

3. The Status of the Principle of Charity

The principle of charity has always occupied a key role in Davidson's theory of interpretation. Roughly, the principle states that in order to understand another being, one has to assume that the other being is in accordance with himself, and that he speaks the truth. In this way, the principle describes a condition for understanding other linguistic beings. Davidson's theory is not yet a theorization of actual everyday communication. His theory operates on an assumed underlying level, and the theory takes a thought experiment as its vantage point. The thought experiment is named "radical interpretation", and describes an interpretational situation in which an interpreter is trying to understand another being with an unknown culture and an unknown language, that is, unknown to the interpreter. Radical interpretation involves a basic position, in which no linguistic behavior is ascribed a specific meaning in advance. The key to understanding of the other being is to be found in a connection between truth and meaning. An assumption is that the interpreter is able to recognize the speaker's attitude of "holding true", and the interpreter reaches the meaning of the utterances on the basis of the truth conditions of the utterances. It is initially in this imagined situation that the principle plays its role as a possibility condition for understanding. In this case, the principle is a requirement for successful communication and even for the very possibility of getting communication started. Davidson anticipates that the principle is inevitable.

In this article I discuss ascription of epistemic status to the principle of charity.¹ I give an account of Davidson's version of the principle, partly by comparing it to Quine's version, and, in particular, I stress the function that the principle of charity had in the first place in Davidson's theory.

As radical interpretation is an imagined case, and in one sense, an idealized situation, I will analyze it further to get a hold of the possibilities and restrictions that the situation prescribes to the principle. I will involve Quine and Davidson's understanding of classical epistemic distinctions, and consider how this understanding influences ascription of epistemic status to the principle. I argue that our present epistemic categories are unfit for

¹ The epistemological turn that Davidson's philosophy has taken with the introduction of the concept of triangulation has actualized the question of the epistemic status of the principle.

the principle of charity; yet, at the same time, that an understanding of the function of the principle is epistemically enlightening nonetheless. Moreover, I show that the epistemic potential of the principle in Davidson's theory is elucidated by the introduction of the concept of triangulation.

1. Davidson's principle of charity

The basis for Davidson's principle is to be found in Quine's *Word and Object*, where Quine introduced the principle as a rule of "radical translation". Quine's principle states that obviously false and absurd utterances from the person you are trying to understand are more likely to be due to poor translation than that he or she actually holds the view that your translation attributes to him. He says: "The maxim of translation ... is that assertions startlingly false on the face of them are likely to turn on hidden differences of language. ... The common sense behind the maxim is that one's interlocutor's silliness, beyond a certain point, is less likely than bad translation—or, in the domestic case, linguistic divergence."² If we hear the other person say something that seems founded on a very unreasonable belief, we must, according to Quine, assume that there could be something in the other person's language that we have misunderstood. In a related note he continues: "Cf. Wilson's principle of charity: "We select as designatum that individual which will make the largest possible number of ... statements true" ().³ According to Wilson, the point is to choose as starting point the individual [the translation] that makes the largest possible number of utterances true⁴, and we must assume that Quine counts this as a point in favor of his own characteristic of the principle (that utterances which appear untrue are likely to originate in our deficient understanding of the language).⁵ If we are to believe that our informant truly means what he says, according to our manual, this may compromise many established translations and even make obsolete many of the reasonable beliefs we had

² In Willard Van Orman Quine: *Word and Object*, The MIT Press, Cambridge, Mass. 1960, p. 59.

³ *Ibid.*

⁴ Wilson's example anticipates that we have five possible "versions" or alternatives to deal with, and when four out of these alternatives points towards the same "individual", while one alternative points towards another individual, we will assume that the individual pointed to by the four concurrent versions is the right one. In Wilson's example we thus deal with a predominantly quantitative assessment. See page 532 in N. L. Wilson: "Substances without Substrata", in *The Review of Metaphysics*, Vol. 12, nr 4, 1959, p 521-539.

⁵ We can add that when we run into absurd utterances of the informant, these utterances can be absurd in at the least two different ways; they are absurd because they are so in the light of our own beliefs, or they are absurd on the basis of the beliefs we until now have ascribed the informant, which in its turn *also* is based upon our system of beliefs.

previously considered true beliefs of the informant. Therefore, we will not pick the translation of the utterance that makes a large number of the other utterances false. Instead, we will assume that we have bumped into a previously hidden difference between the language of the informant and our own language. Ten pages on in the book, Quine approaches another methodic point by saying: "...bold departure, [is] to be adopted only if its avoidance would seem to call for much more complicated analytical hypotheses. For certainly, the more absurd or exotic the beliefs imputed to a people, the more suspicious we are entitled to be of the translations; the myth of the prelogical people marks only the extreme. For translation theory, banal messages are the breath of life."⁶

Having simple analytical hypotheses is something to reach for, and ascription of ordinary, or even banal, beliefs to our informant, is according to Quine "the breath of life" for translation theory. Simplicity is Quine's only mentioned reason for putting aside the principle of charity. If we would have had to operate with a much more complicated hypothesis, the cost of employing the principle of charity could turn out too high.

Summarized, the function of the principle in Quine's theory is to refuse a translation that ascribes exotic beliefs to the informant. This means that the assumed meaning of an utterance is being weighed against the belief that must be ascribed the speaker, when we take this particular meaning to be a correct translation. This is Davidson's starting point for bringing the principle into the theory of radical interpretation. In radical translation, or, to be precise, radical interpretation, Davidson starts off with the operation of weighing meaning and belief. In order to be able to understand someone else, one has to understand which beliefs he or she has, as well as what his or her utterances mean. In "Radical Interpretation" Davidson says that it all turns on "...holding beliefs constant as far as possible while solving for meaning"⁷, and the principle of charity has the function of holding on to the belief.

