In: Wittgenstein and Davidson on Thought, Language, and Action,
ed. C. Verheggen, Cambridge University Press 2017: 69-96.

Rule-Following and Charity: Wittgenstein and Davidson on Meaning Determination

Kathrin Glüer

1 Introduction

Both Ludwig Wittgenstein and Donald Davidson were deeply concerned with language. Most deeply, they cared about foundational, or meaning theoretic questions, questions of the kind: What is meaning? What is it for words to have meaning? And how do we know what words mean? Not only did they care about the same questions, they also shared some very fundamental ideas about their answers. The opening of Quine's *Word and Object* – "language is a social art" (Quine 1960, viii) – would in fact be an excellent motto for the meaning theoretic thought of all three of them. Whether investigating rule-following, radical translation, or radical interpretation, these thinkers take a distinctive perspective on the metaphysics of meaning: They recognize the constitutive role of its epistemology. In Davidson's words:

> 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 (Davidson 1990b, 56).

But while it has become comparatively standard, not least due to the immense influence of Kripke's *Wittgenstein* (1982), to read the rule-following considerations as concerned with meaning determination, it is less common to also see their epistemological strands in this

light.[1] It is in precisely this light, however, that I think the deep affinity between Wittgenstein and Davidson on meaning determination is best appreciated.[2] Moreover, while it has become fairly standard to structure the debate in terms of the contrast between reductivism and non-reductivism, this very landscape will come to look somewhat different from the perspective suggested here.

In this chapter, I shall assume that the rule-following considerations are – in a sense to be made more specific, however – about meaning determination. My question will be whether the principle of meaning determination used in the early Davidson's account of meaning determination – the principle of charity – provides an answer to what I will call "Wittgenstein's paradox". I shall proceed as follows. In section 2, I shall work out the connection between the rule-following considerations and meaning determination and isolate the aspect of Wittgenstein's paradox I want to explore: the "problem of objectivity". Then, I shall run us through the basics of the radical interpretation account of meaning determination in section 3. In section 4, I shall argue that the principle of charity does seem to fall prey to the problem of objectivity, a verdict I shall ultimately endorse after unsuccessfully trying to rescue objectivity by means of Lewisian natural properties in section 5.

2 Rule-Following and Meaning Determination

The rule-following considerations (roughly, *Philosophical Investigations* 138-242 and *Remarks on the Foundations of Mathematics*, section VI) culminate in what I shall call "Wittgenstein's paradox":

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule (PI 201).

Two very basic questions concerning this well-known passage are the following: First, how

[1] Wright is an exception to this; cf. Wright 1980; 1989; 2007; 2012. There also is a more recent tendency in the literature on Kripke's *Wittgenstein* to play up epistemological demands on a "straight solution", but these do not concern the publicness of meaning. Rather, Kripkenstein picks up on another epistemological aspect of the rule-following considerations: the idea that meanings must be such that they *justify* the speaker's use of linguistic expressions (cf. e.g. PI 217). Cf. a. o. Jackman 2003; Kusch 2006; Guardo 2010; Merino-Rajme 2015.

[2] On this, I take Verheggen 2003 and myself to be in agreement – even though we disagree on how, precisely, the affinity is best developed (cf.Verheggen 2006; Verheggen 2007, Verheggen 2013, Glüer 2006b, Glüer and Pagin 2003).

is this about meaning? And second, how did Wittgenstein get into this "paradox"?[3]

As Åsa Wikforss and I have argued at length elsewhere (Glüer and Wikforss 2010), it is wrong to think that the rule-following considerations are about meaning because Wittgenstein conceives of speaking a language as an essentially rule-guided activity. While this might still be the received view, it is in fact an idea very much at odds with central tenets of the later Wittgenstein – such as the idea that meaning is, or is determined by, use. The rule-following considerations thus do *not* present a deep and subtle defense of the idea that language is a rule-guided activity against Wittgenstein's own paradox. Rather, they provide some deep and subtle reasons for *rejecting* this idea.[4]

By the time of the *Investigations*, Wittgenstein is no longer thinking of language as a calculus of grammatical rules. Instead, he is investigating meaning by means of the *game analogy*, the analogy between speaking a language and playing a game, between meaning and rules. And the game analogy is precisely that: an analogy, and a rich and rewarding one at that. Its most important aspect in this context is this: the meaning of an expression determines its correct application in much the same way that a rule determines a set of actions as being in accordance with it. Also, both meanings and rules are things speakers and rule-followers are said to 'grasp'. It is probably only in *On Certainty* that Wittgenstein fully and explicitly brings these fundamental elements of his late thought together, but there, he clearly sums up the intimate relation between the game analogy and the idea that meaning is, or is determined by, use:

> A meaning of a word is a kind of employment of it. For it is what we learn when the word is incorporated into our language. That is why there is an analogy between the concepts 'meaning' and 'rule' (OC 662).

But as Verheggen neatly brings out when formulating what she calls the "determination problem", trouble is brewing once we put these things together:

> Wittgenstein has reason to wonder how a sign that is meaningful can determine its applications, for he has been developing the view that all there is to the meaning of a word is its use in a language (Verheggen 2003, 289).

[3] There is also the question of in what sense, if any, the problem Wittgenstein formulates here is a *paradox*. I shall not pursue this question; rather, I shall simply use 'Wittgenstein's paradox' as a name for the problem.

[4] We are, of course, by no means claiming that this is all they do.

So, the rough answer to the two questions above is this: Wittgenstein's paradox is about meaning because in the rule-following considerations he is investigating meaning by exploring the analogy between meaning and rules. And he gets himself into the paradox by thinking of the meaning of an expression as "a kind of employment of it", as use.

It will be useful to distinguish between two aspects of the determination problem, however. Both of these concern what we can call the "determination target", i.e. the second relatum of the determination relation in question. According to the game analogy, the meaning of an expression is like a rule in determining *the correct applications* of that expression. This determination target has several worrisome aspects. First of all, it consists of a potential *infinity* of correct applications. The question then is: How can meaning possibly determine a potential infinity of applications*in advance* of these applications actually being made? This question is quite intriguing in itself, but it becomes rather perplexing once combined with the idea that the meaning of an expression is, or is determined by, its use, i.e. its actual applications. How could its use, which is finite, determine an infinity of correct applications for an expression?[5] We can call this aspect of the determination problem the "infinity in advance problem".

Second, what we can call the "objectivity problem" concerns a different property of the target applications of an expression: their correctness. In many cases, these applications need to be such that their correctness is an objective matter if there is to be any correctness at all. The question then is: How can meaning possibly be such that the correctness of the applications of meaningful expressions is an objective matter? And again, the question becomes rather perplexing once combined with the idea that the meaning of an expression is, or is determined by, its use. Wouldn't whatever we do have to count as correct, thus undermining the very idea of correctness?[6]

Distinguishing these problems provides us with a somewhat more precise idea of how the idea that meaning is, or is determined by, use gets Wittgenstein into his paradox. If meaning is use, it is very hard to see how meaning can possibly determine what it is supposed to determine: a potential infinity of objectively correct applications.

It has become customary, I said above, to read the rule-following considerations as about

[5] Cf. Verheggen 2003, 289, Glüer and Wikforss 2010, 156. See also Pagin 2002.

[6] Cf. Verheggen 2003, 289f. I take it to be obvious that both the infinity in advance and the objectivity problem are amongst those most exercising Wittgenstein in the rule-following considerations.

meaning determination. In the light of the problems just spelled out, this might appear questionable or even false. But it isn't. The rule-following considerations *are* about meaning determination – just not in a totally direct way. To see why, let me first explain why the claim might appear to be false. Questions about meaning determination concern what determines meaning, not what gets determined by meaning. In meaning determination, we might say, meaning is the determination *target*. In Verheggen's "determination problem", by contrast, meaning is the not the target, but the determination *base*. Here, meaning does the determining.[7]

Keeping these determination relations distinct, we get a picture like the following:

(DMD) Use $\longrightarrow_{D_1}$ Meaning $\longrightarrow_{D_2}$ Application

Ultimately these relations are supposed to compose, of course: Whatever determines meaning thereby determines correct application. In what follows I shall therefore both speak of meaning and of use determining correct application. Intuitively, both D_1 and D_2 appear to be relations of the same kind or "strength"; both are *metaphysical* determination relations and their holding appears to be a matter of at least metaphysical necessity.[8] Wittgenstein's paradox is that this seems to be an illusion: Nothing metaphysically determines correct applications for an expression.

