precisely the one that Davidson himself invokes, this means that Charity is not an a priori truth. I therefore recommend concluding that Charity's epistemic status is *a posteriori*. I think that that ipso facto excludes conceiving of it as conceptual. If we nevertheless want to follow Davidson at least in thinking of Charity as a *necessary* truth, we will have to settle for *metaphysical or nomological necessity*. In the remainder of this paper, I shall try out a way of arguing for the stronger of these suggestions, a way surprisingly Davidsonian both in letter and spirit.

6. The modal status of Charity

The idea now is to hang on to the claim that Charity is necessarily true, even though it is neither a conceptual nor an a priori truth. In other words: Can Charity be defended as an a posteriori necessity? What would it mean for Charity to be metaphysically necessary, but a posteriori? It would mean, of course, that charity is the meaning determining principle in all (accessible) possible worlds. But that it takes empirical investigation to determine *which* principle it is that is meaning determining (both in

[7] In conversation with Peter Pagin, Davidson once said that he thought there was 'room for adjustment on both sides' here. This suggests that he thought that in a conflict like the one envisaged above, we can either disregard (some of) the data, or adjust the principle of charity so as to fit the data (or both). As with any empirical theory, where to make the adjustment would then depend on a variety of factors. The main point, however, is precisely that 'room for adjustment on both sides' is what we think the predictions of well-entrenched *empirical* theories leave. By including room for adjusting even the principle of charity itself, its epistemic status thus would clearly not be a priori.

the actual and in all other possible worlds). You could also put this by saying that it takes empirical investigation to find out *what* the meaning determining principle exactly says.

Such a claim we presumably have to pay for in real metaphysical coin, that is, in *essences*. And, surprisingly enough, we already saw that Davidson himself seems quite willing to do that. What we therefore will try to do is divorce Davidsonian essences from their supposedly conceptual source; we are talking *de re* necessity now. Before actually presenting a Davidsonian argument for the a posteriori necessity of Charity, I would like to draw your attention to how many of the things Davidson says about Charity we can make excellent sense of on this interpretation. Suppose that Charity is a metaphysical necessity. In that case, the possibility of counterexamples is merely epistemic. We can illustrate that claim by saying that there are no (accessible) possible worlds in which meaning and belief are not determined by Charity. We can also put it in more Davidsonian jargon by saying: "What could not be arrived at by these methods is not thought, talk, or action" (1985a, 92). And analogously for the Davidsonian talk of 'inevitability', 'methodological necessity' etc. Even the claim that the truth of Charity "cannot be a factual question" (1985b, 245), which I argued is false on an aprioristic reading, can be rescued along these, roughly Carnapian, lines: There is no real, no metaphysical possibility excluded by Charity. Last, but not least, there is the Davidsonian claim that Charity is *constitutive* of belief and meaning. This might seem like the hardest nut to crack, if only because it is so hard to understand what it means in the first place. Here is, again, what Davidson says:

> The satisfaction of conditions of consistency and rational coherence may be viewed as *constitutive* of the range of applications of such concepts as those of belief, desire, intention and action (1974b, 237).

It is this very claim that he in other places puts in terms of *essences*, for instance here:

> 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 (1985a, 92).

To reconcile constitutivity with the metaphysical line proposed here, we need to free constitution or essences from their (exclusively) conceptual fetterings. For Davidson, belief is essentially charity determined, and so it is on the interpretation suggested. For Davidson, that seems to amount to the same as saying that charity is constitutive

of belief; so, in that sense, charity is constitutive of belief on our interpretation, too. As Davidson himself puts it: "unless there is some coherence in a mind, there are no thoughts" (1999, 343). The difference is *not*, and this is important, that Davidson shies away from making claims about essences. He doesn't. They are all over the place. The difference is that his claims about essences are primarily claims about what is essential for certain *concepts* of ours, and only secondarily, in consequence of that, claims about what is essential for certain kinds of *objects* or *states*. This does not leave room for a posteriori claims about essences. On a metaphysical reading, you use the concepts to 'hook onto' the objects or states, and then it is those objects or states themselves that (somehow) are the source of the essences. Whatever this exactly amounts to (and by no means do I mean to suggest that this is perfectly clear and unproblematic), it does seem to leave room for a posteriori claims about essences.

So, let's try out this strategy. Let's try to give a Davidsonian argument for the claim that Charity is no mere truth, but one that is metaphysically necessary. To do this, we have to get back to radical interpretation. As we saw earlier, Davidson conceives of radical interpretation as a test that any acceptable theory of meaning and belief must pass. Why?

The significance of radical interpretation for Davidsonian theory of meaning derives from yet another claim about essences. It derives from the claim that language is *essentially public* or *social*. Here is what Davidson says:

> [W]hat has to do with correct interpretation, meaning, and truth conditions is *necessarily based on available evidence.* 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* (2005, 56, all emph. mine).

