DRAFT.

To appear in *the Southern Journal of Philosophy*

Davidson on Self-Knowledge:

A Transcendental Explanation

Abstract

Davidson has attempted to offer his own solution to the problem of self-knowledge, but there has been no consensus between his commentators on what this solution is. Many have claimed that Davidson's account stems from his remarks on disquotational specifications of self-ascriptions of meaning and mental content, the account which I will call the "Disquotational Explanation". It has also been claimed that Davidson's account rather rests on his version of content externalism, which I will call the "Externalist Explanation". I will argue that not only are these explanations of self-knowledge implausible, but Davidson himself has already rejected them. Thus, neither can be attributed to Davidson as his suggested account of self-knowledge. I will then introduce and support what I take to be Davidson's official and independent account of self-knowledge, that is, his "Transcendental Explanation". I will defend this view against certain potential objections and finally against the objections made by William Child.

Keywords: Self-Knowledge; Disquotational Explanation; Externalist Explanation; Transcendental Explanation; Davidson; Wright; Child.

1. Introduction

Davidson has famously taken an anti-Cartesian, third-personal point of view to be essential to the study of meaning and linguistic understanding, the view which manifests itself in his extensive use of the notion of interpretation. According to this view, "[w]hat a fully informed

interpreter could know about what a speaker means is all there is to learn; the same goes for what the speaker believes" (Davidson 1983, 148).[1] What is the consequence of this view for the discussion of first-person authority, the authority which we, with a strong intuition, concede a speaker has over the content of her semantical and mental states? This paper investigates Davidson's answer to this question. He, perhaps contrary to Quine,[2] does not deny the existence of such a difference between the speaker, as the first-person, and the interpreter, as the second-person, with regard to their knowledge of what the speaker means and believes. The problem of self-knowledge concerns explaining the fact that we know ourselves directly and non-inferentially, while others' knowledge of our attitudes is indirect and inferential. The commentators on Davidson disagree on what his solution to the problem of self-knowledge is. My aim is to settle these controversies by answering the question what Davidson's *official* account of first-person authority concerning mental and semantical content is.

I begin by an outline of Davidson's remarks on the problem of self-knowledge. I will then introduce two main accounts of first-person authority which have been attributed to Davidson, that is, the "Disquotational Explanation" and the "Externalist Explanation". First of all, I will argue that these explanations are implausible. This, however, would not raise any problem for Davidson because, second of all, I will argue that Davidson himself has already rejected these explanations. I will conclude that these explanations have been wrongly attributed to Davidson as his *official* account and that his relevant remarks on disquotation and meaning-determination are to be considered as his description of the phenomenon, rather than his explanation of it. Finally, I will introduce a "Transcendental Explanation" as Davidson's actual explanation of self-knowledge and argue that this account has been overlooked by the commentators on Davidson as his official, independent account of self-knowledge.

[1] This view can be called "Interpretationism". See, e.g., Byrne (1998), Dennett (1987 , Chapter 2), Child (1994, Chapter 1), Hossein Khani (2020a) and Bernecker (2013).

[2] Whether Quine's project of radical translation, and his thesis of the indeterminacy of translation, result in a denial of first-person authority is a matter of controversy. See, e.g., Searle (1987), Blackburn (1984, 281), Glock (2003, 201-206), and Hylton (1990/91).

2. Davidson on Self-knowledge

In his paper, "First Person Authority" (1984a), Davidson introduces the problem of self-knowledge as follows:

> When a speaker avers that he has a belief, hope, desire or intention, there is a presumption that he is not mistaken, a presumption that does not attach to his ascriptions of similar mental states to others. … What accounts for the authority accorded first person present tense claims of this sort, and denied second or third person claims? (1984a, 3)

Davidson seeks an explanation of such an asymmetry between the attributions of certain attitudes to ourselves and the attributions of similar attitudes to others, the existence of which we intuitively concede. In particular, Davidson is after an answer to this question: "What explains the difference in the sort of assurance you have that I am right when I say 'I believe Wagner died happy' and the sort of assurance I have?" (1984a, 11). He calls this asymmetry in knowledge of belief the "basic asymmetry" (1984a, 12).

Davidson's explanation of thin basic asymmetry, or the asymmetry in the sort of guarantee my interpreter and I have of the correctness of my self-ascriptions, involves the process of interpretation, meaning-belief relation and holding-true attitudes. For Davidson, "to speak is to express thoughts" (1975, 155). Having granted that, he states that when a speaker utters a sentence, the interpreter assumes that the speaker holds her sentence to be true on that occasion and she does so for two reasons: "A speaker holds a sentence to be true because of what the sentence (in his language) means, and because of what he believes" (1973, 134).[3] Now, as Davidson continues, "if you or I or anyone knows that I hold this sentence true on this occasion of utterance, and she knows what I meant by this sentence on this occasion of utterance, then she knows what I believe – what belief I expressed" (1984a, 11). This means that if my interpreter knows that I hold my uttered sentence to be true and if he knows what I mean by it, he would know what belief I have expressed by uttering it. Davidson's first assumption is that "we can assume without prejudice that we both know … that on this occasion I do hold the sentence I uttered to be true" (1984a, 12). For him, no interesting asymmetry in knowledge of mental content has yet emerged because my interpreter and I are both granted knowledge of the fact that I hold this sentence to be true on this occasion. Things are otherwise, however,

[3] See also Davidson (1974a, 142, 144-145, 152), (1975, 161-162, 167) and (1983, 147).

with respect to our knowledge of what I believe: "On these assumptions, I know what I believe, while you may not" (Davidson 1984a, 12). This is the basic asymmetry that he aims to explain.

Davidson's next step to explain the belief-asymmetry is this: "The assumption that I know what I mean necessarily gives me, but not you, knowledge of what belief I expressed by my utterance" (1984a, 12). There is something that assures us of the fact that I am almost always right about what I mean by my utterance and if so, I would be almost always right about what belief I express by it, while there is no such an assurance for the interpreter. The question, however, is what does explain the difference in knowledge of meaning between me and my interpreter? Let's call this asymmetry the "meaning-asymmetry".

At this point, the consensus on how Davidson's account proceeds disappears. In what follows, I will first introduce and criticize a reading of Davidson's explanation of the meaning-asymmetry which construes it as essentially relying on his remarks on disquotational specifications of meaning. This reading has been offered by a majority of the commentators on Davidson's account, such as Wright (2001, 348-350), Thöle (1993), Picardi (1993), Beisecker (2003) and Hacker (1997). I call it the "Disquotational Explanation" and will argue that not only is such an explanation implausible in general, but Davidson himself has argued against it. Such a construal of Davidson's remarks on self-knowledge fails to distinguish between the situations in which he is *describing* the problem of self-knowledge and the situations in which he is offering his own *explanation* of it. There is also another reading of Davidson's explanation offered by, for instance, Child (2007, 2013), Beisecker (2003), Jacobsen (2009), Shoemaker (1996), Macdonald (1995) and Gallois (1997), who claim that Davidson's explanation actually arises from his remarks on semantic externalism. I will discuss this explanation in Sections 5 and 6.1 and argue that this explanation suffers from more or less similar problems. Finally, I will seek for an alternative explanation in Davidson's works and argue that the sort of explanation that Davidson has had in mind has been a sort of "Transcendental Explanation", which can explain the meaning-asymmetry in a way free from the problems which the above two readings face. I will end the paper by responding to Child's objections to the attribution of such an explanation to Davidson. Let's start by the Disquotational Explanation.

3. The Disquotational Explanation of Self-Knowledge

How does Davidson explain the difference between me and my interpreter in the sort of assurance we have about the fact that I am right in my self-ascriptions of meaning? Davidson, at some point, states that

> the speaker, after bending whatever knowledge and craft he can to the task of saying what his words mean, cannot improve on the following sort of statement: "My utterance of 'Wagner died happy' is true if and only if Wagner died happy". An interpreter has no reason to assume this will be *his* best way of stating the truth conditions of the speaker's utterance. (1984a, 13).[4]

This passage has given rise to a sort of reading of Davidson's explanation of the meaning-asymmetry which can be introduced as follows. Suppose that a speaker, X, utters "two plus two equals four". In the case of an interpreter's attributing meaning to X's utterance, we will have the following sort of specification of what X's utterance means:

(I) X means *two plus two equals four* by her utterance of "two plus two equals four".[5]

Even this disquotational specification of what X means by her utterance can be overturned by further evidence of interpretation. For example, by borrowing the example Kripke's Wittgenstein's sceptic works with in the second chapter of Kripke's book (1982), it is possible that the interpreter later finds out that X answers by "5", rather than "125", to the question "57 + 68 =?". The speaker seems to mean something else, e.g., *quus*, and not *plus*, by "plus". In this case, the evidence leads the interpreter to modify his interpretation and use a different sentence in order to give the truth-condition of the speaker's uttered sentence:

(II) X means *two quus two equals four* by her utterance of "two plus two equals four".