Wittgenstein himself, identifying meaning with, or at least very closely tying it to, use, does not seem to blame the problem on D_1. The culprit, in his opinion, is D_2. Thus, his problem is how there could be meaning at all even though "no course of action could be determined by [it]" (PI 201). In Kripkenstein's version of the paradox, on the other hand, the problem does not seem to be that meaning does not determine correct application, but that there is nothing that determines meaning. According to Kripke's skeptic, what we need to give up is D_1:[9]

> This, then, is the sceptical paradox. When I respond in one way rather than another to such a problem as '68 + 57', I can have no justification for one response rather than another. Since the sceptic who supposes that I meant quus cannot

[7] Essentially the same distinction is drawn in Pagin 2002, 156f.

[8] Wittgenstein himself arguably thought of these relations as non-contingent and "internal". The obtaining of an internal relation probably amounts to something stronger than metaphysical necessity.

[9] I am following Pagin 2002, 157, fn. 9 here.

> be answered, there is no fact about me that distinguishes between my meaning plus and my meaning quus. Indeed, there is no fact about me that distinguishes between my meaning a definite functions by 'plus' (which determines my responses in new cases) and my meaning nothing at all (Kripke 1982, 21).

But as I said, insisting on this distinction by no means amounts to saying that the rule-following considerations are not about meaning determination. They are, just in less direct and more complicated, more subtle ways. For one thing, Wittgenstein gets into his paradox because of the idea that meaning is, or is determined by, use. It is meaning-as-(determined-by)-use that threatens to fall short of delivering infinity in advance and objectivity. If this result can be made to stick, it tells us something quite stunning about meaning determination: If meaning is determined by use, infinity in advance and objectivity go by the board.

But this is not all. At this point, we might well feel like asking: Why not construe this result as a *reductio ad absurdum* of the idea that meaning is use? Why not go looking for a better idea about meaning determination? The answer suggested by the rule-following seems to be: Because it wouldn't help. Wittgenstein's paradox would still be with us. This even more dramatic conclusion can be seen as motivated in terms of the development of Wittgenstein's thought through the middle and later periods.[10] But the upshot of this development also finds concentrated expression in the rule-following considerations itself. Thus the intense discussions of explanations of rules, of mental states (or processes) of understanding or mental states (or processes) of grasping rules, as well as of calculating machines implementing rules.

Now, if we understand these discussions as proceeding by the method of exclusion, it seems reasonable to doubt that the list of potential "meaning determiners" is exhaustive. But there is more to the method Wittgenstein employs: There is system. A recurring complaint – most prominently concerning explanations and mental states – is that a vicious regress results. If, for instance, we explain the meaning of a word in other words, the explanation will consist of just more linguistic items – items just as much in need of explanation as the one we started with. The basic idea is that whatever item we provide as a candidate for meaning determiner, this item will turn out to be itself in need of getting its meaning

[10] In Glüer and Wikforss 2010, we provide a more detailed reconstruction of this development as it concerns precisely rule-following and meaning.

determined, or of getting "interpreted". Whatever the candidate, it will just bring us back to Square One, back to the very same kind of problem we started with.

To see the underlying mechanics, we do not need to consider all the particular candidates for meaning-determiner. We can instead look at a generalisation of our determination scheme (DMD):

(DMD′) Base $\longrightarrow_{D_1}$ Meaning $\longrightarrow_{D_2}$ Application,

where 'Base' now is a variable ranging over (sets of) candidates for meaning-determiner. The first observation then concerns determination relations quite in general. It is this: Whatever you put into a determination base, it will not, just by itself, determine your target. Basically, determination relations are (one-one or many-one) functions from one domain or set of entities S_1 to another domain or set of entities S_2.[11] And if all I tell you is that something is a function of a certain kind, or set, of items S_1, there is a sense in which that doesn't tell you much at all: For any S_1, S_2, there are sufficiently many such functions to make it the case that for any pair of items s_1 from S_1 and s_2 from S_2 there is a function mapping s_1 onto s_2. Translated into Wittgensteinian terms, this amounts to the following claim: Whatever kind of item we put into Base, "every course of action can be made out to accord with [it]", that is, every possible application can be "made out" to be correct. Providing a determination base therefore is only part of the answer to any determination question. We also need to know the relevant function or principle of correlation: We need to know by what *principle* the items in the determination target are determined by those in the determination base.

And the crux of the rule-following considerations is that this need for a principle itself poses a problem, a problem of the very same kind as the one we started with. As soon as there is more than one candidate for being the relevant principle of correlation, the worry is, we are back to Square One. For then, it needs to be determined which principle is the right one. As Pagin puts it with respect to states of understanding: "if meaning is to be determined by a state of understanding, then the state of understanding must also select the principle of correlation, and this brings us back to the first problem again" (Pagin 2002,

[11] Relations of metaphysical determination have, of course, properties over and above being such functions, properties such as having a direction (from the metaphysically more fundamental to the metaphysically less fundamental) and, possibly, providing a special kind of explanation, but these properties do not matter at this point.

160).[12] As long as we do not provide *justification* for picking a particular such principle our account of meaning determination is in trouble.[13]

It is at this point that turning to Davidson seems so promising. For as we already began to see in the introduction, Davidson's account of meaning determination aims at providing precisely the ingredients we need now: A determination base, a determination principle, and, crucially, justification for both the choice of base and the choice of principle. Moreover, there is that deep affinity between, on the one hand, these Davidsonian choices and their justifications and, on the other hand, the answers at least suggested by the later parts of the *Investigations*, answers turning on the social character of language and crucially involving agreement (in primitive reactions and judgments) as well as a common way of understanding, and explaining, action.[14,15] In the next section, we shall therefore turn to David-

[12] As Pagin points out, this holds equally well for *dispositions*. In the context of the rule-following considerations, the real problem with meaning determination by dispositions is neither their (supposed) finiteness nor that we are disposed to make mistakes. The problem is that, like any other items we might put into the determination base, dispositions by themselves do not "pick" a principle of correlation. Unless such a principle is specified, we can map any meaning/correct applications whatsoever onto any expression from any given set of dispositions regarding that expression. In Kripke's discussion of dispositions, a rather simple such principle is in fact implicitly assumed: According to this principle, the uses a speaker is disposed to make of an expression simply are the correct uses. But why this principle rather than any other principle linking dispositions to correct applications? This question *cannot* be answered by merely complicating the principle of correlation. What is needed is *justification* for picking a particular principle instead of any of the others. Cf. Pagin 2002, 161.

[13] This holds largely independently of the question of whether we think of an account of meaning determination as in the business of *reduction* or not. As long as what we put into the determination base does not, somehow, already select the principle of correlation, the problem recurs. Arguably, it recurs for both reductive and non-reductive, but informative accounts. The only kind of "account" that might remain unaffected is a sort of *quietism* about meaning determination insisting that meaning facts determine meaning facts. Note, however, that this is not because the principle of determination is identity. There are many ways of mapping a domain onto itself. Rather, the thought would have to be that for a fact to be a meaning determining fact it needs to be such that it can obtain while the meaning fact it determines obtains. This would not be the case for a mapping from meaning facts to meaning facts which does not map every meaning fact onto itself.

[14] In PI 206, Wittgenstein asks us to imagine coming "as an explorer into an unknown country with a language quite strange to [us]". And he refers to the idea of the "common behavior of mankind" ("die gemeinsame menschliche Handlungsweise") as a "system of reference by means of which we interpret an unknown language". Cf. also PI 243. One commentator who reads these passages in very Davidsonian terms, indeed, is Hopkins (cf. Hopkins 1999; Hopkins 2012). According to Hopkins, Wittgenstein solves his paradox by construing meaning as *constituted* by ideal radical interpretation, i.e. by adopting a form of *judgment-dependence* about meaning. There are commentators ascribing this kind of "interpretivism" to Davidson, too (cf. for instance Byrne 1998 and, possibly, Williamson 2004, 137). As Wittgenstein is concerned with meaning *and* content determination, such an interpretivism would be rather obviously question-begging, however (cf. Gross 2015). And the same goes for Davidson. In both cases, exegetical charity rules these interpretations out. See Boghossian 1989, 546f for discussion of the more general idea (proposed in Wright 1989) of construing meaning or content as response-dependent.