The problem of finding a way of weighing belief and meaning against each other is a reminder of the problem Davidson had been involved in earlier when trying as he tried to find a way of separating "utility" and "subjective probability" in experimentally oriented

⁶ Willard Van Orman Quine: *Word and Object*, The MIT Press, Cambridge, Mass. 1960, p. 69, my addition.

⁷ "Radical Interpretation" (1973), in Donald Davidson: *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984, p. 137.

decision theory.⁸ The problem in question was how to find an empirical basis for identifying and separating the two entities, and how a theory could contribute in the separate measuring of subjective probability and expected utility. When an individual chooses between alternatives involving uncertain outcomes, at least two factors arise in the decision: The relative desirability of the possible outcomes, and the extent to which the different outcomes are considered probable. Generally, one would assume that a rational individual chooses, among high-risk alternatives, the alternative that maximizes expected utility.⁹

Davidson's inspiration came from the strategy that Frank Ramsey had chosen when trying to figure out the concept of subjective probability.¹⁰ Ramsey had noticed a problem with subjective probability, or what we might refer to as "the degree of belief in beliefs", the problem that others cannot observe it directly. Unquantified preference can, however, be read of in behavior in concrete situations of choice, which can be partly observed. Behavior in situations of choice gives empirical evidence, and, based on this, Ramsey conceives that a theory of subjective probability can be given empirical content. He imagines setting up a sequence of options, which, when we have the theory, may give the information about the degree of belief in beliefs and the relative strength of desired preferences. 'The degree of belief in beliefs' and 'the relative strength of preferences' will thus be theoretical concepts that will have their meaning through their role in explaining and predicting decisions in situations of choice.

Having this in mind, Davidson construes a utility function based on a random event with a subjective probability of one half, which can be utilized to define the utility of the outcomes, and this utility function can subsequently be used in the measurement of

⁸ First published as Donald Davidson and Patrick Suppes: "A Finitistic Axiomatization of Subjective Probability and Utility" in *Econometrica*, Vol. 24, 1956, p. 264-75. "Utility" or "expected utility" will here involve the question of to what extent one will value that a certain incident will occur (corresponds to "desire"/wish), while "subjective probability" will involve the question of how likely one estimates it to be that this incident will occur (corresponds to "belief"/view). Davidson has often called attention to this resemblance between theory of translation/interpretation and decision theory.

⁹ See p. 1-3 in Donald Davidson, Patrick Suppes and Sidney Siegel: *Decision Making, an Experimental Approach*, Stanford University Press, Stanford 1957.

¹⁰ See p. 84 in Davidson's "Reply to Jerry Fodor and Ernest LePore" in Stoecker, R. (ed.): *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers*, De Gruyter, Berlin 1993, pp 77-84. Davidson did however find the utility function later to be mentioned on his own, and only later did he discover that Ramsey had found the same function. For an example, see p. 32 of his "Intellectual Autobiography" in Lewis Edwin Hahn (Red.): *The Philosophy of Donald Davidson*, in the series The Library of Living Philosophers, Open Court, Chicago and La Salle 1999.

subjective probability. This means that by fixing the subjective probability at one half, one can calculate the relative utility of the different outcomes, and the function thereby construed can again be employed to measure subjective probability. Hence, it was now possible to “determine a person’s subjective probabilities and relative degrees of preference simultaneously, assuming neither in advance.”¹¹

The similarity between theory of interpretation and decision theory is this difficult situation where we are facing two factors that are both obviously important in a situation where we have to decide what to choose, but where none of the two are more basic than the other, and where the value of both factors mutually influence each other. Thus, it is necessary to weigh up the two factors, and it is all about finding the right place (or reference mark) on the lever by using one of the factors to “measure” the other. Hence, one of the factors is given a specific value; in decision theory the utility function is given a subjective probability of one half, thereby deciding the utility, and in theory of interpretation the beliefs of the other person is given according to my beliefs and according to the other person’s other beliefs. This fictitiously specific reference mark (a specific utility value/a specific belief) is thereafter employed to, respectively, measure subjective probability and to determine the meaning of an utterance. By inserting a specific hypothetical value for one of the factors, a specific value for the other factor appears.¹²

But what is really achieved by this move? In the case of radical interpretation, in which Davidson has sought a way to separate meaning and belief, the objective has been to gain

¹¹ Ibid.

¹² Davidson made it clear that he brought this solution along from decision theory to radical interpretation, and the solution came to be the beginning of a unified theory on language and action. Davidson originally took as starting point the assumption that the interpreter must be able to recognize when the speaker holds something true. This is being described in “Radical Interpretation” from 1973, but already in “Belief and the Basis of Meaning” (see pages 147-48) from 1974 (both in Donald Davidson: *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984) he notices that it would have been better if we could employ a speaker’s preferences between sentences as a starting point for interpretation. That the speaker holds something true (“holding true”) were in an expanded theory substituted by that the speaker prefers the truth (“preferring true”) of one sentence to the truth of another sentence. See also “Towards a Unified Theory of Meaning and Action” in *Grazer Philosophische Studien*, 11, 1980. This substitution lead to the unification of theory of meaning and decision theory (or, more generally, theory of action) and thus a theory of interpretation, where all forms of human intentional action can be handled. A comprehensive account of this unified theory is given in “The Structure and Content of Truth” from 1990, in *The Journal of Philosophy*, Vol 87, p. 279-328. The extended theory is also radical, since it has the same kind of ignorance constraint, and the principle of charity is being used unlimited. In this article I nevertheless maintain the focus on linguistic behavior and the role of the principle of charity in this context, as the unified theory, as far as I can see, does not bring about a changed epistemic status for the principle of charity.