Any specification of what X means by her words, from the interpreter's point of view, is hostage to evidence and apt to further "improvement" or "modification". However, and this is the main point of the Disquotational Explanation, the same cannot be true in the case of *self-ascriptions* of meaning, for instance, when we have the following specification:

[4] See also Davidson (1989, 66).

[5] Or "X's utterance of 'two plus two equals four' is true if and only if two plus two equals four".

(III) I (presently) mean *two plus two equals four* by my utterance of "two plus two equals four".

There is no possibility of mismatch between the right-hand side and the left-hand side sentences in (III): any evidence which appears to be suggesting a re-interpretation of the right-hand side sentence will supposedly suggest the same re-interpretation of the left-hand side sentence. Any evidence suggesting that "plus" means *quus* applies to the mentioned sentence (on the right) as well as the used sentence (on the left). This is the reason why, for instance, Wright construes Davidson's account as follows: while the interpreter's attribution of meaning to me, "even one that uses that very sentence to specify the content in question, is hostage to the evidence of interpretation", "when I use a sentence to specify disquotationally what *I* mean by it, there is no such hostage" (Wright 2001, 348). Wright continues:

> For whatever interpretation may teach you about the content I attach to the sentence in question will apply to both the mentioned and the used occurrences in my specification: whatever I mean by a sentence, S, I am guaranteed to be able to say with perfect accuracy what that sentence means merely by using it. ... I am uniquely assured of no error in the specification of *what* I believe. (2001, 348)

In a similar vein, Thöle also states that "we should certainly agree with Davidson that [the disquotational] way of telling the meaning of a sentence provides the speaker, but not the interpreter, with a *secure* method for saying what he means" (1993, 245). Thus, the difference in the sort of assurance which me and my interpreter have of my being right about what I mean by my utterances comes from the fact that disquoting my uttered sentence and using it to specify what it means guarantees my success in correctly specifying what I mean by it, while there is no guarantee that doing the same thing in the case of the interpreter results in a correct interpretation of my utterance. The speaker's best and only way to specify what she means by her utterance is to specify it in the disquotation way, while the interpreter's best and only way to specify the meaning of the speaker's utterance may not always be to specify it in the disquotational way. And supposedly this makes the speaker immune to errors of the sort the interpreter is always susceptible to.

In what follows, I will argue for two claims: (1) The Disquotational Explanation is an implausible account of self-knowledge. (2) More importantly, Davidson himself has rejected such a sort of explanation of self-knowledge.

4. On the Disquotational Explanation

The first point to note is that the Disquotational Explanation results in the infallibility of the speaker's authoritative knowledge of what she means and believes, while this is a claim that Davidson explicitly rejects: "First person authority is not infallible" (Davidson 1984a, 13). Davidson denies what Wright's and Thöle's readings attribute to him, i.e., that "I am uniquely assured of no error in the specification of *what* I believe", as Wright said above, or that "the speaker, but not the interpreter, [is provided] with a *secure* method for saying what he means", as Thöle said. Rather, for Davidson, I am *not* guaranteed with any immunity to error about what I mean and believe. The speaker is not always right because she may fail to speak in a way understandable to her interpreter: "It is possible for the evidence available to others to overthrow self-judgements" (Davidson 1984a, 4). According to Davidson, a speaker can be said to be successfully meaning something by her utterance if her utterance is interpreted as she intends.[6] Otherwise, the speaker fails to mean anything at all and we can thereby say that the speaker has just thought, or it just seemed to her, that she meant something by her words. As Davidson puts it, "the speaker may fail in this project [of remaining interpretable to others] from time to time; in that case we can say … that he does not know what his words mean" (1984a, 13). It is also important to note that Davidson's claim here is not that the words have a meaning but the speaker fails to know it. Rather, there would be no meaning to be known at all. This is part of the reason why I think Child's (2007) interpretation of Davidson's account of self-knowledge is problematic, according to which Davidson believes that there are cases in which my utterance has a meaning but I have no knowledge of what that meaning is, such as the cases in which the speaker intends to use her words in accordance with a socially accepted way. For Child, in this situation, the speaker knows her intention to use her words in that way, but since "she has no special, authoritative way of knowing what the word does mean on others' lips, she has no special, authoritative knowledge of what it means on her lips, either" (2007, 163). This is not a claim that Davidson endorses. For Davidson, it does not matter how a speaker uses her words, that is, whether she uses them in a socially accepted way or not; what matters is that the speaker fails to mean anything by her words *if* she fails to speak in an interpretable way. I will say more about this below and in Section 6.1.

[6] According to Davidson, "if the speaker is understood he has been interpreted as he intended to be interpreted" (1986, 93). See also Davidson (1986, 97, 99, 101, 1994, 120, 1991), and (1992, 111-112, 116, 1987, 28).

More importantly, there are two main questions regarding the Disquotational Explanation which need to be answered by Davidson *if* it is the view that he officially holds: (I) Suppose that when the speaker makes a self-ascription of meaning, she does so in the disquotational way. Is doing so *sufficient* to show that she knows what she means by her utterance? (II) For the speaker, in order to specify what she means by her utterance, is it *necessary* that she does so in the disquotational way? Let's start by the question about sufficiency.

Consider this statement:

(IV) I mean *Está lloviendo* by my utterance of "Está lloviendo".

This statement is always true. But the point is that I can always give the meaning of *any* sentence like this by simply disquoting it on the left-hand side of the statements like (IV). I, however, *ex hypothesi* do not know what "Está lloviendo" means. Although I do not know what the sentence means, I will always be right in so specifying the truth-condition, that is, the meaning of that sentence.[7]

One may object that here, *ex hypothesi*, the speaker does not know Spanish – she is an English speaking speaker – and thus she does not *understand* what that sentence means. But this is the whole point of the example. It may be better to rewrite (IV) as follows:

(V) I mean *Está lloviendo* in Spanish by my utterance of "Está lloviendo".

The speaker may not know the language but she is still capable of using the disquotational method to *correctly* specify what that sentence means. The speaker's ability to disquotationally specify the meaning of such sentences does not show that she has any authoritative, non-inferentially knowledge of what her utterance means. For instance, Child (2013) has claimed that Davidson's point on disquotational specifications of meaning in the case of self-ascriptions is to be read as claiming something like the following: "Suppose … I *know* what it means. Then

[7] This problem can be put in terms of the difference between knowing-how and knowing-that, that is, the difference between the speaker's knowing that "S" means *S* and the speaker's knowing how to use S. Knowing how to use S to specify what "S" means is one thing, knowing what "S" means is another. The Disquotational Explanation reduces knowing-that to knowing-how: *as if* once I know how to use S to specify the meaning of "S" by simply disquoting "S", I thereby know what "S" means; *as if* knowing how to use S is all we need to explain the speaker's authoritative *knowledge* of what "S" means. As indicated above, this would not suffice to show that the speaker knows what "S" means. I am thankful to an anonymous referee for this journal for drawing my attention to this point. On this distinction, see also Jacobsen (2009).

I can use that sentence to state its own meaning in a way that is proof against the kind of error to which I am vulnerable when I use my words to state the meanings of someone else's words" (2013, 538, emphasis added). This claim, however, is question-beginning because it presupposes what was promised to be explained, that is, the speaker's knowledge of what her utterance means. This explanation fails to explain the essential difference which we already concede there is between the speaker's knowledge of the meaning of her utterance and the interpreter's knowledge of what the speaker's utterance means. We were after explaining what makes it the case that I have non-inferential *knowledge* of what I mean. But, instead of explaining that, we have presupposed that she does already know what she means and that if she does know what she means, she can specify it by using the disquotational method. Not only is this explanation question-begging, but it also results in taking the speaker to be *infallible*, immune to error in specifying the meaning of her utterances; as previously indicated, this is a claim that Davidson rejects.

One may insist that the speaker, but not the interpreter, is bound to specify the meaning of her utterance in the disquotational way *because* the speaker is to use the same language, her own language, to specify the meaning of her own words, while the interpreter has his own idiolect, or in this sense, his own language to specify the meaning of the speaker's utterance. But this move does not work either. For it would not be impossible to imagine that the speaker's and the interpreter's languages are the same, even their idiolects.[8] For instance, assume that they are identical twins who have learnt their first language under the same kind of (triangular) situation. On this scenario, they both mean the same thing by the same words: they both mean *arrangement* by "arrangement", *green* by "green", and so on. We can say that they are almost alike in being immune to errors about the meaning of one another's utterances because they are both capable of disquotationally specifying what the other's utterances mean. They both know what the other's sentences mean because their languages are the same. The point is that, even in this situation, we still *need* an explanation of self-knowledge because the asymmetry does not disappear in this case. The reason is that although supposedly they speak the same language, mean the same thing by the same words, and know what those words mean, the speaker knows what she means by her utterance in a different way. The interpreter lacks such non-inferential authoritative knowledge, though he, like the speaker, is capable of specifying what the speaker means by her utterance by employing the disquotational method. The original asymmetry is

[8] On this possibility see Ludwig (1994, 388-389).

left unexplained. These points are related to my second question as to whether it is *necessary* that self-ascriptions of meaning are specified disquotationally.