[15] It is of course, by no means obvious what Wittgenstein's own solution, if any, to his paradox is supposed to precisely consist in. I shall not enter into these discussions here. For a comparison between Davidson and Wittgenstein on the role of agreement, see Glüer 2000, however.

sonian meaning determination, more precisely, to the account of meaning determination that Davidson offers in a series of papers concerned with radical interpretation.

3 Radical Interpretation

In this section, I will run us through the basics of Davidsonian radical interpretation as an account of meaning determination. My focus will thus be on the "early" Davidson and a series of papers we can call "the radical interpretation papers".[16] My aim here is to provide just enough background for understanding the Davidsonian answers to the following questions: What is in the determination base for meaning? What is the principle of meaning determination? Why are these the base and principle of meaning determination – or: what justifies the Davidsonian choices of base and principle?

To fully appreciate the Davidsonian choices and their justification, we need to be clear about the basic perspective he takes on meaning. According to Davidson, meaning is a theoretical concept. Its main purpose is the explanation of successful communication by language. The same holds for concepts like those of reference, predicate, or sentence; their "main point ... is to enable us to give a coherent description of the behavior of speakers, and of what speakers and their interpreters know that allows them to communicate" (Davidson 1992, 108f).

What speakers and their interpreters know when they successfully communicate by language are things like the following: what someone said, what the uttered expressions mean, and how to express a certain thought in language. This knowledge is the output or result of what we can call our linguistic ability or competence. Concentrating on the interpretive side of this ability, Davidson suggests approaching the fundamental meaning theoretical question – the question "What is it for words to mean what they do?" – indirectly: by means of two others. Classically, these are formulated in the course of the opening paragraph 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

[16] The most important of the radical interpretation papers are Davidson 1973, Davidson 1974, Davidson 1975, Davidson 1976. A useful overview can be found in Davidson 2005. By contrast,Verheggen (1995; 2000; 2006) draws on affinities between the later Wittgenstein and the *"later"* Davidson on meaning and triangulation. For more on triangulation, see also Pagin 2001; Glüer 2006b; Verheggen 2007; Verheggen 2013.

> 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?* (Davidson 1973, 125, emph. added.)

And some years later, in the introduction to the collection *Inquiries into Truth and Interpretation*, Davidson recapitulates his project:

> What is it for words to mean what they do? ... I explore the idea that we would have an answer to this question if we knew how to construct a theory satisfying two demands: it would provide an interpretation of all utterances, actual and potential, of a speaker or group of speakers; and it would be verifiable without knowledge of the detailed propositional attitudes of the speaker (Davidson 1984, xiii).

As is well known, Davidson proposes that the theory we are after is a formal semantic theory for a natural language L. According to him, such a theory is compositional, and takes the form of a Tarskian truth-theory (t-theory).[17] A t-theory for a language L is supposed to give the meaning of each sentence of L by specifying its truth-conditions. And according to Davidson, the meaning of an expression of L is precisely its systematic contribution to the truth-conditions of the sentences it occurs in, a contribution spelled out by the correct t-theory for L.[18] By using a t-theory as a formal semantic theory for a natural language L, Davidson submits, we can describe or model the linguistic competence that allows for interpreting utterances in L.[19] The resulting knowledge is empirical knowledge. And a formal semantic theory for a natural language L is an empirical theory – it is an empirical question whether any particular such theory is correct for L, i.e. gives the right meanings for utterances in L. Such knowledge is based on evidence, justified by empirical data. What are the data supporting formal semantic theories for natural languages?

[17] The reader unfamiliar with this framework might for instance consult my (2011), where a t-theory for a fragment of English is provided in the Appendix.

[18] There is no need, Davidson submits, to assign *entities* – such as propositions – as meanings to expressions: "My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use" (Davidson 1967, 21).

[19] It is important to distinguish between two potential objects of knowledge here: According to Davidson, what speakers know is the output or result of their linguistic competence, i.e. what utterances mean. This does *not* mean that they know the theory by means of which we model that competence (cf. Davidson 1986, 96). Thus, the hypothetical form the question takes in "Radical Interpretation": "What could we know that would enable us to do this?"

According to Davidson, there are two major restrictions on the foundationally interesting data for semantic theorizing. First, to learn something about what meaning is, we must be able to formulate these data in terms not presupposing meaning or any other semantic notions – or "without essential use of such linguistic concepts as meaning, interpretation, synonymy, and the like" (Davidson 1973, 128; see also Davidson 1974, 142f). To learn something about what meaning is, we thus are after the "ultimate evidence" (Davidson 1973, 128) for any correct theory of interpretation. And second, since we are talking about knowledge that any competent speaker has, the data must be "plausibly available to a potential interpreter" (Davidson 1973, 128), where the interpreter is just another competent speaker. What kind of data is there that could do this job? This is the question that leads us directly to the idea of radical interpretation.

In radical interpretation, the interpreter is to construe a t-theory for a radically foreign language, a language she doesn't know anything about at the start. The only evidence available to her consists of data about the behavior of the speakers and the observable circumstances in which it occurs. Such data, Davidson holds, allow for the identification of a certain kind of attitude the speakers hold to uninterpreted sentences: the attitude of *holding an uninterpreted sentence s true* (at a time *t*). This *is* an intentional attitude, in fact, a belief, but it is an attitude of the kind that Davidson calls "nonindividuative" (Davidson 1991, 211): the interpreter can know that a speaker holds this attitude towards *s* at *t* without knowing what *s* means and, thus, without knowing *which* belief the speaker thereby has. Thus, no meaning (or content) theoretical questions are begged when using data about this attitude in the account of meaning determination.

The only thing special about the radical interpreter then is that, in contrast to an ordinary competent speaker, she has huge amounts of such data at her disposal: "We may as well suppose", Davidson writes, "we have available all that could be known of such attitudes, past, present, and future" (Davidson 1974, 144).[20]

The radical interpreter thus collects vast amounts of data like the following:

(E) Kurt belongs to the German speech community and Kurt holds true 'Es regnet' on

[20] What this precisely means is not so easy to understand, however. On the one hand, Davidson is concerned with making sure that the available evidence can support sufficiently many of the differences in meaning it is pre-theoretically plausible to think we can detect. But on the other hand, we need to be careful not to stretch the limits of the evidence available to the radical interpreter beyond recognition. After all, her evidence is supposed to be accessible to an everyday interpreter.

Saturday at noon and it is raining near Kurt on Saturday at noon.

Sufficient numbers of observations like (E) then are supposed to support t-theories from which theorems like the following t-sentence can be derived:

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

Subscribing to Quinean confirmation holism, Davidson construes the relation between theory and evidence as a holistic one. It is whole t-theories that are supported by the data to varying degrees. "[T]he method," Davidson explains, "is (...) one of getting a best fit" (Davidson 1973, 137).[21]

But what is it for data like (E) to "fit" t-theories entailing (T)? The basic idea is to assign the conditions under which speakers hold sentences true as the truth conditions of those sentences. But speakers hold all sorts of things true under all sorts of circumstances. More precisely, whether a speaker holds a sentence true under given circumstances depends crucially on the (further) beliefs of the speaker. Take a speaker who (erroneously) believes that there is an elaborate system of sprinklers on the roof. Upon looking out of the window on a rainy Saturday at noon, such a speaker might not only fail to believe that it is raining, but even form the belief that someone must have turned those sprinklers on. This is just one example of the pervasive phenomenon Davidson calls the "interdependence of belief and meaning". The interdependence

> is evident in this way: a speaker holds a sentence to be true because of what the sentence (in his language) means, and because of what he believes. Knowing that he holds the sentence to be true, and knowing the meaning, we can infer his belief; given enough information about his beliefs, we could perhaps infer the meaning (Davidson 1973, 134f).

Just by themselves, observations like (E) thus do not provide any evidence whatsoever for a t-theory. As long as the interpreter can ascribe any old belief, be it ever so weird or absurd, all such observations can be squared with any old t-theory. It is here that the *principle of charity* is supposed to kick in. In one of its earliest formulations, Davidson formulates the principle as follows:

[21] For more detailed accounts of the method of radical interpretation, see Lepore and Ludwig 2005 and Glüer 2011.

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

It provides a method for solving the problem of the interdependence of belief and meaning

> by holding belief constant as far as possible while solving for meaning. This is accomplished by assigning truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right (Davidson 1973, 137).