the ability to explain what is involved when an utterance means what it does, or in other words, what is the significance of saying that the utterance has a meaning. Davidson says that he initially wanted an answer to the question of what meaning is, and that the radical interpreter thought experiment was motivated by the question of what linguistic meaning is, and what it can be when we do *not* base it on a metaphysically heavy loaded concept of meaning.¹³ But, according to Davidson, the question of what meaning is cannot be given an answer; we don't even know where to begin, or what would count as an answer. Instead, Davidson poses the following question: "What would it suffice an interpreter to know in order to understand the speaker of an alien language, and how could he come to know it?"¹⁴ We are dealing with the unknown speaker and his physical environment. We don't know anything about the speaker's character; we don't know the culture in which he takes part, and we have no preconception of the meanings of words and expressions in his language. The radical interpreter enters the situation with an aim of getting to grips with the beliefs of the speaker, and the hypotheses that he forms will not be filled with a specific understanding of the meanings of the speaker's utterances. In radical interpretation, we want evidence that will *suffice* for the interpreter to be able to form a substantiated theory about the beliefs and meanings of a speaker. The basis for this is that the interpreter sometimes knows when the speaker affirms a given sentence, even if the interpreter doesn't understand the sentence. This is a decisive premise for the possibility of radical interpretation, where interpretation as the possibility of recognizing that somebody holds something true enables our entry into the mutual dependence between belief and meaning. The beliefs of the speaker, which the interpreter partly copies from his own system of beliefs, function as the key to finding the meanings of the speaker's utterances. Partly, the aim is to acquire a theory of meaning, a theory about the meanings of a certain speaker's sentences in a limited period of time. The principle of charity, which, in this case, performed the function of enabling the interpreter to hold onto (what he sees as) the speaker's reasonable beliefs, enabled the interpreter to reach the meanings in a legitimate way.¹⁵ By assuming the rationality of the speaker when an alternative interpretation would

¹³ In "Reply to Jerry Fodor and Ernest LePore", p. 83 in Stoecker, R. (ed.): *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers*, Berlin 1993 we find Davidson saying: "Like many others, I wanted answers to such questions as "What is meaning?", and became frustrated by the fatuity of the attempts at answers I found...."

¹⁴ *Ibid*, p. 83.

¹⁵ Not knowing the speaker and his language is a condition that is supposed to "prevent smuggling in into the foundations of the theory concepts too closely allied to the concept of meaning". (From the introduction to his *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984, p. xiii) The legitimacy of

involve ascription of obviously false assertions to him, (all of course according to the interpreters own standards of rationality and truth) the interpreter has found the meaning of the speaker's utterances. This means that the use of the principle of charity involves ascription of belief and meaning based on unusually low evidential coverage. We cannot know for sure whether or not the speaker is, in fact, telling the truth, or whether or not he is in accordance with himself, and therefore, we cannot be certain that the meaning that we ascribe to the utterances is the meaning of the utterances at the time. By further re-examining the theory, we can obviously test the results at a later time, and thereby increase evidential support. But we have brought forward an initial hypothesis about the meaning of the utterance, which is basis for proceeding.¹⁶

Which role and status should now be attributed to the fictitious specific reference mark? It is important to stress that the beliefs that we, hypothetically, have been holding on to, are precisely temporary hypotheses about the speaker's beliefs; they are a means to reach the meanings of the utterances. For this reason, the result must be temporary as well. Suggestively, Davidson describes this as "passing theory"¹⁷. Yet again, through our estimation of them as such by our usage of the principle of charity, the beliefs are neither true nor consistent. We have utilized the principle in order to choose the interpretation that we consider most reasonable, and by doing this, we assume that we have found the best interpretation. There is nothing in our employment of the principle that *give* beliefs a permanent or metaphysic status of truth, and this applies to both speakers' and interpreters' beliefs.

For Davidson, as pointed out earlier, the principle of charity was initially some sort of lever for "measuring" or uncovering the meanings of the speaker's utterances in radical interpretation. In interpretation, employment of the principle results in a theory of meaning and a theory of belief in a single move. In Quine's version, the principle functions in a

the theory of meaning is secured by not presupposing a specific concept of meaning, and by letting the radical interpretation rest heavily on empirically occurring interaction between speaker, surroundings and an, as far as possible, unbiased interpreter.

¹⁶ The truth conditions for the utterances of the speaker and that the speaker is holding sentences true, is yet not the sole support for the interpreter. Davidson is pointing out that the interpreter also must consider "...simplicity, hunches about the effects of social conditioning, and of course our common-sense, or scientific, knowledge of explicable error." ("On the Very Idea of a Conceptual Scheme" (1974) p. 196 in Davidson: *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984.)

¹⁷ See p. 442 in "A Nice Derangement of Epitaphs" in E. LePore: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford 1986.

similar way as the rule of aspiration to simplicity, and it does not have a special status in that respect. Having this in mind, it is easy to see that the principle of charity has another role to play in Davidson's theory. The plain naturalist will not regard the principle any differently from Quine, and it is therefore important to take a look at how Davidson's use of the principle is different from Quine's. To Davidson, the principle is necessary for accessing the system of beliefs of the speaker, in order to make out what he means by what he says, and for finding out which beliefs he has. In the very beginning of radical interpretation, the principle cannot be put aside to the benefit of considerations of simplicity. The principle enables us to pick a first hypothesis and to choose between different alternatives of interpretation. In Davidson's version, the principle is also constitutive in the sense that it contributes to giving meaning to concepts such as belief, meaning, wish, etc. In this sense, it is internally constitutive in the theory of interpretation. This is so because we cannot make out what the meaning of the speaker's utterance or his beliefs are unless we can utilize the principle, and because talking about meaning and belief is meaningful only when we interpret a speaker in relation to his surroundings. Quine's version of the principle was never constitutive in character, and it is reasonably clear that Davidson commits himself more in his usage of the principle than did Quine. In Quine's theory, the principle is a methodological guide that, level with the demand for simplicity, regulates the process of translation.