It seems that the asymmetry exists even when the way in which the speaker can specify what she means by her words is *not* best specifiable in the disquotational way. Recall Davidson's example of Mrs. Malaprop: she meant *a nice arrangement of epithets* when she uttered "A nice derangement of epitaphs".[9] According to Davidson, the hearer usually has no problem understanding what Mrs. Malaprop means by her utterance if there is enough evidence and clue for him to reach Mrs. Malaprop's *intended* interpretation of the utterance. Surely there is still this difference: Mrs. Malaprop non-inferentially knows what she means by her utterance, while the interpreter, as Davidson says, relies on evidence, observation, and much general information to successfully interpret Mrs. Malaprop's utterance. The sentence that Mrs. Malaprop uttered was "A nice derangement of epitaphs". When could we say that the interpreter has *correctly* interpreted this utterance? According to Davidson, we can say so only when the interpreter interprets Mrs. Malaprop's utterance as meaning *a nice arrangement of epithets*, since, *ex hypothesi*, this is the way Mrs. Malaprop intends her utterance to be understood. Now, can't we say that Mrs. Malaprop knows herself to mean *epithet* by "epitaph", or *arrangement* by "derangement"? I believe, on Davidson's view, we can and, in what follows, I will argue for why it is so.

For Davidson, we use the same words to mean different things every day and our hearers have no trouble understanding what we say.[10] Mrs. Malaprop has non-inferential knowledge of what she means by "epitaph"; and note that what she means by it is not *epitaph*. We can say that she cannot generally *improve* on the following sort of statement:

(a) I mean *epithet* by "epitaph".

There is no problem in saying that her language (idiolect) includes both such words. But the point is that if she uses the disquotational method, she *fails* to specify what she really intended to mean by this utterance. She is not bound to specify what she means *only* in the disquotational way. This is so while her interpreter has no immediate reason to take either of the following statements as offering his best interpretation of Mrs. Malaprop's utterance of "epitaph":

[9] See Davidson (1986, 90, 94-95, 103-104) and (1994, 115).

[10] See Davidson (1986, 89-90, 98-99) and (1994, 115).

(b) Mrs. Malaprop means *epitaph* by "epitaph".

(c) Mrs. Malaprop means *epithet* by "epitaph".

(d) Mrs. Malaprop means *epiphany* by "epitaph".

The interpreter might choose one of the above or other interpretations of Mrs. Malaprop's utterance and check its plausibility against further evidence. However, this is not what happens in the case of Mrs. Malaprop's self-ascriptions of meaning. This is exactly the kind of asymmetry that was supposed to be explained and this asymmetry has nothing to do with the claim that Mrs. Malaprop's self-ascriptions of the meaning of her utterances are to be disquotational. This claim is perfectly compatible with, and actually follows from, Davidson's later view of meaning. In order to see how, let's focus a bit more on this view, especially on the Davidsonian distinction between "first (or literal) meaning" and "speaker-meaning".

Davidson believes that "meaning … gets its life from those situations in which someone intends (or assumes or expects) that his words will be understood in a certain way, and they are" (1994, 120). This notion of meaning is what Davidson calls that of "literal" or "first" meaning of utterances: "How he intended to be understood, and was understood, is what he, and his words, literally meant on that occasion" (Davidson 1994, 120). It is called "first meaning" (Davidson 1986, 91) because this is the meaning that the interpreter should grasp first if he is to be able to understand other meanings, intentions, subsequent effects and ulterior purposes the speaker has in mind when she utters those words. Let's call these the "speaker-meanings" (Davidson 1986, 91). For instance, her friend may use Mrs. Malaprop's utterance of "That's a nice derangement of epitaphs" to make an irony, a complement, and the like.[11] None of these can be understood if he fails to interpret Mrs. Malaprop's utterance as meaning that *that's a nice arrangement of epithets*. First meaning comes first in the order of interpretation and it is what the uttered words literally mean on that occasion. Davidson clarifies this claim by distinguishing between "passing" and "prior" theories of interpretation:

> For the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he *does* interpret the utterance. For the speaker, the prior theory is what he *believes* the interpreter's prior theory to be, while his passing theory is the theory he *intends* the interpreter to use. (1986, 101)

[11] On this, see Ludwig and Lepore (2005, 265-266).

For Davidson, "what two people need, if they are to understand one another through speech, is the ability to converge on passing theories from utterance to utterance" (1986, 106). This is just to restate that the interpreter must interpret the speaker's utterance in the way the speaker intended it to be understood. This is the conclusion of Davidson's argument against the sort of view which claims that the existence of, and conforming to, certain conventions about the meaning of words is "necessary to the existence of communication by language" (1984b, 265).[12] Davidson, however, believes that in order for communication between two people to be successful, conforming to such conventions – or following certain rules determining the correct use of words – is *neither necessary nor sufficient*. It is not necessary because we can understand what Mrs. Malaprop means by "epitaph" on this specific occasion without appealing to the conventional meaning of the word. After all, the conventional, standard, or dictionary-based meaning of "epitaph" is *epitaph*, not *epithet*. Nor is knowledge of the conventional meaning of the words sufficient for the communication between them to be successful because in order to understand what Mrs. Malaprop means by "epitaph", even if she means what "epitaph" conventionally means in her speech-community, we need knowledge and information over and above mere knowledge of what "epitaph" conventionally means: "It is derived by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely" (Davidson 1986, 107). For instance, we need to know that Mrs. Malaprop now intends her utterance to be interpreted in the standard way and this requires a sort of knowledge additional to mere knowledge of the conventional meaning of the words.[13] As Davidson says, "even when a speaker is speaking in accord with a socially acceptable theory he speaks with the intention of being understood in a certain way" (1994, 122).

Davidson believed that the interpreter's and the speaker's passing theories must converge if their communication is to be successful. In the case of Mrs. Malaprop's utterance, as Davidson puts it, "Mrs. Malaprop's theory, prior and passing, is that 'A nice derangement of epitaphs' means a nice arrangement of epithets" (1986, 103). In this passage, Davidson's own example has just provided us with a *non-disquotational specification of meaning* in the case of Mrs. Malaprop's *self-ascription of meaning*. If, according to Mrs. Malaprop's passing theory, she

[12] Davidson characterizes this view differently in several of his writings. Cf. Davidson (1986, 102), (1994, 110), and (1991, 143).

[13] For more on this point, see Hossein Khani (2020b).

means *that's a nice arrangement of epithets* by her utterance of "That's a nice derangement of epitaphs", Mrs. Malaprop's self-ascription of meaning *cannot* be specified disquotationally in this case. From these points it follows that Davidson does not believe that it is necessary that self-ascriptions of meaning are specified disquotationally. As he takes malapropisms and similar phenomena to be "ubiquitous" (1986, 89), the Disquotational Explanation cannot be regarded as his intended explanation of self-knowledge. Attributing this explanation to Davidson as his official account is to misconstrue his view. This is the reason why Davidson confesses that his "article 'First Person Authority' perhaps did not sufficiently emphasize that my 'solution' to the problem about self-attributions of attitudes depended on my theory of meaning" (1993b, 250).

What I have argued for so far would also rule out another parallel misconstrual of Davidson's account, according to which self-ascriptions of meaning are to be disquotationally specified because it is a *convention* of the speaker's language to do so in such cases. Davidson, as previously indicated, is generally against the views that take the existence of such conventions to be necessary or sufficient for success in understanding what the speaker means by her utterances. His arguments against such Conventionalist (and Communitarianist) views appear first in his paper "Communication and Convention" (1984b), followed mainly by "A Nice Derangement of Epitaphs" (1986), "The Second Person" (1992) and "The Social Aspect of Language" (1994). The limitations of space do not allow me to unpack his criticisms of this view here. But Davidson, in criticizing Strawson's account of self-knowledge, makes a similar point, which also helps to clarify why his explanation of self-knowledge cannot be the Disquotational one.