The radical interpreter, that is, tries to hold belief constant both between himself and the alien speaker, but also for the alien speaker over time. Now, we can see why Davidson proposes to "take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima facie evidence that the sentence is true under those circumstances" (Davidson 1974, 152). If beliefs, and belief ascriptions, are restricted by the principle of charity, as Davidson argues they are, then a sentence's being held true under certain circumstances does provide evidence that it is true under those circumstances.

The evidence is *prima facie*, however. That is, it can be overridden by other, stronger evidence. People do make mistakes, and some of the vast number of data that the radical interpreter collects will have to be considered as overridden by others. But which? Davidson: "The basic methodological precept is (...) that a good theory of interpretation maximizes agreement. Or, (...) a better world might be *optimize*" (Davidson 1975, 169).[22]

It is not only truth that is important here, however. Mistakes can also come in the form of incoherence, in the form of drawing the wrong inferences from what one believes. Beliefs, Davidson maintains, come in coherent clusters, if they come at all (cf. Davidson 1977, 200). Take Fido, the dog. While we might very well be in a situation where it is plausible to think that someone believes that Fido is a car, this very hypothesis is instantly made much less

[22]Optimizing agreement involves weighting mistakes or disagreements and minimizing a theory's overall score. This is better than simply counting the number of mistakes a theory ascribes (which might not even be possible given that the number of sentences in a language is infinite) because "some disagreements are more destructive of understanding than others" (Davidson 1975, 169). Usually, being wrong on simple observational matters such as whether it rains around one is more destructive than disagreement on highly theoretical matters. Being wrong about one's own mental states or about how things look to one is worse than being wrong about other's mental states or about how things are. The general idea is that a mistake is the more weighty the more epistemologically basic it is, the more basic, that is, to the totality of our knowledge: "The methodology of interpretation is, in this respect, nothing but epistemology seen in the mirror of meaning" (Davidson 1975, 169).

plausible if we also think that they at the same time fail to draw obvious inferences – such as that Fido is an artifact, not an animal – from it. This pressure towards coherence extends all the way to a subject's actions – the hypothesis that you believe that Fido is a car comes under pressure, too, if we observe you carrying a leash while muttering 'Fido needs to be taken out a bit'.[23]

The maxim to make the speaker right when plausibly possible encompasses both these elements: To make the speaker right when plausibly possible is to optimize the beliefs ascribed in such a way that they are, at least in basic cases, mostly true and coherent. Charitably interpreted speakers therefore always come out as persons of a certain, basic *rationality*. In later writings, Davidson sometimes explicitly separates the two components of charity – truth and coherence:

> 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 logic, the other endows him with a degree of what the interpreter takes to be true belief about the world. 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 (Davidson 1991, 211).

We shall have occasion to get back to this formulation of charity later. Here and now, we can summarize charity's role in radical interpretation. Given the methodology of best fit and

[23] Here we have, then, those elements of meaning determination crucial also to the later, post-rule-following Wittgenstein: agreement in judgement (and primitive reactions) and a common way of making action intelligible. Davidson in fact came to think that we ultimately needed "A Unified Theory of Thought, Meaning and Action" (Davidson 1980).

the interdependence of belief and meaning, the principle of charity is supposed to fulfil two essential functions: It

1. restricts belief ascription so that observations like (E) provide data for t-theories, and
2. ranks t-theories by fit to the totality of available data such that the best is/are correct.[24]

So here are, in rough outline, the answers to the two questions Davidson poses at the beginning of "Radical Interpretation". According to him, Tarskian t-theories can be used as formal semantic theories for natural languages. The data supporting them ultimately is data about the behavior of speakers and its observable circumstances. To round off the picture, we need to connect two more dots: We need to connect the epistemology of meaning with its metaphysics. Why do we learn something about what meaning *is* from learning how to justify semantic knowledge? Why would the epistemology of meaning tell us something about its metaphysics?

Davidson's – Quine-inspired – answer to this question is encapsulated in the following claim: "What a fully informed interpreter could learn about meaning is all there is to learn" (Davidson 1983, 148). There are no meaning facts beyond those that can be known on the basis of evidence available to the interpreter.

> Quine revolutionized our understanding of verbal communication by taking seriously the fact, obvious enough in itself, that there can be no more to meaning than an adequately equipped person can learn and observe; the interpreter's point of view is therefore the revealing one to bring to the subject (Davidson 1990b, 62).

But intuitively, it is quite unusual for evidence to have such epistemico-metaphysical double significance; with respect to most objects or properties we do not think that the facts about them are exhausted by the evidence available to us. Why would meaning be different? Because meaning is *essentially public*:

> 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...

[24] According to Davidson, it is pretty much inevitable that there will be more than one "best" theory.

> 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 (1990, 56, emph. added).

Language is essentially social: Meanings are such that they can be understood. This, for Davidson, is the most fundamental thing about language. And, as I said right at the beginning of this section, in Davidson's hands, meaning is nothing more than a theoretical notion used to explain linguistic communication. So, he argues, there cannot be more to meaning than what we can know about it. The data for the t-theory the radical interpreter is after therefore have an epistemico-metaphysical double nature. They epistemically support the theory, but at the same time, Davidson tells us, they "entirely determine" the very thing the theory is a theory of. The data available to the radical interpreter thus form the determination base for meaning. Meaning is an evidence-constituted property.

By taking this perspective Davidson is able to suggest justifications for both his choice of base and his choice of principle for meaning determination. Strictly speaking, Davidsonian meaning determination is two-step: First, observable behavior in observable circumstances determines attitudes of holding true towards uninterpreted sentences. And in the second step, attitudes of holding true towards uninterpreted sentences determine, via the principle of charity, meanings for those sentences (and, simultaneously, contents for beliefs (and other propositional attitudes)). Davidson focuses almost exclusively on the second step, and so shall we. The base we are concerned with thus contains attitudes of holding uninterpreted sentences true, and meanings are determined for those sentences by means of the principle of charity. According to Davidson, these choices are justified by what he argues is a "constitutive aspect of language": the public nature of meaning.

That meanings are knowable is indeed a constitutive aspect of language, it seems to me. And that the semantic notions are theoretical notions whose purpose is the understanding or explanation of successful linguistic communication also seems extremely plausible. Moreover, these two insights go hand in hand, and once fully adopted, they do seem to

unlock resources unavailable from the austere, "platonist" perspective of the rule-following considerations, resources for constraining both what is a reasonable candidate for meaning-determiner and what is a reasonable candidate for meaning determining principle. It is thus rather natural to wonder whether adopting this Davidsonian point of view might not make answers available, answers congenial to Wittgenstein, to the very questions the rule-following considerations are tempting us to despair of.

4 Casey, Alien, and the Problem of Objectivity

We have seen why it is tempting to think that charity might be an answer to the rule-following considerations. A thorough investigation of this idea would require looking into more aspects of charity than I have space for here. One question concerns the epistemic and modal status of charity. The rule-following considerations seem to be premised on the relevant determination relations being very strong modal relations. And while some commentators argue for Davidsonian charity being a conceptual necessity (and knowable a priori – cf. a.o. Lepore and Ludwig 2005), I find that both implausible and exegetically strained. It is better, or so Peter Pagin and I have argued elsewhere, to construe Davidsonian charity as an aposteriori necessity, most plausibly a nomological one.[25] But in that case, how much of an answer to the rule-following considerations can we hope to get from charity?

Another question concerns the justification of charity. How, precisely, does the publicness of meaning justify the choice of charity as the meaning determining principle? A common assumption amongst commentators appears to be that this is relatively straightforward: Charity is justified because it makes meaning knowable (cf. a.o. Lepore and Ludwig 2005). But to avoid this being an instance of affirming the consequent, we would also need to argue that no other principle would make meaning knowable. In radical interpretation, this means arguing that no other principle would make holding uninterpreted sentences true into data for t-theories. But it seems fairly obvious to me that any principle that assigns truth-conditions as a function of holding true attitudes would do precisely that.[26] Of

[25] See our companion papers "The Status of Charity I & II": Glüer 2006a; Pagin 2006.