Furthermore, it is important to notice that Davidson's focus on truth in the process of interpretation has an effect on the function of the principle of charity. For Davidson the principle is to ensure truth and consistency in a first approach to radical interpretation since, without a critical mass of shared belief, understanding cannot be attained. In Quine's theory, the principle did of course not have this function; one reason being that Quine's radical translation lacked an objective of obtained understanding.

2. Radical Interpretation: "Ideal" or "Normal"?

How are we then to understand Davidson's concept of radical interpretation? As mentioned, it is a thought experiment not intended to theorize on actual interpretation.¹⁸ Is

¹⁸ Quine is talking about it as a thought experiment on p. 28 in his *Pursuit of Truth*, (Harvard University Press, Cambridge, Mass. 1990) while Davidson speaks about radical interpretation as a "conceptual exercise"

it hence meant to be a description of the ideal interpretational situation? Davidson has explicitly denied that the radical interpreter is an ideal interpreter.¹⁹ According to Davidson, the theory of radical interpretation and the radical interpretation situation is not meant to be “ideal” but “normal” (representing a normal rather than an ideal situation). Radical interpretation has recently been referred to as “ideal interpretation”; the radical interpreter called an “idealized interpreter” and the speaker said to possess the greatest extent of “cognitive and moral perfection”.²⁰ It is clear that this might make a difference for the epistemic status of the principle of charity. The concept of ideality is not leading us in any specific direction, and, as opposed to the concept of normality, doesn’t anchor the principle in (human) nature. Ideality of communication lead our thoughts towards transcendental pragmatics, and it can lead us into thinking Davidson’s project into such a frame. The concept of “normal” can lead us in the direction of Darwin, in the direction of human dispositions, and into thinking that humans are linguistic beings in their biological constitution. Facing these two alternatives, it undoubtedly makes more sense to associate a tradition of evolutionary theory, or naturalism, to Davidson’s philosophy.

The concept of ideality can, however, be understood in two different ways. It can be perceived as “prefigurative”, hence as an exemplary, or it could on the other hand be understood as “conventionalized”, “simplified”, as a model or a typical form. Davidson has disavowed the former. However, he does agree that radical interpretation is a theoretical situation, and a fictitious situation. The situation can be likened to a scientific experiment, or a model, which is simplified in the sense that potentially interfering elements are left out. The situation is cleansed, made transparent and construed for a specific purpose, and is methodologically ideal. Furthermore, the situation cannot be an actual historic occasion. But the radical interpreter’s set of beliefs is neither more ideal than are yours and mine, nor is the speaker made ideal or perfect by the interpreter. Some commentators have

(page 324-325 in his “The Structure and Content of Truth” in *The Journal of Philosophy*, 87 (6), 1990, pp 279-328) which was not “meant to throw any direct light on how in real life we come to understand each other, nor on how we master our first concepts and our first language.”

¹⁹ Personal comment, Berkeley, May 2000.

²⁰ See page 1 and page 10-16 in Bjørn Ramberg’s “Naturalizing Idealizations: Pragmatism and the Interpretivist Strategy” (at <http://folk.uio.no/bjoerntr/Natideal.html>). Ramberg is making it clear that he means to follow Davidson in his explanation of “the interpretative strategy” as he is laying it out in the article, but that he is not claiming to do Davidson-exegesis. He is putting up something like an interpretational machine, IDA, which is very similar to the radical interpreter, but he points out that he is aware of that this is not Davidson’s radical interpreter. I therefore neither argue against Ramberg’s interpretation as an interpretation of Davidson.

apprehended Davidson's principle of charity as a principle of speaker idealization. There can be a considerable difference between saying that a first hypothesis is that which is constituted by use of the principle of charity, as I have emphasized, and saying that "what people actually believe and mean is constituted by what [] it is ideally rational for them to believe and mean." as Simon Evnine²¹ explains Davidson. I will maintain that this is not about the "ideally rational", but about what a normal, fallible interpreter deems "normal".

A preliminary hypothesis about meaning and belief is what is constituted by the application of the principle of charity in the first place. Evnine's phrasing in the mentioned quote ("what people *actually* believe and mean") might lead us to think that other people/other creatures ontologically or metaphysically constitute the beliefs and meanings occurring in speakers. Either way, this is not the case. The perspective of the radical interpreter is methodological and epistemological in the sense that it points out the possibilities and limitations for ascription of meaning from a third-person perspective. This third-person perspective, which is a fictitious case, determines the possible theories of meaning and belief for a speaker, or maybe even the best possible theory of meaning and interpretation. But an interpreter cannot alone *determine* what a speaker means and believes; this is settled in a larger context, where the speaker's general natural dispositions and his causal and social learning history form part of the picture.

Moreover, Davidson has emphasized that both the interpreter and the speaker are fallible subjects that make mistakes, are wrong and also sometimes contradict themselves, even if mistake and inconsistency cannot be the main rule, because a consistent set of analytical hypotheses could not then be generated. The radical interpretation situation must therefore be intended to consist of a normal interpreter and a normal speaker, where the limits are set by the way in which natural language works and by the psychology of normal, linguistically equipped humans (for instance, normal humans are fallible and not omniscient), but it is a situation in which our overview is better and we are more in control of what is going on than we are in actual interpretation. We make use of the situation to find out and clarify the requirements of understanding another speaker. We want to illuminate the underlying presuppositions for catching someone else's beliefs and the meaning of his utterances.