For Davidson, it is just an "uninformative and unexplained claim that it is a convention of language to treat self-ascriptions with special respect" (1984a, 10), that "a speaker who sincerely uses a certain sort of sentence must be presumed to be right in what he says" (1984a, 10). Such claims are question-begging because they presuppose what was promised to be explained: they just state that we "should interpret self-ascriptions in such a way as … to assign a special priority to their truth" (Davidson 1984a, 10). However, what reason do we have for treating them in this way? To say this is just to repeat, or *re-describe* the problem: "Self-ascriptions have special authority: true; and that is where we began" (Davidson 1984a, 10). To claim that the speaker is bound to disquotationally specify what she means by her utterance, as a matter of conforming to some convention, takes us back to the problem of explaining *why*

there must be such different criteria in the application of the terms like "means that" and "believes that". For Davidson, "we may postulate different criteria of application for the key concepts or words ('believes that', 'intends to', 'wishes that', etc.). But these moves do no more than restate the problem" (1984a, 11). This forms Davidson's main objection to what he calls the "Wittgensteinian views", such as that of Strawson's, which aim to explain the meaning-asymmetry by appealing to the claim that self-ascriptions of meaning (and thus of belief) are not based on evidence or observation, while others' ascriptions of meaning (and belief) to the speaker are.[14] It is the role evidence plays in such ascriptions that supposedly explains the asymmetries in question. Davidson correctly detects a deep problem with such accounts: "This feature of first person authority, suggestive as it may be, does not help explain the authority" (Davidson 1984a, 5). He cites two reasons for this claim. According to the first,

> it is a strange idea that claims made without evidential or observational support should be favored over claims with such support. Of course, if evidence is not cited in support of a claim, the claim cannot be impugned by questioning the truth or relevance of the evidence. But these points hardly suffice to suggest that in general claims without evidential support are more trustworthy than those with. (1987, 16)[15]

Nonetheless, I do not think that this claim by itself suffices to rule out the aforementioned Wittgensteinian view since the view after all offers *some* explanation of why we should distinguish between self-ascriptions of meaning and others' ascriptions of meaning to the speaker. Davidson's more important objection to these views is that they invite scepticism about other minds. For him, identifying a concept as the concept it is, for the most part, relies on the sort of application conditions we have for it. "A table" is the concept it is partly because it is (as intended by the speaker on a specific occasion) applicable to certain things only (i.e., tables) and not to other things. If we have two radically different criteria – or application conditions – for supposedly the same concept or expression, or, as Davidson puts it, "if what is apparently the same expression is sometimes correctly employed on the basis of a certain range of evidential support and sometimes on the basis of another range of evidential support (or none), the obvious conclusion would seem to be that the expression is ambiguous" (1987, 16). For Davidson, similar considerations apply to the application of expressions like "believes

[14] Wittgenstein's account of the meaning-asymmetry is based on the differences between the rules governing the application of the expressions like "means that" and "believes that" as they are used in different language games. Strawson's account is based on such a Wittgensteinian view.

[15] See also Davidson (1984a, 5).

that" and "means that". He asks: "Why then should we suppose that a predicate like '*x* believes that …', which is applied sometimes on the basis of behavioral evidence and sometimes not, is unambiguous?" (1987, 16). And if it is ambiguous, we have no reason to assume that it preserves its meaning when used by the speaker in the case of self-attributions of attitudes and by the speaker in the case of attributing them to others. This would invite the sceptic who questions whether the attitudes we are attributing to ourselves are the same as those we attribute to others. This is the reason why Davidson thinks that "Strawson and Wittgenstein had *described* the asymmetry, but had done nothing to explain it" (1993a, 211).[16] Therefore, for Davidson, all the views which claim that the asymmetry is explained by appealing to the fact that self-ascriptions of meaning are not based on evidence, while others' ascriptions of meaning to the speaker are, "merely restated the asymmetry without explaining it" (1993b, 249). It is easy to see how this objection can be re-applied to the Disquotational Explanation: this account does nothing to explain the asymmetry; it rather offers two radically different criteria for the application of "means that".

If Davidson's remarks on self-knowledge is to be read as so far suggested, how should we interpret Davidson's claim that the speaker cannot generally "improve" on the statements like "My utterance of 'Wagner dies happy' is true if and only if Wagner died happy"? I think Davidson here is *describing* the meaning-asymmetry in certain cases, rather than explaining it. He is describing the sort of situations in which self-ascriptions of meaning can best be specified by using the disquotational method. Normally, the best I can do is to give the meaning of my utterance by using that sentence. I utter "It's raining" and if I mean *it's raining* by it, this is best specifiable via simply disquoting the uttered sentence. I may, however, mean something different by my utterance. In such cases, as Davidson says, "the speaker … cannot wonder whether he generally means what he says" (1984a, 12). This is, as indicated above, the way Davidson *describes* the asymmetry since the question remains as to what explains such an asymmetry in knowledge of meaning. Therefore, we should carefully distinguish between the situations in which Davidson is offering his own explanation and the situations in which he is merely describing the asymmetry by, for instance, appealing to the notion of evidence, or that of disquotational specifications of meaning.

[16] See also Davidson (1993b, 248-249). For Strawson's account, see Strawson (1959).

So far, I have argued that the Disquotational Explanation is not plausible; nor is it Davidson's official explanation of self-knowledge since it is incompatible with Davidson's fundamental remarks on meaning, communication and self-knowledge. The Disquotational Explanation is not really part of Davidson's *explanans*; rather a re-description of his *explanandum*, though it was not even a comprehensive description of the *explanandum*. Those like Wright (2001, 348-350), Thöle (1993), Aune (2012, 216-217), Child (2007, 159, 2013, 536), Ludwig and Lepore (2005, 353-354) and many others who claim that Davidson's official explanation comes from his remarks on the disquotational specification of self-ascriptions of meaning and belief have misrepresented Davidson's account: they failed to distinguish between the situations in which he describes the asymmetry and those in which he is explaining it. Before introducing the explanation which I think Davidson has actually had in mind in his discussion of self-knowledge, it is worth considering a different reading of Davidson's account which construes it as an Externalist Explanation of self-knowledge.

5. The Externalist Explanation

It has been claimed that the speaker is always capable of disquotationally specifying what she means by her utterance because the *cause* of her utterance and that of the sentence she *uses* to specify the meaning of that utterance is the same. For instance, Child (2007) states that, according to Davidson, "I will always be able to supplement the disquotational statement by identifying objects or events to which the word applies" (2007, 162). This Externalism Explanation and the Disquotational Explanation are not meant to be mutually exclusive, though the Externalist Explanation does not need to provide support for the Disquotational Explanation or to entail it. As we will see, it can be regarded as an independent explanation claiming that speakers know what they mean by their words because the meaning of their words is determined by the things in the world which typically cause them to apply those words to. However, for similar reasons previously discussed, I think it is already clear that this attempt would not be successful in explaining the meaning-asymmetry and that it cannot be taken to be Davidson's official explanation of self-knowledge. For one thing, Davidson says that "the agent herself … is not in a position to wonder whether she is generally using her own words to apply to the right objects and events" (1987, 37). But, as he emphasizes, by so describing the asymmetry, we have "done nothing to explain it" (1984a, 8) since "it remains to show why there must be a presumption that speakers, but not their interpreters, are not wrong about what

their words mean" (1984a, 12) and, considering Davidson's externalism about content, we can add that it remains to show why there must be a presumption that speakers, but not their interpreters, are not wrong in identifying the objects or events to which they apply their words. I agree with Child on his claim that "[e]verything depends on what explains the presumption" (2007, 164); I do not, however, agree with him on the claim that Davidson's explanation of this presumption comes from his (externalist) considerations about meaning-determination. For consider the situation in which two subjects are learning their first language. At least in such basic situations, we can assume that both of them, i.e., the speaker and the interpreter, speak the same language. By way of Davidson's remarks on externalism, this is just to say that they have been (typically) caused by the same things – the same stimuli in the world – to respond in the same way. They both respond by "table", for instance, in the presence of the whole table in view and hence, according to this explanation, they both can be said to mean *table* by that utterance. The problem is that, given all these, the asymmetry is still left unexplained: What does explain the sort of difference there is between the speaker's knowledge of what she means by "table" and the interpreter's knowledge of the meaning of the speaker's utterance? This is a problem very similar to that which was discussed in the case of the Disquotational Explanation.

Child claims that "[Davidson's] *fundamental* justification for the presumption that speakers generally know what their words mean comes from considerations about meaning-determination; it does not come directly from considerations about the *process* or *procedure* of interpretation" (2007, 165). Child too seems to overlook the difference between Davidson's description of the asymmetry and his explanation of it. It is true that, for Davidson, "the agent … [cannot] wonder whether she is generally using her own words to apply to the right objects and events" (Davidson 1987, 37). This passage form Davidson is the main reason for Child to attribute the Externalist Explanation to him. But Davidson's treatment of the asymmetry has shown us that this is *not* the way in which the asymmetry can be explained. At most, we are *describing* it. For the problem still is to explain *why* there is such a difference, such a presumption; why does the agent have such authoritative knowledge of the meaning of her words *so determined*? It is one thing to claim that what you mean by your words is partly determined by what cause you to utter those words – given Davidson's externalism, the external causes contribute to the determination of the content. But it is completely another to *explain* why the speaker has authoritative knowledge of such meanings. Does the speaker have authoritative knowledge of the meaning of her utterances simply because of having direct knowledge of the causes of her responses? It is dubious whether we can make a clear sense of

the claim that one has direct knowledge of what causes one's responses and a strange enough claim to doubt its whole plausibility. Nonetheless, even given that one can be credited with such knowledge, we can still see that our original problem is merely relocated: What does explain such a difference between the speaker and the interpreter with regard to their knowledge of such causes? We have, as Davidson said, *re-stated* the problem. Moreover, there are utterances which are not directly prompted by such external causes. Do we need to track the causes of such utterances back to their external ancestors in order to be credited with authoritative knowledge of the meaning of our utterances? In what sense, then, is our knowledge of the meaning of such utterances *non-inferential* for us? This is one reason why the Externalist Explanation cannot explain the "how" question with regard to self-knowledge either. The "how" question concerns explaining *how* it is that we know ourselves non-inferentially, while others have no such knowledge of what we mean and believe. How can each speaker acquire first-personal authoritative knowledge of what she means and believes? The Externalist Explanation is stuck at the level of explaining whether there is such an asymmetry at all, that is, it even fails to answer the "why" question with regard to self-knowledge. I will come back to this issue again in Section 6.1.