[26] Ludwig disagrees:

> Glüer objects that charity isn't needed to succeed at interpretation. Any principle will do. Is that right though? What about the principle that says that the interpreter should interpret the speaker as massively wrong about his or her environment? How does that enable the interpreter to use her evidence to gain access to a detailed picture of the speaker's meanings and attitudes? It is no guide whatsoever. Or suppose the principle is to assign beliefs about prime numbers on the basis

course, we might feel that alternative such principles would not result in what we intuitively think are the correct meaning assignments for our actual languages, but if that feeling is justified, it is not justified by such principles' making meaning unknowable. The justification for charity therefore must at least in part originate elsewhere. As I have argued (Glüer 2006a; Glüer 2011, esp. ch. 3), Davidson himself locates it in the very nature of belief – belief, he argues in numerous places, "is in its nature veridical" (Davidson 1982, 146).[27] If there are any beliefs at all, that is, they come in largely true and coherent clusters. For Davidson, it is be-

> of the number of words in sentences held true. Does that relate the interpreter's evidence to how she interprets the speaker? Charity gives us a principle that shows us how the evidence marshaled can be brought to bear systematically upon our interpretation of the other as a speaker and an agent. It is not clear that there is another principle that can do the same job (Ludwig 2014, 469).

The second principle Ludwig considers is quite irrelevant, as it is not a principle "determining truth conditions on the basis of holding true attitudes" (as I said was required; Glüer 2011, 143). But the first could be used just as much as charity, it seems to me. Charity makes holding true into evidence because it allows us to "take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima facie evidence that the sentence is true under those circumstances" (Davidson 1974, 152). But if what we could call "anti-charity" holds, i.e. if speakers are (to be construed as) massively wrong about their environment, we can take the fact that they hold a sentence true under observed circumstances as prima facie evidence that the sentence is *false* under those circumstances. I see no reason to think that this method could not be used to construct a "detailed picture of the speaker's meanings and attitudes". The resulting picture might not be correct for actual speakers, but that is irrelevant here. The question is whether meanings would be knowable on the basis of data about holding true, if meanings were determined (on the basis of holding true) by a principle other than charity. Again, it seems to me fairly obvious that the answer is yes for *any* principle assigning meanings to sentences on the basis (or as a function) of holding uninterpreted sentences true (under observable circumstances).

[27] Ludwig agrees, but thinks that this claim is too controversial for charity to rest upon. It needs to be argued for itself, and according to him, the argument Davidson provides for belief's veridical nature is indirect. It goes precisely via radical interpretation and charity's being the only available principle allowing us to use our evidence (cf. Ludwig 2014, 468f). I quite agree that belief's veridical nature requires argument and that Davidson provides such arguments. Here's one that's quite typical:

> A belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about. Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter (Davidson 1975, 168).

But these arguments do not seem to go via radical interpretation; in fact, they seem quite direct to me. Here's another, as far as I can tell clearly arguing from what is possible regarding belief to what it means to ascribe beliefs:

> 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 (Davidson 1977, 200).

cause of this essential fact about beliefs that the radical interpreter can use charity to break into the interdependence of belief and meaning. The choice of charity as the meaning determining principle is justified, according to him, because meaning is essentially public *and belief is essentially veridical.* But isn't this latter claim way too philosophically ambitious, or way too contentious, for basing an answer to the rule-following considerations on?

While intriguing, these are not the questions I am going to focus on here. Rather, I'd like to go back to Wittgenstein's paradox. More precisely, I want to go back to that aspect of the paradox that I called the "problem of objectivity" above. Meaningful expressions have conditions of correct application. And, in many cases at least, the correctness of an application has to be an objective matter if there is to be any correctness at all. The correctness of the use of our expressions must have a certain sort of independence from us – it must, for instance, be independent of our actual applications, judgements, and other relevant kinds of reactions. Assuming that the meaning of an expression is determined by its use, the question is: How could the use of an expression ever "ground" this kind of objectivity? Wouldn't whatever an expression is applied to have to count as correct, thus undermining the very idea of correctness? What I am going to investigate in the remainder of this paper is how *meaning as determined by charity* fares with respect to the objectivity problem.

I am going to take for granted that charity does not preclude the ascription of a plausible amount of mistakes to normal human speakers. It is for instance perfectly possible for the radical interpreter to have sufficient evidence for interpreting a speaker's predicate 'spunk' as expressing the concept *beetle*, thereby ascribing a mistaken belief when the speaker applies 'spunk' to the occasional spider. What I shall look at are rather situations in which the radical interpreter cannot for the life of her figure out what an alien speaker could possibly mean by what seems to be a predicate. For it at least appears to be possible to construe t-theories for such "speakers" that satisfy charity – but without allowing the interpreter to understand the speaker. Here is Pagin's case of *Casey and Alien*:[28]

> Assume that interpreter Casey from Earth embarks on the interpretation of apparent speaker Alien from Outer Space. Casey identifies a candidate predicate

[28] Pagin here makes use of the old idea that the radical interpreter, when faced with the task of interpreting speakers of a language with conceptual resources much more advanced than her own, can learn from the speakers and acquire their concepts in the process of radical interpretation, adding new terms for those concepts to her own, i.e. the meta-language in which she formulates her t-theory. Cf. Harman 2011, 17.

> Φ that seems applied to some objects and withheld from others by Alien, but Casey sees no pattern in the usage. None of the property concepts Casey can come up with matches even approximately the pattern of Alien's applications.
>
> Casey then decides to learn from Alien, and starts defining a new predicate *F* in his own language. It is defined by cases: true of objects that Alien applies Φ to, false of objects that Alien withholds Φ from, and for all objects *b unconsidered* by Alien, *F* is true of *b* just in case *b* is a *rocket*. Clearly, by interpreting Φ to mean *F*, and assuming Casey has identified atomic sentences with Φ as predicate and a demonstrative as subject term, Alien's demonstrative Φ sentences all come out true.
>
> Casey then goes on to do the same with other predicates, and also with what he identifies as grammatical particles, and sentence constructions. For each sentence held true at a time, on a case by case basis, an interpretation is given of the parts and the syntactic operations that makes the sentence come out true at that time. Some arbitrary interpretation is provided for all cases *not* considered by Alien. So Casey's meaning theory is compositional and complete (with respect to the syntax he has identified), and results in only true beliefs being attributed to Alien (Pagin 2013, 236).

Casey's method – though ingenious – is no good as a method of interpretation. It is no good in two respects. First, despite appearances to the contrary, Alien might in fact not be speaking any language at all. Using Casey's method, there is no way an interpreter could come to the conclusion that what initially appears to be speech behavior in fact is not. But, second, even if we assume that Alien is speaking a language, Casey's method is faulty: Casey's theory clearly does *not* allow him to understand Alien.

Take Φ. What Casey is lacking when it comes to Φ is an *independently possessed concept* that would allow him to subsume at least a (weighted) majority of the objects Alien appears to apply Φ to. As long as Casey is unable to come up with, or form, such a concept to interpret Φ as expressing, he will not understand Alien. And Casey will not be able to come up with, or form, such a concept as long as he cannot detect any similarity in the *F* objects.[29]

[29] But isn't the main problem with the theories Casey's method delivers that they don't make the right *predictions*? No, not really. First, it isn't so clear whether Davidsonian t-theories are even supposed to be predictive.

Charity thus appears to allow meaning assignments that do not secure objectivity. An object (that Alien has considered) satisfies Φ precisely in case Alien has applied Φ to it.[30] Objectivity couldn't be much further away.[31] What's special about Casey is the combination of showing that, while this kind of "meaning determination" appears to be compatible with charity, it has the consequence that a predicate need not express anything recognizable as a meaning whatsoever. And the root of the problem here seems to be precisely that the "meanings" assigned by Casey's method do not secure objectivity: there is no sense in which these meanings make the correctness of Alien's applications of Φ an objective matter, independent of Alien's actual applications. Let's call meanings that do secure objectivity "objective meanings". We can then summarize Casey's case as follows: Satisfying charity does not seem to guarantee that the meanings so determined are the genuine, "objective" article. Meanings that are both public and objective, or so it now seems, require *detectable similarity*.[32]

We need to be careful here, however. If we look beyond its early formulations, there is a good question whether Casey's method really satisfies charity. Take the following passage (again):

> The process of separating meaning and opinion invokes two key principles which

'Fit' is a merely retrospective notion, and, as we saw above, it is some supposed totality of data, "past, present, and future", that according to Davidson determines meaning. Once you have that totality, there is nothing left to predict (cf. Pagin 2013, 237). Pagin (1999) emphasizes that the method of applying charity is a matter of getting a best fit to the data and, thus, to achieve accommodation, as opposed to prediction. He argues that because of this, compositionality cannot by justified from charity.

And second, even if Casey had the relevant totality of data, and thus would not need to make any predictions, his theory would be useless. The *main* problem of Casey's theory is not that according to it, unconsidered objects fall under Φ if they are rockets: Even if Casey had observed Alien consider *every* object, he wouldn't understand Alien precisely because he doesn't detect any similarity in the things that Alien does apply Φ to.