²¹ Simon Evnine: *Donald Davidson*, Stanford University Press, Stanford 1991, p. 178.

Hence, Davidson's radical interpreter cannot be an ideal interpreter, even if his competence is a model. Just as little is the speaker in radical interpretation an ideal speaker (Likewise, the speaker in radical interpretation is just as less an ideal speaker), and just as little (less) is the principle an ideal that interpreters of all kinds should strive for.²² If we assume that radical interpretation is theoretical, but has a normal interpreter and speaker, we have to assume that the principle of charity applies to normal interpreters and speakers, even if it does not necessarily apply in the same form in actual interpretational situations. The principle is not designed for ideal interpreters and ideal speakers, but is involved in a theoretic model. If, on the other hand, the radical interpretation thought experiment in its entirety is perceived as an ideal in the prefigurative sense, it will not necessarily apply to normal interpreters and speakers. This interpretation would undermine the scope of the principle and would weaken the value of transmitting it to real-life speech and thought.

Davidson has gradually turned radical interpretation from a one-way interpretational situation, in which an interpreter is trying to understand a speaker based on the speaker's utterances as related to the common physical environment, into a situation in which mutual interpretation exists between the two speakers, though still related to their joint surroundings. In triangulation, the new thing is that the speaker and the interpreter have the same status. Both parties to triangulation are speakers as well as interpreters. Moreover, using the triangulation metaphor, Davidson emphasizes that our beliefs are partly caused by casual influence exerted upon us by a common world; in other words, through casual effects exerted on us by empirically occurring objects and events. Furthermore, ostension and the special role of beliefs related to the genesis of sets of belief have been accentuated. We have seen a shift from focusing on internal constraints in interpretation to, instead, emphasizing causality. I will return to the effects of this change as regards the principle of charity.

²² Davidson is utmost clear that the principle is not an ideal that we can choose to follow if we wish; it is according to him not possible to choose not to follow the principle. See e.g. page 221 in his "Mental Events" (1970) in Davidson: *Essays on Actions and Events*, Oxford University Press, Oxford 1980, where he says "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 irrationality."

3. Apriori? Open to empirical revision? Transcendental?

Is the principle of charity empirical or apriori? The Kantian constitutive apriori seems to be a possible apriori category for the principle. In this case, the apriori is what makes experience possible; in other words, it is the condition for experience without being dependent on experience. The Kantian apriori involves that there are elements of knowledge independent of experience yet closely connected to experience. It is characterized by being necessary and non-revisable. Kant's conception is that the apriori involves knowledge, and that which is knowledge must be true, and therefore non-revisable. Can the principle of charity have an apriori status in this sense? The principle functions as a condition for interpretation and is in principle conceived into the special situation that constitutes radical interpretation. The validity of the principle is presupposed in radical interpretation, and without this presupposition no understanding would come to exist. Through the use of the principle we can understand the other person. Hence, the principle is actively involved in forming understanding in radical interpretation; according to Davidson the principle is inevitable for such formation. Consequently, it must be referred to as constitutive pursuant to normal usage, and it is constitutive in a thought experiment. Davidson indicated as early as in 1970 that the principle of charity is constitutive, without going deeper into the details than we have as yet.²³ In "Mental Events", Davidson likens the constitutive with synthetic apriori. In the article he also says that ascription of consistency and coherence is inevitable.

But could the principle really be apriori in a Kantian sense? The thought experiment precedes experience in its entirety, and so the question does not comport well with the case, at least if we consider it epistemically committing. If we were to say that the principle is apriori, it would be natural to think that there must be something in the situation that is not apriori. Is there anything non-apriori in the thought experiment of radical interpretation? Not everything is internally in radical interpretation is in a flux. How normal (not ideal) people function psychologically, what characterizes natural languages – these are empirical presuppositions on which radical interpretation rests. We cannot avoid these assumptions if the thought experiment is to honor the value of transmittance. The thought experiment is therefore not devoid of presuppositions; it depends on a series of

²³ In "Mental Events" (1970), p. 221, and p. 236-237 in "Philosophy as Psychology" (1974), both printed in Davidson: *Essays on Actions and Events*, Oxford University Press, Oxford 1980.

assumptions that we would generally consider anthropological. The participants in radical interpretation are not machines, and they are not ideal, but normal. The events and actions that take place are still not “normal”; it is a fictitious situation, a model that is operated in order to explore and examine certain philosophical ideas.

If we still were to ask whether the principle is apriori in a Kantian sense, we could ascertain that it does make understanding possible, and therefore, it is a condition for the kind of experience we seek when interpreting others. But does the principle itself depend on experience? Could we have been aware of the existence of these conditions if we had never had interpretational experience? It is hard to imagine that we, at first, could have established the thought experiment if we had no interpretational experience. Is then the principle necessary and non-revisable? The principle constitutes a part of a theory of interpretation of human beings. Obviously, as any other theory, this theory could be made obsolete by a new or more accurate theory. But, given this particular theory, we cannot imagine the theory without the principle since the principle takes part in giving meaning to central concepts, such as belief, meaning and truth. Thus, the principle seems to be theoretically necessary, within the theory. But is it practically necessary? Can what we say about the principle be transferred to practical interpretation [and to our general reflection on thought, language and knowledge], or is it possible that our perception of the principle’s necessity is due to our interpretational practice? We are at loss for a good answer.²⁴

In modern usage of the concept of apriori, the elements of necessity and impossibility of revision is not always closely tied to the concept. In a collection of articles by Paul Boghossian and Christopher Peacocke: *New Essays on the A Priori*, published in 2000, the introduction reads: “Being a priori is to be sharply distinguished from being necessary, from being true purely in virtue of meaning, and from being knowable infallibly.”²⁵ And continues: “For all that, it may still seem that a priori propositions cannot be defeated by

²⁴ An indication which gives reason to talk about the principle as necessary is given by Davidson in his “On the Very Idea of a Conceptual Scheme” (1974, printed as pp 183-198 in his *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984) where he says that in order for something to be counted as a language or a conceptual a scheme, it must be interpretable. Analogical to this one might say that for somebody to be counted a speaker, it must be possible to apply the principle of charity to him or her. This implies that the principle is constitutive of being a linguistic being. This is another form of necessity than the necessity involved when we say that the principle is required for getting started on an interpretation of another being.