What Child takes to be Davidson's "fundamental justification" is indeed what Davidson is concerned with in his discussion of the notion of triangulation, which offers a causal explanation of meaning-determination, or better meaning-emergence. Davidson uses this notion to describe the primitive situations in which meaning and mental content may emerge, that is, the situations in which two creatures are assumed to similarly respond to the same stimulus in the world as well as to each other's responses. Davidson's aim is to say something constructive about what it takes for such creatures to come up with meaningful responses and contentful mental states.[17] But we should note that this explanation adds nothing new to what Davidson has already offered in his discussion of interpretation since what Davidson eventually leads us to, in his discussion of triangulation, is to put even more emphasis on the essentiality of *the process of interpretation*, though this time a causal story has been added to such a process. As Davidson says, in order for the meaning of their responses to get determined, "each [of the triangulators] must speak to the other and be understood by the other … they must each be an interpreter of the other" (1992, 121). However, it is crucial to bear it in mind that

[17] For Davidson's discussion of triangulation, see Davidson (2001c, 8-9), (1997, 26-27), (1992, 117) and (1982, 327). I will return to this discussion in Section 6.1.

regardless of what caused me to utter the words I did, the problem is to explain why I non-inferentially know what I mean by my *so caused* utterance. Suppose again that the whole table in view caused me to utter "table". Why do I know what I mean by my utterance of "table" differently from the way the second-person can know what I meant by it? For Davidson, regular application of words to objects in the world by itself would do *nothing* to determine the meaning of the speaker's words, *unless* the speaker's utterance is successfully *interpreted* by at least another speaker. If so, then once you can eventually be considered as a competent user of a language, "you can change the meaning provided you believe … that the interpreter has adequate clues for the new interpretation" (Davidson 1986, 98). I may intend to mean *chair*, *table*, *tree*, or anything else by my utterance of "table" which is *caused* by the same specific aspect of an object in view. I may even mean something entirely new by that word. The object, e.g., the table in view, in this sense, has never caused me to utter "table" to mean that new thing. If my utterance is interpreted as I intended, that is, to be meaning what I intended it to mean, that meaning is what my word literally means on that occasion and, consequently, I non-inferentially know it. Regular application of words to certain things in the world, hence, does not provide us with any explanation of self-knowledge of the sort Davidson has promised.

However, if the above two explanations that have been attributed to Davidson are implausible and have been already rejected by Davidson, what is his official explanation?

6. Davidson's Transcendental Explanation of Self-Knowledge

I think Davidson's explanation of self-knowledge is a transcendental one stemming from his remarks on the process of interpretation and this "Transcendental Explanation" is what Wright, Aune, Child, Thöle and many others have not considered as Davidson's *official* explanation of self-knowledge.

Davidson's claim was that if I non-inferentially know what I mean, I non-inferentially know what I believe, while the interpreter's knowledge of what I mean and believe cannot be achieved in the same way. The interpreter reaches my belief only after successfully interpreting my utterance, that is, after engaging in the process of interpreting my utterance by utilizing the available evidence and clues. According to Davidson, in order to know what I mean and, hence, what I believe, I *cannot*, as he calls it, engage in such "a difficult inference" (1984a, 13). Davidson's account, as so far introduced, leads to the claim that "a hearer interprets … on the

basis of many clues; … The speaker … cannot wonder whether he generally means what he says" (1984a, 12). These are still descriptions of the asymmetry. Davidson needs to explain why it is so. He says that if I know what I mean by my utterance,

> it follows that I know what I believe, but it does not follow that you know what I believe. The reason is simple: you may not know what I mean. Your knowledge of what my words mean has to be based on evidence and inference: you probably assume you have it right, and you probably do. Nevertheless, it is a hypothesis. (1989, 66)

But, we saw that making such a claim by itself is not enough. The promised explanation is not yet completed since the meaning-asymmetry remains to be explained. According to Davidson, there must be "a presumption that speakers, but not their interpreters, are not wrong about what their words mean" (1984a, 12). What is Davidson's reason for the claim that there *must* be such a presumption?

Davidson's explanation of the meaning-asymmetry comes from his remarks on the necessary conditions on the possibility of interpretation and this is one reason for why this explanation is transcendental. As Davidson says, "there is a presumption – *an unavoidable presumption built into the nature of interpretation* – that the speaker usually knows what he means" (1984a, 14, emphases added); this "presumption *is essential to the nature of interpretation*" (Davidson 1984a, 12, emphases added). He emphasizes again that "the presumption … *is essential ... to my being interpretable at all*" (1989, 66, emphases added). According to this Transcendental Explanation, if speakers were not mostly right about what they mean, interpretation would not be possible: there would be nothing to interpret and no such thing as interpretation at all. It is a necessary condition on the possibility of interpretation that speakers non-inferentially, that is, free from engaging in the process of interpreting themselves, know what they mean and believe.

In the beginning of "First Person Authority" (1984a, 3), Davidson points to a *presumption* regarding the existence of the asymmetry in question, which is already in play in our interactions with others and which needs to be accounted for. Presupposing the asymmetry in order to explain why it exists is one thing, presupposing the asymmetry without explaining it – and, as Wright (2001, 348-350), Thöle (1993) and others have accused Davidson of, taking for granted what was promised to be explained – is another. For instance, Wright says that "normally, we are credited with a special authority for the character of our own intentions", and "a little reflection shows that both these features – non-inferentiality and indefinite 'fecundity' – are simply characteristic of the normal intuitive notion of *intention* [and

meaning]" (2001, 111-112). Of course, Wright believes that we need to explain such features and in order to do so, he offers his judgement-dependent account of meaning and intention.[18] But, *assuming* that speakers are by default credited with such an authority is not a problem for him. It is surprising, however, that he criticizes Davidson for doing exactly the same.[19] The problem with Wright's accusation is that he misses the point that Davidson assumes the meaning-asymmetry in order to explain why it exists. As Davidson emphasizes in his criticism of Ryle, the problem with Ryle's account of self-knowledge is that he "neither accepts nor explains the asymmetry; he simply denies that it exists. … I think it is obvious that the asymmetry exists" (1984a, 6). According to Ryle, we, as first-persons, only occupy a better position than others do to observe ourselves: the differences "are difference of degree, not of kind", as Davidson construes Ryle's view (Davidson 1984a, 5).[20] For Davidson, however, these claims merely describe the symmetry.[21] What Davidson offers in order to explain the meaning-asymmetry (and hence the belief-asymmetry) is his Transcendental Explanation: the speaker's knowledge of the meaning of her own utterances (and the content of her own attitudes) is direct and non-inferential, is not based on evidence and observation, and is not rested on the process of interpreting herself, *because* if it was so dependent, interpretation would not be possible.[22] One may object that the explanation as it stands is still incomplete because it needs to answer

[18] For this account, see, e.g., Wright (2001), (1992) and (1988).

[19] See, e.g., Wright (2001, 349).

[20] For Ryle's view, see Ryle (1949).

[21] Considering Davidson's criticism of Ryle's treatment of the problem of self-knowledge, it would be difficult to sympathize with Aune's (2012) construal of Davidson's account, according to which this account leads to the conclusion that "I am in a better position to know what I believe in holding true 'I believe Wagner died happy' than anyone else" (2012, 215).

[22] It is worth noting that this account can also explain what we may call the "attitude-asymmetry". Davidson's claim was that we can without *prejudice* assume that both the speaker and the interpreter know that when the speaker utters a sentence on an occasion, she holds it to be true on that occasion. Davidson states that he credits *both* the speaker and the interpreter with such knowledge of the speaker's holding-true attitude (see, e.g., Davidson (1993b, 250)). Thus, it seems that no asymmetry between the speaker and the interpreter should arise with regard to knowledge of this attitude of the speaker. Call it the "attitude-asymmetry". Elsewhere, he claims that it would "make the account circular to explain the basic asymmetry by assuming an asymmetry in the assurance you and I have that I hold the sentence I have just uttered to be a true sentence. There must *be* such an asymmetry, of course, but it cannot be allowed to contribute to the desired explanation" (1984a, 12). Here, his claim seems to be that there *is* such an "attitude-asymmetry" but it does no play any constitutive role in his account. The first claim seems implausible to me since there certainly *is* the "attitude-asymmetry" since after all the speaker knows her *own* attitudes, whatever they are, differently from the way her interpreter knows them. But I think Davidson is right in his second claim that such an asymmetry plays no constitutive role in his explanation since the attitude-asymmetry is intended by Davidson to be explained by his Transcendental Explanation.

the question why interpretation would not be possible if it was not the case that speakers have non-inferential knowledge of what they mean and believe.