[30] If Alien hasn't considered an object, whether it satisfies Φ depends, not on Alien's application, but on whether it's a rocket or not. So that part is fine.

[31] This is because the "meanings" Casey's theory assigns do not reflect "real", independently possessed, concepts of his. There are perfectly kosher concepts in the vicinity, of course: For Φ, there is the concept of *either being such that Alien has considered and applied* Φ *to one or such that Alien has not considered one and being a rocket*. That, however, clearly is *not* the concept, if any, Alien expressed when using Φ. Nor is it the concept expressed by *F* as defined by Casey, even though these are co-extensive.

[32] Once we recognize the need for using our own concepts in interpretation, considerations like these can be turned into an argument *for* semantic holism: Even if we require our semantic theory to be compositional, and it's composition rules to be projectible, we would need to accept intuitively senseless interpretations if we didn't construe the determination relation between data and meanings as one of *global* best fit. It is only if meaning is determined holistically, and by means of a many-to-one determination relation allowing for the ascription of false belief, that the possibility of interpretation in terms of our own concepts is secured. For an argument like this see my (2001).

> 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 logic, the other endows him with a degree of what the interpreter takes to be true belief about the world.(Davidson 1991, 211, emph. added).

Here, the principle of correspondence can plausibly be read as building detectable similarity right into charity: It 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".

But if charity requires detectable similarity, it might seem as though we have hopped out of the frying pan – and right into the fire. This is because detectable similarity here is similarity detectable *by the interpreter*. This, it might seem, would take the objectionably "subjective" element from Casey's method and replace it with another, potentially equally objectionable "subjective" element: the interpreter's own sensitivity to the similarity in question. If charity requires detectable similarity, the worry is, the interpreter plays an essential role in meaning determination. She couldn't be, as David Lewis once put it, "a way of dramatizing our problem – safe enough, so long as we can take it or leave it alone" (Lewis 1974, 334). And if the interpreter plays an essential role in meaning determination, we are stuck with the objectivity problem. Or so the worry goes.

5 Objectivity and Sensitivity

I think it is very plausible to think that Davidson did incorporate detectable similarity into charity.[33] And ultimately, that does mean that charity falls short as an answer to the rule-following considerations. Meanings, as determined by charity, won't be fully and pristinely

[33]He more and more came to stress the causal (and social) externalist element in his account of meaning and content determination, construing the "features of the world" that both speaker and interpreter react to as the "common causes" of these reactions: "Communication begins where causes converge" (Davidson 1983, 151) could be the slogan of this development. This kind of externalism requires shared sensitivities, commonly detectable similarities across speakers and their interpreters:

objective. But it is important to understand precisely what kind of contribution the interpreter makes to meaning determination.

Most significantly, what requiring detectable similarity amounts to is a restriction – not on the determination base, but *on the "things" eligible to be meanings*. Think of Casey again. What his theory assigns to Φ doesn't intuitively qualify as a meaning – in the sense of something we can mean, i.e. something that can correctly be assigned to a predicate of natural language as its semantic value. This is not because it's not a property – for all I care, it's a property alright. Nevertheless, it does not seem to be the right kind of property – for there to be detectable similarity, a property needs to be such that the interpreter is sensitive to its being instantiated. To be eligible to be meant – i.e. such that a predicate is satisfied by an object iff the object has it – a property needs to be what I shall call "interpreter-sensible".[34] But that would seem to imply that we get "objective" meanings only if we restrict the realm of what can be meant to the properties we are sensitive to. That is, the correctness of the applications of our words can be "objective" only to the extent that we have already restricted what can be meant to those properties we can detect. And this would mean that correctness of application ultimately is not fully objective, not totally independent from our judgements. This is a failure of objectivity in a more subtle sense than the most flagrant one. Objectivity fails in the most flagrant sense if, given that we mean something by an expression, correctness coincides with whatever we apply it to. It fails in a more subtle sense if what is eligible to be meant (at least in the basic cases) coincides with what we can detect, i.e. have an ability to judge correctly. Of course, an ability to detect the instantiation of a property does not require being correct in every single ascription of it. Nevertheless, if we restrict the eligible to the detectable, correctness of application will no longer be fully independent from our judgments: there wouldn't be correctness without the ability to judge correctly.

> What makes these the relevant similarities? The answer again is obvious; it is we, because of the way we are constructed (evolution had something to do with this), who find these responses natural and easy to class together. If we did not, we would have no reason to claim that others were responding to the same objects and events (i.e. causes) that we are. It may be that not even plants could survive in our world if they did not to some extent react in ways we find similar to events and objects that we find similar. This clearly is true of animals; and of course it becomes more obvious the more like us the animal is (Davidson 1990a, 202).

Similarly, Quine came to speak of "a preestablished harmony, between your standards of perceptual similarity and mine" that is required for any "meeting of minds" (Quine 2000, 2; see also Quine 1995, 21f).

[34] This is not meant to suggest that it has to be what usually is called a "sensible property", i.e. a property directly detectable by perception. The interpreter just needs to have some sort of ability to "detect" it, not necessarily a purely perceptual one.

Talking of eligibility here is of course meant to direct our thoughts to David Lewis. If what we are after is a restriction of the things eligible to be meant, we ought to consider whether Lewis doesn't have a solution on offer, an impeccably objective solution moreover. The idea would be to use Lewisian *natural properties* to secure detectable similarity: Properties need to have a certain degree of naturalness to be eligible to be meant, and properties such as those assigned by Casey are too unnatural.

In his 1983 paper "New Work for a Theory of Universals", Lewis at least seems to be suggesting such a constraint on meaning determination when he writes

> that the saving constraint concerns the referent – not the referrer, and not the causal channels between the two. It takes two to make a reference, and we will not find the constraint if we look for it always on the wrong side of the relationship. Reference consists in part in what we do in language or thought when we refer, but in part it consists in *eligibility of the referent*. And this eligibility to be referred to is a matter of *natural properties* (Lewis 1983, 371, see also Lewis 1984, 226ff).

That naturalness might help precisely with detectable similarity is further suggested when Lewis gives general job descriptions for the natural properties like the following:

> Natural properties would be the ones whose sharing makes for resemblance, and the ones relevant to causal powers. (...) Let us say that an adequate theory of properties is one that recognises an objective difference between natural and unnatural properties; preferably, a difference that admits of degree (Lewis 1983, 347).

Eligibility then is supposed to be one of the measures by which to rank assignments of reference, meaning, or content. And for Lewis, just as for Davidson, the constraints or measures determining the ranking are constraints on whole theories (or assignments, interpretations), so we need a way of determining the degree of eligibility a whole theory has on the basis of the eligibility of the individual assignments of content or reference.[35]

[35]When it comes to how precisely to do that, Lewis doesn't say much more than the following:

> Ceteris paribus, an eligible interpretation is one that maximises the eligibility of referents overall

It is, I take it, fairly clear that the account of meaning determination Lewis suggests to save by means of naturalness is not the one he endorses.[36] He does, however, suggest introducing an analogous constraint into his own account of the determination of propositional attitude content (cf. Lewis 1983, 373ff). And, most intriguingly in our present context, he also suggests dealing with "Kripkenstein's puzzle" by means of an appeal to natural properties:

> [W]e must pay to regain our naiveté. Our theory of properties must have adequate resources to somehow ratify the judgement that instances of adding are all alike in a way that instances of quadding are not. The property of adding is not perfectly natural, of course, not on a part with unit charge or sphericality. And the property of quadding is not perfectly unnatural. But quadding is worse by a disjunction (Lewis 1983, 376).

Thus it is fairly clear that Lewis considers constraints using the naturalness of properties perfectly fine to be used in accounts of meaning or content determination if needed. My question is whether a Davidsonian account of meaning determination can use naturalness to restrict the properties eligible to be meant. More precisely, my question is whether such an account can use naturalness to restrict the eligible properties in a way that does not essentially involve the interpreter.

I have investigated this question at greater length elsewhere (cf. Glüer 2016). Here, I shall only give the bare bones of that discussion. The crucial point is the following: In the context of Davidsonian meaning determination, constraints need to be if not motivated by, then at least compatible with, the public nature of meaning. Restricting the eligible meanings now appears to be motivated in precisely this way: What Casey teaches us is that for meaning to be public, it is not enough that the determination base be a certain way. It is also required that only certain meaning assignments are eligible. The question is whether restricting what can be meant to the *natural properties* can be motivated in this way – while keeping the interpreter out of the picture, so to speak.