²⁵ Paul Boghossian and Christopher Peacocke (Red.): *New Essays on the A Priori*, Oxford University Press, Oxford 2000, p. 3.

wholly empirical information; that is, that they may still be experientially indefeasible. It may be natural to wonder: if something is empirically defeasible, how can it be known justificationaly independently of experience?"²⁶ It still seems that the concept of aprioricity is characterized by that the validity of an apriori assertion can be established independently of experience. This seems to be "the smallest common denominator" in a modern usage of the concept. But, according to Boghossian and Peacocke, the apriori also seems immune to plain empirical invalidation. The interesting question is whether or not we consider the apriori empirically revisable. If the principle of charity is non-revisable, this seems to be in accordance with the view that we can know its truth validity independently of experience, but if it *is* revisable, how can we then know its validity independently of experience? In other words, it won't be easy to uphold a concept of the apriori if we say that what is apriori is also revisable. Still, claiming that we can have knowledge of something independently of experience, and yet, that this something is subject to empirically conditioned revision is not incongruent. The authors talk about this form of aprioricity as "non-conclusive a priori justification or entitlement".²⁷ This means that the epistemic basis is not considered sufficient for an adequate or "clean" apriori status, but they maintain that these cases should still be referred to as apriori. This concept of aprioricity is consequently substantially different from Kant's notion of the apriori, as the authors focus on justification, and as they do not assume that aprioricity involves truth.

Boghossian and Peacocke maintain that much work remains before we can fully understand these cases and before we know how they should be evaluated in relation to non-empirically-revisable cases. For our purpose, the relevant question is whether or not the principle of charity can be a case of "non-conclusive a priori entitlement". In the sense that we don't have to wait for an empirical confirmation in order to determine the principle's entitlement, we can know that the principle is valid, and that it must work independently of experience. But is the principle immune to plain empirical invalidation or not? Can we imagine a situation where our experiences of interpretation and communication could show that the principle must be revised on empirical basis? If the answer were yes, we would be able to ascribe the status of "non-conclusive" or incompletely apriori to the principle. A second option would be to call the principle quasi-

²⁶ Ibid, p. 4.

²⁷ Ibid, p. 4f.

apriori, but that would be a less precise term.) Should we decide to categorize the principle in this way, we would really only be saying that it is a principle that we can validate independently of experience, and that it is a principle that can be revised on empirical basis. The first part of the characteristic is straightforward, but as yet, we haven't discussed the question of empirically conditioned revision.

In radical interpretation, as mentioned earlier, something obviously depends upon empirical affairs, and this also concerns the principle of charity. First of all, in radical interpretation there is the premise that the language spoken by the speaker is a natural language, involving certain empirical opportunities and restrictions. We can imagine that if we were wrong about the properties of natural language, we could also be wrong about our assumptions about the behavior of the principle of charity. Moreover, when it comes to the anticipated normal psychological inventories of speakers and interpreters, there are connections between radical interpretation and the empirical. If our conception of how human beings work is psychologically incorrect and it could be shown that we are generally unable to recognize that somebody holds some sentence true, this could facilitate a revision of the principle. If interpreters in general were unable to recognize assent to utterances, it would have been impossible for an interpreter, as the principle assumes, to keep the beliefs of the speaker constant. Or, what if we as interpreters are unable to distinguish between grades of belief in our own belief, and in this way cannot distinguish between more or less central beliefs in our own system of beliefs, and therefore neither can estimate when the beliefs of the speaker are absurd? Or, what if we are merely lacking grades of belief in our own beliefs, to such an extent that we consider all of our own beliefs to be true? This last state is the supposed condition for autistic persons.²⁸ I don't think that these very broad assumptions we maintain concerning the properties of known natural languages and the normal psychological inventory of people could turn out to be to such an extent that empirically based revision of the principle of charity would enforce itself. Neither is it obvious what would be the consequence if anything for the principle if our suppositions about such empirical matters were exposed.

²⁸ Regarding the principle of charity and autistic people, Peter Pagin and Kathrin Glüer argues that people with autism still does not represent empirical counter-evidence. See pp. 44-46 of their "Meaning Theory and Autistic Speakers" in *Mind and Language*, Vol.18 No. 1 February 2003, p. 23-51.

In asking whether the principle can be revised on an empirical basis, it could be that we really meant to ask whether the beliefs we ascribe to the speaker by help of the principle really by help of the principle are by and large true and consistent. We must then assume that there is a way in which we can independently discover the beliefs of other people, and then check whether or not this way is in accordance with the result obtained through employment of the principle of charity. But what does the principle really state? It states that in order to interpret others in a situation of radical interpretation, we must *assume* that speakers more or less tell the truth, and that they are by and large in accordance with themselves. It does not state that most of our own and the other person's beliefs *are* true and consistent.²⁹ Moreover, it is almost impossible to imagine that an independent check could be initiated without employment of the principle.