Davison, I think, offers two reasons. He states that "what a speaker says can be misinterpreted by others, but it cannot be misinterpreted by the speaker because no content can be given to the idea of interpreting one's own words" (1993b, 250). Here, the idea is that it simply makes no sense to claim that in order to know what the speaker means by her words she has to interpret herself. We can add that it does not make sense because doing so leads to a vicious regress of interpretations and, as a result, the speaker would never be able to intend or to mean anything at all: there is no finite list of conditions the speaker should take into account in order to intend her words to mean something specific and, thus, to reach something stable which can be interpreted by her interpreter at all.[23] However, one may insist that even if we agree with Davidson on this point, it has not yet provided us with an answer to our question because although it does not make sense to claim that the speaker has to interpret herself in order to understand what she means and believes, it does not yet explain why this makes *interpretation* impossible in general. Davidson's reply can simply be that, in that case, there would thereby be nothing for the interpreter to interpret. If the speaker cannot be said to be capable of intending her utterance to have a determinate interpretation, there would be nothing for the interpreter to interpret; he would have no reason to treat the speaker as a rational agent at all. Interpretation is possible, after all, if something interpretable is available. To be a rational agent, a subject is to have a rich set of mental states with determinate content and utterances with determinate meaning.[24] Let's focus a bit more on this claim.

For Davidson, even when speakers make an utterance in the absence of any audience, for example, when they say to themselves "What a nice day!", "yet it matters what words are used, what they mean. ... there must be some *reason* for using those words, with their meaning, rather than others" (Davidson 1984b, 272). What is important in our present discussion is not the speakers' "reason" for choosing the words they utter – maybe they have just learnt, or been conditioned, to use the words in the way they do, or as Davidson says, perhaps "it just comes naturally" to talk in this way, for "what magic ingredient does holding oneself responsible to

[23] Compare this with Davidson's similar remarks on "pure intending" and the contrast between prima facie (conditional) vs. all-out (unconditional) judgements about the desirability of an action. See Davidson (1978).

[24] See Davidson's "Rational Animals" (1982) for additional remarks on this matter, such as the argument from the holism of the mental.

the usual way of speaking add to the usual way of speaking?" (Davidson 1994, 117). What matters here is the meaning the words have for the speaker, i.e., their intended interpretation, regardless of how she has been particularly taught, conditioned, disposed, or caused to use the words in the way she does. The words have a meaning for the speaker (and not necessarily a conventional or standard one, but what the speaker intends them to mean) and the speaker chooses those words with the meaning they have for her to express her beliefs. Without presuming the speaker's knowledge of how she intends her words to be interpreted, there would be no meaning to be interpreted, no ground for the interpreter to even start her interpretation of the speaker's responses at all. Without the assumption that the speaker knows what she means and believes, there would be *no answer* to the question why the speaker uttered the words she did and hence no reason for the interpreter to treat the speaker's responses as utterances with a determinate meaning and generally to treat her as a rational agent. As Davidson clarifies, "[w]hat is impossible is that she should be wrong most of the time. The reason is apparent: unless there is a presumption that the speaker knows what she means, i.e. is getting her own language right, there would be nothing for an interpreter to interpret" (1987, 38). This leads to Davidson's famous claim that "whether we like it or not, if we want to understand others, we must count them right in most matters" (1974b, 197).

Granted that this is Davidson's official explanation of self-knowledge, let's consider Child's potential objections to my defence of the Transcendental Explanation as Davidson's official explanation.

6.1. Child's Objections

At some point, Child considers the claim that Davidson's remarks on interpretation are to be treated as the real source of Davidson's explanation of self-knowledge and attempts to resist the objection that his Externalist Explanation "misrepresents Davidson's view" (Child 2007, 164). He makes three claims, which I think are all implausible, considering my discussion of the Externalist Explanation in Section 5. In this part, I will respond to these objections and accordingly develop my criticisms of the Externalist Explanation.

According to his first attempt, he agrees that Davidson puts a huge emphasis on the process of interpretation, but he denies that this can show that his Externalist Explanation is thereby wrongly attributed to Davidson. He claims that Davidson has also put an emphasis on the

process of meaning-determination and has said that what a speaker means by her words depends on, or at least is partly determined by, the things in the world causing the speaker to apply those words. From this claim, he aims to conclude that "so my account is certainly true to at least one strand in Davidson" (2007, 164). This attempt, at most, would make his Externalist Explanation a rival interpretation of Davidson's account. This conclusion is weak enough to raise no problem for my claim that Davidson's account must alternatively be taken to be the Transcendental Explanation. However, in Sections 5 and 6, I showed why Child's Externalist Explanation cannot be treated as an acceptable interpretation of Davidson's account. I argued that not only is it an implausible explanation in general, but Davidson has also rejected it as an inadequate explanation of the meaning-asymmetry.

Child's second attempt is to claim that the presumption that speakers are not usually wrong about what they mean and believe cannot be supported by Davidson's remarks on the process of interpretation. According to him, in order to see whether this is true, we should see what it is about the process of interpretation that can sustain our presumption. It is worth noting that even if Child could show that Davidson's discussion of the process of interpretation fails to support the presumption that speakers are often right about what they mean, this by itself could not show that his Externalist Explanation is correct and correctly attributable to Davidson. Nonetheless, Child continues by claiming that Davidson's discussion of interpretation fails to offer such a support for our conceded presumption that speakers are usually right about what they mean *because* we have a way of ascribing meanings to the speaker which can violate Davidson's presumption. The alternative way Child has in mind is simply this: "interpret the speaker's words as meaning what those words mean on the lips of others in her community" (Child 2007, 165). This is just to say that, regardless of what the speaker intends to mean by her words, interpret those words as meaning what others mean by them, or as what the words standardly, normally or conventionally mean. However, one of the most significant parts of Davidson's later view of meaning, as discussed in Sections 4, 5 and 6, has been dedicated to resisting any view that takes following some rule, convention, norm, standard, institution, and generally any sort of socially agreed-on way of using words to be essential to the existence of successful communication. Considering Davidson's extensive remarks against such a view, attributing it to Davidson is completely implausible. Child himself immediately confesses that "[o]f course Davidson thinks that way of ascribing meanings would be wrong" (2007, 165). If so, why should he attribute this view to Davidson at all?

I think the reason why Child does so is that he aims to justify inferring the following conclusion from it: the aforementioned way of ascribing meanings to the speaker would be wrong, for Davidson, because the meaning of the words gets determined by the things causing the speaker to use them, or as Child puts it, the interpreter's so "ascriptions of meaning would be unacceptable in the light of other principles about meaning – specifically, the principle that the meanings of a speaker's words are determined by the nature of the things that normally cause her to hold those words applicable" (2007, 165). Let's pause for a moment and focus on Child's claim. His aim was to show that the Transcendental Explanation cannot support the presumption that speakers are often right about what they mean. The reason he offered was that we can ascribe meanings to speakers without using the presumption. Not only did I show, especially in the first part of Section 6 above, why Davidson thinks this cannot be done, but Child himself agrees that Davidson is against such a view. What reason, hence, does he provide to persuade us that the Transcendental Explanation cannot support the presumption? No new reason. He rather claims that, for Davidson, speakers are often right about what they mean by their words because what they mean by their words is determined by what causes them to use those words. This is just to restate that Davidson's account of the meaning-asymmetry is the Externalist Explanation, rather than the Transcendental Explanation. And, again, I already argued for why this reading of Davidson is implausible.