(Lewis 1984, 227).

Nothing here will hang on the details, but those interested in how this might work can look at Williams 2007. He suggests a method for determining eligibility values for whole interpretations, and provides discussion.

[36] For relevant discussion, see Schwarz 2014; Weatherson 2013.

I have argued that this is not the case. The attempt to do so faces a dilemma. What we are trying to do is to keep the interpreter out of determining the eligible properties while at the same time securing detectable similarity, i.e. making sure that the interpreter is sensitive to the eligible properties. As far as I can see, there are three ways in which Lewis characterized the natural properties. On two of them, detectable similarity can be secured, but the interpreter is very much involved in characterizing the properties as natural. On the third, the interpreter is out of it – but so is the guarantee for detectable similarity. Thus, it seems that it is either-or: We can either keep the interpreter out or secure detectable similarity. We can either have objectivity, that is, or sensitivity. This seems to be an instance of a more general dilemma. Let's call it the "dilemma of objectivity and sensitivity".[37]

As I said, Lewis makes use of three ways of characterizing properties as more or less natural. One is by providing examples:

> The mereological sum of the coffee in my cup, the ink in this sentence, a nearby sparrow, and my left shoe (...) is an eligible referent, but less eligible than some others. (I have just referred to it.) Likewise the metal things are less of an elite, eligible class than the silver things, and the green things are worse, and the grue things are worse still – but all these classes belong to the elite compared to the counted utterly miscellaneous classes of things that there are (Lewis 1984, 227).

A second way of characterizing the distinction proceeds by reflecting on what is is more or less rational to believe and desire (cf. e.g. Lewis 1983, 375). And third, there are passages where Lewis suggests that there are "perfectly natural properties" which it is up to physics to discover. Examples he gives are mass, charge, and quark colour and flavour (cf. Lewis 1984, 228). Other properties can then be defined in terms of the fundamental physical properties. And degrees of naturalness can be characterized as follows: the longer the chain of definability between a property and the perfectly natural properties, the less natural it is.

It should be clear that the first two methods of characterizing naturalness rely on ourselves and our reactions as a kind of black box. So understood, the natural properties would indeed seem to be properties an interpreter would be sensitive to, but by the same token our understanding of naturalness remains dependent on precisely those sensitivities. Not so on

[37] In my (2016), I construe the verdict regarding Lewis as an instance of what I there call " the dilemma of purity and sensitivity".

Lewis's third way: No reference to ourselves and our sensitivities appears to be involved in characterizing naturalness in terms of the length of chains of definitions. But by the same token, there is no longer any guarantee for these properties' being interpreter-sensible. Once we identify the perfectly natural properties as the fundamental physical properties, sensitivity to naturalness becomes a metaphysical coincidence.

The dilemma we face when trying to restrict eligibility in terms of naturalness can plausibly be expected to generalize: Whatever makes for eligibility needs to be such that detectable similarity does not become mere coincidence. This would seem to require characterizing eligibility in a way that essentially relates it to an interpreter's sensitivities. And this, in turn would mean that the problem of objectivity is with us to stay. It was precisely its epistemico-metaphysical nature that seemed to provide the radical interpretation account of meaning determination with the additional resources required for answering the rule-following considerations. And indeed, the account does have resources towards justifying choices of determination base and principle, choices that otherwise might seem totally arbitrary. Nevertheless, it is its very epistemico-metaphysical double-nature that ultimately prevents the account from escaping the problem of objectivity. Interestingly, it is not primarily the determination base – the form of "use" the account works with – that turns out to be problematic. Rather, objectivity remains elusive because publicness not only requires the data to be public. Publicness also turns out to require restrictions on what is eligible to be meant. Ironically, it turns out that we get "objective" meanings only if we restrict the realm of what can be meant to the properties we are sensitive to. Unless we find a way of preventing the dilemma of objectivity and sensitivity from generalizing, the problem of objectivity therefore cannot be solved by means of charity.[38]

Once we adopt the Davidsonian epistemico-metaphysical take on meaning, the prob-

[38]While this means that charity ultimately does not answer the rule-following considerations, it doesn't have to mean that a radical interpretation account of meaning determination is doomed. This depends on what precisely we expect such an account to be like.

Importantly, this result also does not mean that the eligible properties have to be construed as relational properties, for instance as dispositions to elicit certain reactions in (suitable) interpreters (under certain circumstances). It does mean that we *cannot* just identify them with the categorical bases of these dispositions. But we *can* instead construe them as *the properties that make objects disposed to elicit those responses* – where what that description refers to, or is satisfied by, varies across possible worlds, depending on which property is the categorical base of the relevant disposition in the world in question. For details, see Glüer 2016. There is consequently no sense in which meaning – whether in the sense of 'having a meaning' or in the sense of 'being a property meant' – has to be understood as response-dependent (in Wright's sense; cf. Wright 1988, see also Johnston 1992) on a Davidsonian account. See above, fn. 14.

lems generated by the rule-following considerations look somewhat different from what we have become used to. The main concern is meaning determination, but what appears threatened is not primarily the "naturalist" or "reductivist" character of the base – what looms large is rather the problem of objectivity. What the rule-following considerations teach us is precisely to appreciate the metaphysical consequences of meaning's peculiarly epistemic nature.[39] We might need to learn to live with meanings not being fully and totally objective – in the "platonist" sense the rule-following considerations are premised upon.[40] Wittgenstein himself, of course, went further than anything suggested here – he seems to have concluded that meaning requires the "common behavior of mankind" as a frame of reference (PI 206), quipping

> If a lion could talk, we could not understand him (PI II, X, 223).

Nothing I have argued here forces so drastic – or so paradoxical – a conclusion. What I have argued, however, is that a radical interpretation account of meaning determination does require a background of shared sensitivity, of commonly detectable similarity.[41]

References

Boghossian, Paul (1989). "The Rule-Following Considerations". In: *Mind* 98, pp. 507–549.

Byrne, Alex (1998). "Interpretivism". In: *European Review of Philosophy* 3, pp. 199–223.

Davidson, Donald (1967). "Truth and Meaning". In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 17–36.

— (1973). "Radical Interpretation". In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 125–139.

— (1974). "Belief and the Basis of Meaning". In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 141–154.

— (1975). "Thought and talk". In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 155–170.

[39] As Åsa Wikforss argues in her contribution to this volume, this is precisely one of the fault lines between the early and the later Wittgenstein; cf. Wikforss 2016, ?.

[40] Cf. Some commentators take Wittgenstein's main point to be the more radical one that the very "idea of determination is incoherent" (Goldfarb 2012, 78) and that, thus, no account of meaning whatsoever is possible. I am with Verheggen 2003 here in thinking that Wittgenstein rather points to the need to gain a better understanding of what kind of account of meaning is possible.

[41] For most helpful discussion and comments, I would like to thank Claudine Verheggen, Peter Pagin, Åsa Wikforss, Bill Child, and audiences in St. Andrews, Stockholm, Barcelona, and Toronto.

Davidson, Donald (1976). “Reply to Foster”. In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 171–179.

— (1977). “The Method of Truth in Metaphysics”. In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 199–214.

— (1980). “A Unified Theory of Thought, Meaning and Action”. In: *Problems of Rationality*. Oxford: Clarendon Press 2004, pp. 151–166.

— (1982). “Communication and Convention”. In: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 265–280.

— (1983). “A Coherence Theory of Truth and Knowledge”. In: *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001, pp. 137–153.

— (1984). *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

— (1986). “A Nice Derangement of Epitaphs”. In: *Truth, Language and History*. Oxford: Clarendon Press 2005, pp. 89–108.

— (1990a). “Epistemology Externalized”. In: *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001, pp. 193–204.

— (1990b). “Meaning, Truth, and Evidence”. In: *Truth, Language and History*. Oxford: Clarendon Press 2005, pp. 47–62.

— (1991). “Three Varieties of Knowledge”. In: *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001, pp. 205–220.

— (1992). “The Second Person”. In: *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press 2001, pp. 107–121.

— (2005). *Truth and Predication*. Cambridge, MA: The Belknap Press of Harvard University Press.

Glüer, Kathrin (2000). “Wittgenstein and Davidson on Agreement in Judgment”. In: *Wittgenstein Studies* 2, pp. 81–103.