It is generally a problem to distinguish between the empirical and the apriori in an immanent naturalism of the Davidsonian kind. There could be several reasons for this. Davidson mentions that, in connection with attempts at testing the empirical validity of decision theory, he ran into a variant of this problem: "I became discouraged with attempts to test the empirical validity of decision theory because I could not think of a well-grounded way of separating the empirical and a priori demands of rationality. Too many apparent deviations from what rationality required were explicable as in accord with theory when reinterpreted."³⁰ This seems to be a steadily repeating experience, that when we relate to the intentional behavior of humans, we relate to variables that are almost impossible to empirically fix in unambiguous ways.³¹ The picture that we have of rationality beforehand and our empirical interaction with intentional behavior mix up in such a way that it is hard, and maybe impossible, to know where to draw fundamental

²⁹ Still, Davidson says that the fact that interpretation is possible is due to "the fact that we can dismiss a priori the chance of massive error." ("Thought and Talk" (1975), printed as pp 155-70 in Davidson: *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984, p. 168-9.) We can then, according to Davidson, apriori rule out the possibility of speaker's thorough mistake, and this means that most of the speaker's beliefs must be true. But, it is here not made clear whether this is intended to be a methodological, ontological or metaphysical claim. I have taken the claim to be purely methodological, as a consequence of the method of interpretation from the third person perspective, and my interpretation of Davidson is here that he does not maintain that most of the beliefs therefore actually *are* true. Or rather, I would say that the veridical nature of beliefs couldn't be based on the principle of charity alone.

³⁰ In Davidson: "Reply to J.J.C. Smart" in Lewis Edwin Hahn (Ed.): *The Philosophy of Donald Davidson*, in the series The Library of Living Philosophers, Open Court, Chicago and La Salle 1999, p. 125.

³¹ This is of course built into Davidson's philosophy in a general form via his endorsement of Quine's thesis of "the indeterminacy of translation". See e.g. p. 227 in "The Inscrutability of Reference" (1979), in Davidson: *Inquiries into Truth and Interpretation*, Oxford University Press, Oxford 1984. The indeterminacy is increasing the further away from "observation sentences" we get. (Davidson is not employing this term.)

lines. In practical interpretation, we handle this in a straightforward manner, and for theoretical ends we can avail ourselves of a pragmatically adapted use of a division between the apriori and the empirical. My point is, however, that the distinction between the apriori and the empirical cannot carry a substantial epistemic weight in a holistic theory where such differences are regarded gradual.

If we ask ourselves whether the principle could be a transcendental condition for understanding, we must ask if the principle could appear as a result of a transcendental deduction; or in other words, if it could have its justification and legitimacy via such a deduction. Provided that one finds transcendental deductions to be possible operations, one must then ask whether the principle can be brought forward by such means. I would say that even if the principle is a *condition* for our understanding, this doesn't make the principle a *justified* principle. If a transcendental deduction is characterized by justifying the subject for deduction, the principle probably cannot be deduced transcendently. Hence, it is possible that the principle is a quasi-transcendental condition for understanding, since it on one hand appears to be a possibility condition, but on the other hand that this still isn't enough to justify the principle? As long as we cannot give necessary and sufficient conditions for something to be characterized as quasi-transcendental, such a categorization isn't much of help. A third alternative is to say that the principle is transcendental in a weak sense, owing to the fact that the "transcendental condition" concept is to be understood in the sense of a "possibility condition". But in that case, we would be just as well off by simply employing the term "possibility condition". I cannot see any obvious reasons for saying that the principle of charity is a transcendental principle, nor will we have much use of the term "transcendental" in a philosophical mindset where it doesn't make sense to utilize the concepts of "apriori" and "constitutive" in an epistemic binding fashion.

4. The possible epistemic status of thought experiments.

What status could a principle in a thought experiment such as radical interpretation have? Could a principle in a thought experiment actually be revised? Radical interpretation is not purely theoretical; it is a thought experiment that relies on a long list of empirical assumptions. One could hardly claim to be in control of the individual factors and external

conditions for this type of experiment. Empirical experiments are often used with an aim to control or manipulate the external (or boundary-) conditions, even though we cannot even in this case be certain that all prerequisites are known for us (for instance in the form of theoretical supporting hypotheses). In radical interpretation our overview of premises is hardly better than in experiments in general. Therefore, I will assume that we lack control of the premises and presumptions on which this thought experiment relies, and that the principles we postulate as conditions will therefore not have a different status than other, more or less empirically based premises. Furthermore, to think that we in thought experiments can draw fundamental lines between empirical revision and revision on basis of change in theory, as Robert Sinclair, for instance, seems to assume, can hardly fall into place in a Davidsonian frame.³² By and large, Davidson endorses Quine's holism, and adheres to the Duhem-Quine-thesis. A main point here is that individual assertions do not have their own stock of privileged observational sentences that can uniquely verify or falsify statements within a theory.³³ According to Quine, a theory will have a shared basis of empirical and theoretical assumptions, where the single parts of the theory cannot meet "the tribunal of experience" alone, as separated from the rest of the theory.³⁴ Davidson operates readily with a distinction between empirical and apriori, but only on an utterly pragmatic basis.³⁵ The radical interpretation thought experiment is theoretical, but it relies on the properties of natural languages and on an understanding of the workings of normal, psychologically equipped people. Still, the point is that any revision of a theory will be marked by holism in such a way that a distinction between theoretic and empiric revision, according to a basic Davidsonian view, cannot be attributed a decisive epistemic importance. This indicates that, founded on the assumption that the principle is "empirically non-revisable"; it will be difficult in any case to ascribe a particular epistemic status to the principle of charity within Davidson's philosophy.