Child's last attempt, which is not new and independent, is to restate his previous claim by appealing to Davidson's use of the term "presumption". He says that Davidson does not use the term "assume", that is, Davidson does not say that the interpreter must assume that the speaker is right about what she means; rather, he uses the term "presumption" and what he means by it, according to Child, is that "there is no *guarantee* that a speaker is right about the meaning of a given word; but there is a *presumption* that she is. Seen this way, the presumption that a speaker knows what her words mean is part of the *explanandum* in the discussion of first person authority; it is not itself part of the *explanans*" (2007, 165). It is difficult to see how this claim by itself can add anything new to his previous claim and especially how it can be used as an argument against the Transcendental Explanation. Of course, as I indicated earlier in Section 6, Davidson's concern, from the beginning, has been to *explain* this presumption. He emphasizes that the existence of the presumption is intuitively conceded; the problem is to explain *why* it must exist. As he clarifies, "it remains to show why there must be a presumption that speakers, but not their interpreters, are not wrong about what their words mean" (Davidson 1984a, 12). Moreover, I emphasized that presupposing the meaning-asymmetry without

explaining it (and hence taking for granted what was promised to be explained) is one thing, presupposing the meaning-asymmetry in order to explain why it exists is another. How can one explain why something exists without presupposing that it exists? What I have argued for was that Davidson's explanation of why this presumption necessarily holds is his Transcendental Explanation, rather than the Disquotational or the Externalist Explanation. Given my arguments against the Externalist Explanation in Section 5, nothing new has been offered by Child that can stand against the Transcendental Explanation as Davidson's official account of self-knowledge. Nonetheless, there is still more to say about Child's claims.

Child agreed that, for Davidson, the interpreter cannot work without relying on the essential presumption, that is, by simply taking the speaker's words to mean what others mean by them in her speech-community and his reason was that doing so is wrong "in the light of other principles about meaning", specifically on the basis of "the principle that the meanings of a speaker's words are determined by the nature of the things that normally cause her to hold those words applicable". The question is what Child means by Davidson's "principles about meaning". If by such principles he had *the necessary conditions on the process of interpretation* in mind, he would accept the Transcendental Explanation as Davidson's explanation of the meaning-asymmetry. But, Child clearly does not intend to do so. He rather maintains that these are the principles about meaning-determination and insists that, for Davidson, the most important one of them is that the speaker's words have the meaning they do because of what cause the speaker to use them, which, for Child, forms the *foundation* of Davidson's explanation of the meaning-asymmetry. This leads to a further problem: it seems that Child blurs the crucial distinction between Davidson's remarks on the emergence of meaning and mental content and Davidson's discussion of the process of interpreting such content. This is a crucial distinction, failing to appreciate which would have serious destructive consequences. I will end this part by a brief discussion of one of such consequences.

Child's reliance on Davidson's discussion of the process of meaning-determination, which appears in Davidson's discussion of triangulation, and his insistence on taking it to form the foundation of Davidson's account of self-knowledge, would provide us with a further reason for why the Externalist Explanation is implausible and is not Davidson's official account of self-knowledge. Recall that Child's reason for why it is wrong to interpret the speaker's words simply as meaning what others mean by them was that, for Davidson, it is what causes the speaker to apply her words to certain things in the world that gives meaning to the speaker's

words. But, first of all, this reason fails to imply that the interpreter does not have to interpret the speaker's words as meaning what others mean by them. The reason is that Davidson's discussion of how meaning *emerges* and gets determined confines the speaker to mean just what others mean by the words, while Davidson's discussion of the process of interpretation frees the speaker from such a constraint. For if, following Child, we concede that it is simply what causes the speaker to respond in the way she does that gives meaning to her words, nothing can prevent us from concluding that the speaker's words mean what other members of her speech-community mean by them because, *ex hypothesi*, the same things which cause the speaker to apply her words have supposedly caused others to apply those words too. As Davidson argues for, especially in his famous paper "On the Very Idea of a Conceptual Scheme" (1974b), the speakers of a language share the same world, a common ontology: similar things in the world prompt them to respond in the way they do. In his discussion of the notion of triangulation, Davidson introduces an even more restricted condition: unless there are others who similarly respond to similar things in the world, as well as to each other's responses, their responses would have no chance to acquire any determinate meaning.[25] At this basic level of learning a first language, the triangulators' "innate similarity responses … – what they naturally group together – must be much alike" (Davidson 1992, 120). For, otherwise, the first creature will respond differently to what the second creature takes to be similar. Thus, as Davidson says, a "condition for being a speaker is that there must be others enough like oneself" (1992, 120). To be similar, in this sense, implies that the triangulators must respond by, for instance, "tree" to certain things – i.e., trees in the world – which they group together similarly. Now, this means that the interpreter, say, the teacher, can ascribe meaning to the learner's uttered words basically on the basis of the fact that the similar things which cause her to utter those words cause the learner to utter the same words. Thus, the learner's utterance of "table" means *table* apparently because the same thing causes both the teacher and the learner to utter "table", and the teacher ascribes meaning to the learner's utterance accordingly. Child himself admits that Davidson rejects this way of ascribing meanings.

[25] Introducing Davidson's detailed discussion of the notion of triangulation would go beyond the purpose of this paper. The reason why the creatures' responses can have no determinate meaning in this case is that no genuine disagreement between them can emerge and without making sense of such disagreements, that is, of the fact that there is a distinction between what is the case and what seems to be the case, error cannot emerge. As a result, we will be trapped in the Wittgensteinian paradox, that is, that whatever seems right to the creature would be right, no matter what it is. For a recent discussion of this and of Davidson's reading of Wittgenstein, see Hossein Khani (2019) and (2020b).

The problem is that the Externalist Explanation appears to be compatible with such a view, contrary to the Transcendental Explanation which takes the conditions on *the process of interpretation* to be fundamental. This forms another reason for why the Externalist Explanation cannot be considered as Davidson's explanation of the meaning-asymmetry: it allows for a way of ascribing meaning to the speaker that *violates* our fundamental presumption and Davidson by no means would allow for such a violation. Not only this, but Davidson also intends to provide the speaker with freedom of choice in the practice of meaning something by words, that is, Davidson wants to allow for the situations in which the speaker successfully means *chair* by "table" and her use of "table" would not be regarded as incorrect or meaningless. Davidson's remarks on the process of interpretation, contrary to his discussion of triangulation, grant such a freedom. The speaker is allowed to mean whatever she may by her utterance, provided that there is enough evidence for the interpreter to interpret the utterance in the way the speaker intended and in order for this to be possible, speakers must be treated as rational agents credited with knowledge of what they mean, believe and intend – unless, of course, there is evidence to the contrary. This is the reason why Davidson emphasizes that "meaning something requires that by and large one follows a practice of one's own, a practice that can be understood by others" (1994, 125). The problem with Child's claims considered above has its roots in failing to distinguish between Davidson's remarks on meaning-emergence and meaning-interpretation.

Davidson has carefully distinguished between mere dispositions to respond to (or mere ability to discriminate between) different things in the world in certain ways and making *judgements* about them. He takes the latter to be fundamental: "A creature does not have the concept of a cat merely because it can discriminate cats from other things … You have the concept of a cat only if you can make sense of the idea of … judging that something is a cat" (Davidson 1999a, 8). A parrot can be trained or conditioned to respond by "cat" whenever a cat is present. For Davidson, however, it would not be acceptable to claim that the parrot means *cat* by such sounds: to be caused to respond in a certain way to certain things in the world would not be enough for making the claim that the creature means or believes something; the creature should be capable of judging that such and such a thing falls under a specific concept, such as that of a cat. Once the creature comes up with such a rich set of concepts and propositional attitudes, it is then free to apply its words in whatever way it may, provided that its utterances remain interpretable. This allows for the situations in which the same object, e.g., an emerald, causes the interpreter to utter "green" to mean *green* and the speaker to utter "green" to mean *blue*.

The mere fact that the emerald causes the speaker to respond by "green" does not help the interpreter to interpret her utterance correctly, that is, to determine what the speaker intends to mean by it. The availability of evidence, the interpreter's familiarity with the speaker's habits, life, attitudes, and environment, as well as luck, wisdom and so forth, would help the interpreter to understand what the speaker intends to mean by "green"; otherwise, the speaker has simply failed to mean anything at all. All of this is possible only if the speaker can be viewed as knowing what she means, believes and indents. This is a necessary condition on the possibility of interpretation. The Externalist Explanation misses such a crucial distinction between meaning-emergence and meaning-interpretation.

As I tried to argue for in this paper, it is the Transcendental Explanation, rather than the Externalist or the Disquotational one, that is to be considered as Davidson's official explanation of self-knowledge and this explanation is compatible with Davidson's well-known remarks on interpretation, the role of evidence in explaining self-knowledge, as well as the important distinction between describing and explaining self-knowledge. Of course, this does not mean that Davidson's Transcendental Explanation is free from any problem.[26] It is also important to note that my discussion in this paper focused on Davidson's answer to the "why" question, rather than the "how" question, regarding self-knowledge. My aim has been to support Davidson's Transcendental Explanation of *why* the meaning-asymmetry – and thus the belief-asymmetry – must exist. The question as to "how" it is that speakers non-inferentially know what they mean and believe is an entirely different question. I believe Davidson is to provide an answer to this question too.[27] Nonetheless, insofar as answering the "how" question is concerned, the Transcendental Explanation does not suffer from an extra or special problem: all the explanations discussed in this paper face the same problem, that is, the problem of explaining how each speaker obtains authoritative knowledge of what she means and believes. Neither the Externalist Explanation nor the Disquotational Explanation could successfully explain *how* speakers know what they mean and believe. The reason is that if the source of the speaker's authoritative knowledge of what she means and believes is neither her ability to

[26] For some potential problems, see, e.g., Jacobsen (2009), Gallois (1997, Chapter 9), Shoemaker (1996, Chapter 3), Ludwig (1994) and Beisecker (2003).