If you will, this is a statement of a weak, but very basic behaviorism about meaning, a behaviorism reminiscent of that expressed by Quine in the very opening lines of *Word and Object*:

> Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings, unless in terms of men's dispositions to respond overtly to socially observable stimuli (Quine 1960, ix).

Such behaviorism claims:

(B) Meaning is entirely determined by observable behavior.

According to Davidson, (B) follows from the very nature of language, from its essential publicness. The argument runs something like this:[8] Grant essential publicness. That is, if someone speaks a language then others can, in principle, *know* what he means by his utterances. What does it mean to have such knowledge? According to Davidson, such knowledge is empirical knowledge, that is, knowledge based on *empirical evidence*. What kind of evidence is relevant here? For the answer to be an answer of any interest for the theory of meaning, such evidence not only needs to be itself in principle publicly accessible, it also must "not assume in advance the concepts to be illuminated" (2005, 56), that is the concepts of meaning, propositional content etc. Therefore, Davidson concludes, the relevant evidence consists in the speakers' publicly observable behavior. Thus, (B).

It is important here to be clear about the nature of the determined property: It is what one might call 'evidence-constituted'. There is nothing to meaning over and above what is determined by observable behavior; meaning is *entirely* determined by observable behavior. This construction explains charity's double role noted above: For any evidence-constituted property, it holds that the principle metaphysically determining it at the same time provides a method for the property's epistemology. Thus, also the significance of radical interpretation: Radical interpretation is the situation in which exactly such evidence is available as is meaning determining. If meaning is entirely determined by observable behavior, then it must be possible for the radical interpreter to determine meanings on the basis of observable behavior alone. If he can't, something has gone wrong.[9]

Unfortunately, the Quine-Davidson argument as stated above does not go through as it stands. For instance, the claim that knowledge of meaning necessarily is evidence-based knowledge would need to be defended against reliabilist rivals.[10] But let's assume either that matters can be repaired or that (B) can be defended some other way. For our question is what (B) does for Charity. Can the claim that charity is the meaning determining principle be defended on the basis of (B)? (B) does seem to provide some support for Charity: if (B) holds, the meaning determining principle is a

[8] Cf. Pagin 2000 for a similar reconstruction of the Quinean argument from publicness to behaviorism.
[9] For a little more on Davidsonian behaviorism, esp. in relation to current mainstream externalism, see Glüer 2006.
[10] This complaint is Pagin's, see Pagin 2000, 171.

principle using observable behavior as its determination base. That is not strictly true of charity, but nothing in (B) prevents the overall determination of meaning from being two-step; charity determines meaning on the basis of attitudes of holding sentences true, but these in turn are determined by observable behavior. Thus, while (B) certainly excludes some ideas about how meaning might be determined, it does not exclude Charity. Nevertheless, (B) is completely silent about *how* observable behavior determines meaning. And the argument from the public nature of language does *not* get us any further; all we know from it is that there necessarily is *a* relation of determination between observable behavior and meaning. Why think it is Charity (or that Charity is part of it)?

At this point, it needs to be remembered that what we are after is defending an *a posteriori* necessity. As I have argued in the previous section, *which* principle actually determines meaning is ultimately an empirical matter. It is in perfect keeping with this result that the public nature of language does not provide us with an argument for precisely Charity. That it is precisely Charity that does determine meaning is a matter that ultimately requires some empirical argument to establish. Suppose, then, that we have good, ultimately empirical reason to think it is Charity.[11] Our problem right now is not how to establish the *truth* of Charity, but how to get from its truth to its *necessity*. In other words: Why should we think that the principle actually determining meaning does so in all (accessible) possible worlds?

We *do* have some modal intuitions here, I submit. Take a set of very nearby possible worlds *W*. In all worlds in *W*, almost everything is as in the actual world, most importantly, all the evidence in the determination base for meaning, all the data about observable behavior and the attitudes of holding sentences true, is exactly the same. Now the question is: Do we think that that allows for different meanings? In other words, do we think that there is a world *w* in *W* in which everything is the same as in the actual world except for the principle of meaning determination and, consequently, for what the speakers say (and, thus, believe)? I don't think so. If what is in the determination base is the same, the meanings necessarily are the same (up to indeterminacy, of course, if any).[12]

[11] See Pagin's companion to the present paper for a suggestion of such an argument.
[12] Given that the *truth* of Charity is assumed to already be established (empirically), all we need to get to necessity is intuitions of this general kind. The question, therefore, is *not* whether these intuitions single out Charity (as opposed to other candidates for the meaning determining principle), but whether these general intuition reach far enough, that is, all the way to *metaphysical* necessity.

But that is not enough. We need the conclusion that the principle actually determining meaning does so with metaphysical necessity, that is, in all (accessible) possible worlds. But what about all those (accessible) worlds outside *W*, all those worlds where what is in the determination base is not the same as in the actual world? Or, to put this differently: couldn't meaning be such that it is not determined by observable behavior (in observable circumstances)? It certainly seems possible to not even subscribe to the truth of (B) (its truth in the actual world, that is). And even if that is simply believing something false, it certainly seems *conceivable* that meaning is not determined by observable behavior. And even if this might merely indicate epistemic possibility, I don't know of any argument to the effect that it does. My modal intuitions and metaphysical convictions give out at this point. This is a serious gap in our argument -- suggestions as to how to fill it are very welcome.