³² In Robert Sinclair: "What is Radical Interpretation?" pp 161-84 in *Inquiry*, 45, 2002, p. 177.

³³ Or, in Davidson's formulation: "...the problem of error cannot be met sentence by sentence, even at the simplest level." Page 152 in "A Coherence Theory of Truth and Knowledge" (1983), in Davidson: *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford 2001.

³⁴ Quine is giving an account of this for the first time in "Two Dogmas of Empiricism", p. 42-46 in *From a Logical Point of View*, Harvard University Press, Cambridge, Mass. 1953.

³⁵ See for instance Davidson's comment to Peter Pagin in "Comments on Karlovy Vary Papers", page 293 and 294 in Petr. Kotatko, Peter Pagin and Gabriel Segal (eds.): *Interpreting Davidson*, CSLI Publications, Stanford 2001.

5. Conclusion

What is, strictly speaking, the objective of the principle of charity? Within radical interpretation, the original aim was to produce an empirically based and legitimated theory of meaning, a theory of the meanings of the sentences of speaker A by the language L, within the time T. This took the form of a truth theory; a theory in which T-sentences are analytical hypotheses, the relation between the uttered sentences and the surrounding world situation is a relation of satisfaction when it is fulfilled.³⁶ Thus, the theory of meaning consists of preliminary analytical hypotheses that are constantly revised as we gradually reach greater insight and understanding. In connection to this, the objective is to use the principle of charity to hold on to speakers' beliefs, while we on this basis try to figure out the denotation, or meaning, of her sentences. The point of the principle is that, by supposing that the speaker is right and is in accordance with herself, we can make out what the speaker means by what she says. Still, it is important to remember that radical interpretation describes an imagined case of interpretation, and is meant to clarify the conditions of achieving understanding in a situation where the person and his language are originally unknown. The conditions of understanding that we find through the thought experiment are assumed valuable for an inquiry into the general possibility conditions for successful communication and understanding in natural language. This generalization, through the transference to natural language, makes the status of the principle an interesting question. If the principle describes a general possibility condition for communication, and, in the form of exchange of propositional content, communication is a condition for knowledge, then a clarification of the epistemic status of the principle is obviously important.

If we were to say that the principle must be either transcendental or naturalistic, we would also be saying that there is a principal difference between the formal and the empirical or between that which is as it is because of theory and that which is as it is because of empirical factors. Or, we could say that some things are true, "come what may", and that some other things can be revised on empirical grounds, or that something is "scheme" and constant and something else is "content" and varies with the scheme. It is evident that it would be a mouthful to argue that the principle of charity, as it appears in Davidson's

³⁶ We imagine for instance that Maria say "it's snowing". "It's snowing" is true if and only if it snows in the vicinity of Maria when she utters it.

thinking (or in Quine's for that matter), is a transcendental or purely naturalistic principle; and this will particularly be so for those who maintain the first standpoint, but also for those who by "naturalistic" means "utterly empirically based". Davidson is not a naturalist in this way; he is not of the opinion that rationality and normativity can be unpacked as fundamentally physical or empirical. Thus, the Davidsonian version of the principle cannot be clarified as naturalized in this sense. If the ambition of this article was to assign the principle of charity epistemic status, where the alternatives were naturalistic or transcendental, or perhaps apriori but quasi-empirical, it is plain to see that the goal has not been reached. We could not find a reason to include the concept of "non-conclusive apriori entitlement". On the other hand, if we sought to shed some light on the epistemic status of the principle of charity, we may have succeeded.

I have argued that using the principle of charity (as it appears in radical interpretation) is no guarantee for the truth of the beliefs of the speaker or the interpreter. This is evidently more so if using the early formulation of the principle. Using the principle of charity in radical interpretation cannot guarantee that our beliefs are mostly true, because that which is established (or constituted) are preliminary hypotheses about meaning and belief. Still, these provisional hypotheses must be internally coherent; we cannot deal with a self-contradictory theory. This means that the principle to a very limited extent has a constitutive function for beliefs, as larger evidential resources will be available for later corrections and more advanced versions of the interpretational theory. Thus, our dependence on the principle is gradual; the principle is more imperative the weaker our general knowledge, and less necessary the more we know in advance. It is always needed in order to secure the best interpretation possible, but to a graded extent. We must therefore be able to consider the principle more or less necessary as well as more or less constitutive.

By the introduction of triangulation, the causal element is emphasized; the focus is set on that our ascription of belief and meaning to others, and other people's ascription of belief and meaning to us are in causal interaction with each other and are also both in causal interaction with objects and events in our common physical surroundings. The causal history of our beliefs, (as based on interaction with the world and with other language users), is partly determining the content of our beliefs. This means that the principle of charity in the form of coherence is not the only factor supporting the reasonableness of our

theory about the beliefs and meanings of others. The principle is obviously not the single actor seeing to an intersubjective agreement between interpreter and speaker. It is not until the entry of the concept of triangulation that it becomes clear to Davidson that the principle of charity can be split into two principles, the principle of coherence and the principle of correspondence.³⁷ The focus on triangulation has accentuated the correspondence side of the principle. This doesn't really make a difference related to the epistemic status of the principle, as Davidson has always assumed that the radical interpreter interprets based on the jointly available objects and events, but it does clarify and underline why it is a misunderstanding to think that coherence is Davidson's sole epistemological rationale.³⁸

³⁷ In "Three Varieties of Knowledge" (1991), in Davidson: *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford 2001, p. 211.

³⁸ I thank Jan Harald Alnes for discussions, suggestions and constructive criticism throughout the work on this article. My thanks go also to Mikael Janvid who has read and given useful comments, and to Peter Pagin who has pointed out mistakes and misunderstandings.