[27] I cannot claim that Davidson's own works on self-knowledge can offer a straightforward answer to the "how" question. But, I believe that this question can be answered at least *on behalf of* Davidson via, for instance, interpreting his remarks on meaning and intention as suggesting a sort of judgement-dependent account of meaning and intention, along the lines proposed by Wright. Nonetheless, this would be the subject of a different investigation.

disquotationally specify the meaning of her words, nor the fact that certain things in the word typically cause her to use those words, the Externalist and the Disquotation Explanation would lack the resources needed for explaining how the speaker knows what she means and believes. The conclusion of my arguments, therefore, would be that if one is to investigate Davidson's explanation of self-knowledge, it is his Transcendental Explanation that should be the focus of such an investigation.

7. Conclusion

I argued that Davidson's official explanation of self-knowledge is his Transcendental Explanation, rather than the Disquotational or the Externalist one. Not only do such explanations fail to explain first-person authority, but Davidson himself has already rejected them as entirely implausible.[28]

References

Aune, B. 2012. "On Davidson's View of First-Person Authority." In *Donald Davidson on Truth, Meaning, and the Mental*,, edited by G. Preyer, 214-227. Oxford: Oxford University Press.

Beisecker, D. 2003. "Interpretation and First-Person Authority: Davidson on Self-Knowledge." *Southwest Philosophy Review* 19 (1): 89-96.

Bernecker, S. 2013. "Triangular Externalism." In *A Companion to Donald Davidson*, edited by E. Lepore and K. Ludwig, 443-445. Oxford: Wiley-Blackwell.

Blackburn, S. 1984. *Spreading the Word.* Oxford: Oxford University Press.

Byrne, A. 1998. "Interpretivism." *European Review of Philosophy* 3: 199-223.

Child, W. 1994. *Causality, Interpretation, and the Mind.* Oxford: Oxford Uniervsity Press.

Child, W. 2007. "Davidson on First Person Authority and Knowledge of Meaning." *Noûs* 41 (2): 157–177.

[28] I am grateful to Alex Miller for valuable comments on earlier drafts of this paper. I also thank Hamid Vahid and the audience members at IPM for helpful discussions. For constructive comments and suggestions, I would like to thank the editor and two anonymous referees for this journal.

Child, W. 2013. "Davidson on First-Person Authority." In *A Companion to Donald Davidson,*, edited by Ernie Lepore and Kirk Ludwig, 533-549. Malden, MA: Wiley-Blackwell.

Davidson, D. 1983. "A Coherence Theory of Truth and Knowledge." In *Kant oder Hegel?*, edited by D. Henrich, 423–38. Stuttgart: Klett-Cotta. In Davidson (2001b), 137-153.

Davidson, D. 1986. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E. Lepore, 433–46. Cambridge: Blackwell. In Davidson (2005), 89-108.

Davidson, D. 1974a. "Belief and the Basis of Meaning." *Synthese* 27: 309–23. In Davidson (2001a), 141-154.

Davidson, D. 1984b. "Communication and Convention." *Synthese* 59: 3-17. In Davidson (2001a), 265-280.

Davidson, D. 2001c. "Externalisms." In *Interpreting Davidson*, edited by P. Kot'atko, P. Pagin and G. Segal, 1-16. Stanford: CSLI.

Davidson, D. 1984a. "First Person Authority." *Dialectica* 38: 101–12. In Davidson (2001b), 3-14.

Davidson, D. 1984a. "First Person Authority." *Dialectica* 38: 101–12. Reprinted in Davidson (2001b), 3-14, to which page references apply.

—. 2001a. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press.

Davidson, D. 1978. "Intending." In *Philosophy of History and Action*, edited by Y. Yovel, 41–60. Dordretch: The Magnes Press.

Davidson, D. 1991. "James Joyce and Humpty Dumpty." *Midwest Studies in Philosophy* 16: 1–12. In Davidson (2005), 143-158.

Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441–58. In Davidson (2001b), 15-38.

Davidson, D. 1974b. "On the very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47: 5–20. In Davidson (2001a), 183-198.

—. 2004. *Problems of Rationality.* Oxford: Clarendon Press.

Davidson, D. 1973. "Radical Interpretation." *Dialectica* 27: 314–28. In Davidson (2001a), 125-140.

Davidson, D. 1982. "Rational Animals." *Dialectica* 36: 317–28.

Davidson, D. 1993b. "Reply to Bernard Thöle." In *Reflecting Davidson*, edited by R. Stoecker, 248-250. Berlin: de Gruyter.

Davidson, D. 1993a. "Reply to Eva Picardi." In *Reflecting Davidson*, edited by R. Stoecker, 210-212. Berlin: de Gruyter.

Davidson, D. 1997. "Seeing through Language." In *Thought and Language*, edited by J. Preston, 15–28. Cambridge: Cambridge UP. In Davidson (2005), 127-142.

—. 2001b. *Subjective, Intersubjective, Objective.* Oxford: Clarendon Press.

Davidson, D. 1999a. "The Emergence of Thought." *Erkenntnis* 51 (1): 7-17.

Davidson, D. 1992. "The Second Person." *Midwest Studies in Philosophy* 17: 255–67. In Davidson (2001b), 107-122.

Davidson, D. 1994. "The Social Aspect of Language." In *The Philosophy of Michael Dummett*, edited by B. McGuinness, 1–16. Dordrecht: Kluwer. In Davidson (2005), 109-126.

Davidson, D. 1975. "Thought and Talk." In *Mind and Language*, edited by S. Guttenplan, 7–23. Oxford: Oxford UP. In Davidson (2001a), 155-170.

—. 2005. *Truth, Language, and History.* Oxford: Clarendon Press.

Davidson, D. 1989. "What Is Present to the Mind?" *Philosophical Issues* 1: 197-213. In Davidson (2001b), 53-68.

Dennett, D. 1987 . *The Intentional Stance.* Cambridge, Mass.: MIT Press.

Gallois, A. 1997. *The World Without, the Mind Within: An Essay on First-Person Authority.* Cambridge : Cambridge University Press.

Glock, H. J. 2003. *Quine and Davidson on Language, Thought, and Reality.* Cambridge: Cambridge University Press.

Hacker, P. M. S. 1997. "Davidson on First-Person Authority." *Philosophical Quarterly* 47: 285–304.

Hossein Khani, A. 2020b. "Davidson's Wittgenstein." *Journal for the History of Analytical Philosophy* 8 (5): 1-26.

Hossein Khani, A. 2020a. "Interpretationism and Judgement-Dependence." *Synthese* 1-20. doi:https://doi.org/10.1007/s11229-020-02670-8.

Hossein Khani, A. 2019. "Kripke's Wittgenstein's Sceptical Paradox: A Trilemma for Davidson." *International Journal for the Study of Skepticism* 9: 21-37.

Hylton, P. 1990/91. "Translation, Meaning, and Self-Knowledge." *Proceedings of the Aristotelian Societ* 91: 269-290.

Jacobsen, R. 2009. "Davidson and First-Person Authority: Parataxis and Self-Expression." *Pacific Philosophical Quarterly* 90: 251–266.

Kripke, S. 1982. *Wittgenstein on Rules and Private Language.* Cambridge: Harvard University Press.

Ludwig, K. 1994. "First-Person Knowledge and Authority." In *Language, Mind and Epistemology: On Donald Davidson's Philosophy*, edited by Gerhard Preyer et al., 367-398. Kluwer.

Ludwig, K., and L. Lepore. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality.* Oxford: Oxford University Press.

Macdonald, C. 1995. "Externalism and First-Person Authority." *Synthese* 104: 99-122.

Picardi, E. 1993. "First-Person Authority and Radical Interpretation." In *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers,*, edited by Ralf Stoecker, 197-209. Berlin: de Gruyter.

Ryle, G. 1949. *The Concept of Mind.* London: The Mayflower Press.

Searle, J. 1987. "Indeterminacy, Empiricism, and the First Person." *The Journal of Philosophy* 84 (3): 123-146.

Shoemaker, S. 1996. *The First-Person Perspective and Other Essays.* Cambridge: Cambridge University Press.

Strawson, P. 1959. *Individuals.* London: Methuen.

Thöle, B. 1993. "The Explanation of First Person Authority." In *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers,*, edited by Ralf Stoecker, 213-247. Berlin: de Gruyter.

Wright, C. 1988. "Moral Values, Projection and Secondary Qualities." *Proceedings of the Aristotelian Society* 62: 1-26.

—. 2001. *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations.* Cambridge, US: Harvard UP.

—. 1992. *Truth and objectivity.* Cambridge, MA: Harvard UP.