As long as we don't know how to fill this gap, we have to settle for something weaker than metaphysical necessity. Thus, it might be time to once again fall back on the empirical nature of Charity and to argue that Charity is the one and only empirical claim in the vicinity that has the slightest plausibility, the only one, moreover, with such an enormous amount of confirmatory evidence behind it.[13] In fact, we can go further here and argue for Charity's *lawlikeness*.[14] For a principle like charity certainly would be projectible and counterfactual-supporting, and consequently pretty clearly qualify for psychologico-nomological necessity. That secures some counterfactual worlds where the determination bases for meaning are different from the actual one, but not all of them. Quite possibly, that is all the necessity we can expect a meaning determining principle to have.[*]

Department of Philosophy
Stockholm University

References

Davidson, Donald, 1970, "Mental Events", in Davidson 1980a.

[13] Remember that we still are concerned with how to get from Charity's truth to some sort of necessity, even though weaker than metaphysical. We are still assuming Charity's truth, that is. And if Charity is true, then the evidence for it is in fact massive; then, all the evidence from communicative practice is behind it.

[14] There are, of course, grue-some alternatives, but that is not a problem specific to this putative law. It is therefore not a problem we need to worry about here.

[*] Above all, I would like to thank Peter Pagin for all our discussions of the topics discussed in this paper and its companion. Both papers contain some matter originally contributed by the other, but each of us takes the main responsibility for the paper with our name on. I would also like to thank Åsa Wikforss, Barry Smith, and audiences in Stockholm, London, Prague, and Oslo for their useful comments and criticisms of earlier versions of this paper.

—, 1973, “Radical Interpretation”, in Davidson 1984.
—, 1974a, “Belief and the Basis of Meaning”, in Davidson 1984.
—, 1974b, “Psychology as Philosophy”, in Davidson 1980.
—, 1975, “Thought and Talk”, in Davidson 1984.
—, 1976, “Hempel on Explaining Action”, in Davidson 1980.
—, 1977a, “The Method of Truth in Metaphysics”, in Davidson 1984.
—, 1977b, “Reality without Reference”, in Davidson 1984.
—, 1980a, *Essays on Actions and Events*, 2nd ed. 2001, Oxford: Clarendon Press.
—, 1980b, “A Unified Theory of Thought, Meaning, and Action”, in Davidson 2004.
—, 1983, “A Coherence Theory of Truth and Knowledge”, in Davidson 2001.
—, 1984, *Inquiries into Truth and Interpretation*, 2nd ed. 2001, Oxford: Clarendon Press.
—, 1985a, “A new basis for decision theory”, *Theory and Decision* 18: 87–98.
—, 1985b, “Reply to J.J.C. Smart”, in Vermazen and Hintikka 1985.
—, 1991, “Three Varieties of Knowledge”, in Davidson 2001.
—, 1999, “Reply to Andrew Cutrofello”, in Hahn 1999.
—, 2001, *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press.
—, 2004, *Problems of Rationality*, Oxford: Clarendon Press.
—, 2005, *Truth and Predication*, Cambridge, MA: The Belknap Press of Harvard University Press.
Evnine, Simon, 1991, *Donald Davidson*, Key Contemporary Thinkers, Stanford: Stanford University Press.
Glüer, Kathrin and Peter Pagin, 2003, “Meaning theory and autistic speakers”, *Mind and Language*18, 2003: 23-51.
Glüer, Kathrin, 2006, “Critical Notice: Donald Davidson’s *Collected Essays*”, forthcoming in *Dialectica*.
Hahn, Ludwig E. (ed.), 1999, *The Philosophy of Donald Davidson*, The Library of Living Philosophers Volume XXVII, Chicago and La Salle, Ill.: Open Court.
Lepore, Ernest and Kirk Ludwig, 2005, *Donald Davidson. Meaning, Truth, Language, and Reality*, Oxford: Oxford University Press.
Lewis, David, 1974, “Radical Interpretation”, *Synthese* 27: 331-44.
Ludwig, Kirk, 2003, “Introduction”, in K. Ludwig (ed.), *Donald Davidson*, Contemporary Philosophy in Focus, Cambridge: Cambridge University Press.

Pagin, Peter, 2000, “Publicness and indeterminacy”, in P. Kotatko and A. Orenstein (eds.), *Knowledge, Language and Logic: Questions for Quine*, Boston Studies in the Philosophy of Science, Dordrecht: Kluwer.

—, 2006, “The Status of Charity II. Charity, Probability, and Simplicity”, this volume.

Quine, Willard V.O., 1960, *Word and Object*, Cambridge, MA: MIT Press.

Vermazen, Bruce and Merrill B. Hintikka (eds.), 1985, *Essays on Davidson. Actions and Events*, Oxford: Clarendon Press.