— (2001). “Alter Hut kleidet gut. Zur Verteidigung des semantischen Holismus”. In: *Holismus in der Philosophie*. Ed. by M. Seel; J. Liptow; G. Bertram. Velbrück Wissenschaft Verlag, pp. 114–126.

— (2006a). “The Status of Charity I: Conceptual Truth or Aposteriori Necessity?” In: *International Journal of Philosophical Studies* 14, pp. 337–359.

Glüer, Kathrin (2006b). “Triangulation”. In: *The Oxford Handbook of Philosophy of Language*. Ed. by Ernest Lepore and Barry Smith. Oxford: Oxford University Press, pp. 1006–1019.

— (2011). *Donald Davidson. A Short Introduction*. New York: Oxford University Press.

— (2016). “Interpretation and the Interpreter. On the Role of the Interpreter in Davidsonian Foundational Semantics”. In: *The Science of Meaning*. Ed. by Brian Rabern; Derek Ball. Oxford: Oxford University Press (forthcoming).

Glüer, Kathrin and Peter Pagin (2003). “Meaning Theory and Autistic Speakers”. In: *Mind and Language* 18, pp. 23–51.

Glüer, Kathrin and Åsa Wikforss (2010). “Es braucht die Regel nicht. Wittgenstein on Rules and Meaning”. In: *The Later Wittgenstein on Meaning*. Ed. by David Whiting. Basingstoke: Palgrave Macmillan, pp. 148–166.

Goldfarb, Warren (2012). “Rule-Following Revisited”. In: *Wittgenstein and the Philosophy of Mind*. Ed. by Johathan Ellis; Daniel Guevara. Oxford: Oxford University Press, pp. 73–89.

Gross, Steven (2015). “The Metaphysics of Meaning: Hopkins on Wittgenstein”. In: *International Journal of Philosophical Studies* 23, pp. 518–538.

Guardo, Andrea (2010). “Kripke’s Account of the Rule-Following Considerations”. In: *European Journal of Philosophy* 20, pp. 366–388.

Harman, Gilbert (2011). “Davidson’s Contribution to the Philosophy of Language”. In: *Davidson’s Philosophy. A Reappraisal*. Ed. by Gerhard Preyer. Oxford: Oxford University Press (forthcoming).

Hopkins, Jim (1999). “Wittgenstein, Davidson, and Radical Interpretation”. In: *The Philosophy of Donald Davidson*. Ed. by Lewis Edwin Hahn. Chicago and La Salle, Ill.: Open Court, pp. 255–285.

— (2012). “Rules, Privacy, and Physicalism”. In: *Wittgenstein and the Philosophy of Mind*. Ed. by Johathan Ellis; Daniel Guevara. Oxford: Oxford University Press, pp. 107–144.

Jackman, Henry (2003). “Foundationalism, Coherentism and Rule-Following Scepticism”. In: *International Journal of Philosophical Studies* 11, pp. 25–41.

Johnston, Mark (1992). “How to Speak of the Colors”. In: *Philosophical Studies* 68, pp. 221–263.

Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.

Kusch, Martin (2006). *A Sceptical Guide to Meaning and Rules. Defending Kripke's Wittgenstein*. Chesham: Acumen.

Lepore, Ernest and Kirk Ludwig (2005). *Donald Davidson. Meaning, Truth, Language, and Reality*. Oxford: Oxford University Press.

Lewis, David (1974). "Radical Interpretation". In: *Synthese* 27, pp. 331–344.

— (1983). "New Work for a Theory of Universals". In: *Australasian Journal of Philosophy* 61, pp. 343–377.

— (1984). "Putnam's Paradox". In: *Australasian Journal of Philosophy* 62, pp. 221–236.

Ludwig, Kirk (2014). "Review of: Kathrin Glüer, Donald Davidson: A Short Introduction". In: *Dialectica* 68, pp. 464–473.

Merino-Rajme, Carla (2015). "Why Lewis' Appeal to Natural Properties fails to Kripke's Rule-Following Paradox". In: *Philosophical Studies* 172, pp. 163–175.

Pagin, Peter (1999). "Radical Interpretation and Compositional Structure". In: *Donald Davidson: Truth, Meaning and Knowledge*. Ed. by U. Zeglen. Routledge, pp. 59–71.

— (2001). "Semantic Triangulation". In: *Interpreting Davidson*. Ed. by Peter Kotatko; Peter Pagin; Gabriel Segal. Stanford: CSLI, pp. 199–212.

— (2002). "Rule-Following, Compositionality and the Normativity of Meaning". In: *Meaning and Interpretation*. Ed. by Dag Prawitz. Stockholm: Kungliga Vitterhets Historie och Antikvitetsakademien, pp. 151–181.

— (2006). "The Status of Charity II. Charity, probability, and simplicity". In: *International Journal of Philosophical Studies* 14, pp. 361–383.

— (2013). "Radical Interpretation and the Principle of Charity". In: *A Companion to Donald Davidson*. Ed. by Ernest Lepore; Kirk Ludwig. Oxford: Wiley-Blackwell, pp. 225–246.

Quine, Willard Van Orman (1960). *Word and Object*. Cambridge, Mass.: MIT Press.

— (1995). *From Stimulus to Science*. Cambridge, MA: Harvard University Press.

— (2000). "I, You, and It: An Epistemological Triangle". In: *Knowledge, Language and Logic: Questions for Quine*. Ed. by Alex Orenstein; Petr Kotatko. Dordrecht: Kluwer Academic Publishers, pp. 1–6.

Schwarz, Wolfgang (2014). "Against Magnetism". In: *Australasian Journal of Philosophy* 92, pp. 17–36.

Verheggen, Claudine (1995). “Wittgenstein and ‘Solitary’ Languages”. In: *Philosophical Investigations* 18, pp. 329–347.

— (2000). “The Meaningfulness of Meaning Questions”. In: *Synthese* 123, pp. 195–216.

— (2003). “Wittgenstein’s Rule-Following Paradox and the Objectivity of Meaning”. In: *Philosophical Investigations* 26, pp. 285–310.

— (2006). “How Social Must Language Be?” In: *Journal for the Theory of Social Behavior* 36, pp. 203–219.

— (2007). “Triangulating with Davidson”. In: *Philosophical Quarterly* 57, pp. 96–103.

— (2013). “Triangulation”. In: *A Companion to Donald Davidson*. Ed. by Ernest Lepore; Kirk Ludwig. Oxford: Wiley-Blackwell, pp. 456–471.

Weatherson, Brian (2013). “The Role of Naturalness in Lewis’s Theory of Meaning”. In: *Journal for the History of Analytic Philosophy* 1, pp. 1–18.

Wikforss, Åsa (2016). “Davidson and Wittgenstein: A Homeric Struggle?” In: *Wittgenstein and Davidson on Thought, Language, and Action*. Ed. by Claudine Verheggen. New York: Cambridge University Press.

Williams, J. Robert G. (2007). “Eligibility and Inscrutability”. In: *The Philosophical Review* 116, pp. 361–399.

Williamson, Timothy (2004). “Philosophical ‘Intuitions’ and Scepticism about Judgement”. In: *Dialectica* 58, pp. 109–153.

Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Ed. by G. E. M. Anscombe; R. Rhees. New York: Macmillan.

— (1956). *Remarks on the Foundations of Mathematics*. Ed. by G. E. M. Anscombe, Rush Rhees, and G. H. von Wright. Basil Blackwell.

— (1969). *On Certainty*. Ed. by G. E. M. Anscombe; G. H. von Wright.Translated by Denis Paul and G. E. M. Anscombe. Oxford: Basil Blackwell.

Wright, Crispin (1980). *Wittgenstein on the Foundations of Mathematics*. London: Duckworth.

— (1988). “Realism, Antirealism, Irrealism, Quasi-Realism”. In: *Midwest Studies in Philosophy* 12, pp. 25–49.

Wright, Crispin (1989). “Wittgenstein’s Rule-Following Considerations and the Central Project of Theoretical Linguistics”. In: *Reflections on Chomsky*. Ed. by A. George. Oxford: Basil Blackwell.

— (2007). “Rule-Following without Reasons: Wittgenstein’s Quietism and the Constitutive Question”. In: *Ratio* XX, pp. 481–502.

— (2012). “Replies Part I: The Rule-Following Considerations and the Normativity of Meaning”. In: *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*. Ed. by A. Coliva. Oxford: Oxford University Press, pp. 379–401.