



Speaking the Truth, Conventions, and the Principle of Charity

Daniel Arvage Nagase¹

Received: 15 February 2025 / Accepted: 23 September 2025 / Published online: 6 November 2025
© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract

In this article, I examine Davidson’s surprising argument, in his John Locke Lectures, that *speaking the truth* is what he calls an essentially disengaged act. I show he defends in these lectures a type of semantic conventionalism, and how he took this conventionalism to entail the essentially disengaged character of speaking the truth. I then show how his characterization of the Principle of Charity, in those lectures themselves, suggests a powerful argument against the essentially disengaged character of speaking the truth, and, thus, against semantic conventionalism. Finally, I examine how he could have failed to notice this in the lectures by examining his changing picture of communication and the role the Principle of Charity plays in such exchanges.

Keywords Davidson · Speaking the truth · Conventions · Principle of Charity · Interpretation

1 Introduction

Speaking is an action and, like all actions, it is subject to interpretation. Nevertheless, there seems to be something peculiar about speech, in that at least one of its aspects has its interpretation apparently forced upon us, namely the *meaning* of the person’s words. A reasonable explanation for this is that this aspect is somehow automatic, perhaps because there are *conventions* in force that all but dictate how we should understand the meaning of those words. This gives rise to an attractive picture of linguistic meaning, one in which such meaning is governed by social conventions, and it is the job of the semanticist to uncover those conventions and enshrine them in a semantic theory—perhaps, even, a formal axiomatic one.

As is well-known, Davidson (1984, 1986) is a fierce opponent of such a view, arguing that conventions are not essential for determining linguistic meaning.¹ What is less

well-known is that this was not always the case, and that, in fact, in his John Locke Lectures—recently published as *The Structure of Truth* (Davidson 2020)—, he argued precisely for a picture of linguistic meaning much like the one sketched in our opening paragraph: namely, that there is an aspect of linguistic meaning that *is* governed by conventions, namely the truth conditions of an utterance. Surprisingly, then, it turns out that the early Davidson was a conventionalist, and he believed that his axiomatic theories aimed at capturing the “*conventional semantic core*” (Davidson 2020, p. 26, his italics) of a language.² Why did he then change his mind?

In this paper, I will argue that the main reason is that he came to see more clearly the implications of his radical interpretation thesis, which was already incipient in the last of his John Locke Lectures. In particular, whereas in the first lecture Davidson argued that no interpretation was necessary in order to determine whether a person spoke the truth, and thus that this speech act was entirely conventional, there are already elements in his last lecture that undermine this claim.

Accordingly, this paper is structured as follows. In the next section, I will reconstruct Davidson’s reasoning for claiming that the determination of whether or not a person is speaking the truth needs no interpretation. There, I will show that this reasoning rests on an important insight

¹ That is not to say that they are not relevant. Davidson’s point is that the essential core of linguistic communication is not captured by conventions. Understanding this will be one of the main aims of this paper.

✉ Daniel Arvage Nagase
dan.nagase@gmail.com

¹ Universidade de São Paulo, Dept. de Filosofia, São Paulo, SP, Brazil

² Both Rumfitt (2021) and Verheggen (2022) note the surprising character of Davidson’s defense of conventionalism in their respective reviews of the John Locke Lectures.

regarding the rational structure of our actions and the way this structure interacts with conventions. This insight is retained in the later Davidson, except that he will extract different consequences from it. Whereas in the John Locke Lectures he took it to mean that speaking the truth is not an act that needs interpretation, later on he will defend that speaking the truth is not a conventional affair at all.

I then proceed to reconstruct his arguments in the last lecture in favor of the Principle of Charity, and I try to show that the connections he there establishes between, on the one hand, belief and meaning and, on the other hand, belief and truth, cast serious doubts on the claims developed in the first lecture. Finally, I show that some of his assumptions regarding successful communication and the locutionary dimension of speech acts were a blind spot that prevented Davidson from solving this conflict. Once he recognized that the Principle of Charity made conventions otiose, the way was open for the abandonment of his conventionalism.³

My aims here are not merely exegetical, however. By examining the twists and turns of Davidson's relation to the conventionalism described in the first paragraph, I believe that we can extract important insights regarding the nature of meaning, at least if we assume, as does Davidson, that the meaning of an utterance is given by its truth conditions. In particular, each section will reveal a specific insight: the first, that there is a tight, inverse connection between conventions and the need for interpreting an action, the second, that linguistic meaning is also subject to interpretation (in Davidson's sense), and the third, that the Principle of Charity explains the apparent convergence of meaning required for successful communication. I will then conclude by making some suggestions about a possible justification for the Principle of Charity.

2 Speaking the Truth as Disengaged

In the introduction, I mentioned that, in the John Locke Lectures, Davidson thought that a theory of meaning for a language would capture the "conventional semantic core" of the language (Davidson 2020, p. 26). But what kind of conventions are we dealing with here? Since they are *linguistic* conventions, they must be conventions about speech acts. These linguistic conventions are of a peculiar sort, however. They are not conventions about the acts themselves, *per se*,

but rather conventions about the *signals* that we use to indicate the acts. In other words, they are conventions about the *linguistic signs*, the words and expressions we use to indicate our speech acts. Indeed, Davidson at the time believed that there was a special kind of situation here, in which the mere presence of such linguistic signals *was enough to guarantee that a determinate speech act took place*, irrespective of the wider context in which this act was realized. This is surprising, because it seems to imply that one could have an act without the relevant intention to produce the act. Let us try to unpack how Davidson arrived at this idea.

The first, important distinction to bear in mind here is Austin's distinction between locutionary and illocutionary acts.⁴ Like Austin, Davidson isolates the locutionary dimension of a speech act both from the illocutionary *force* that attaches to the act and the perlocutionary *effects* that follow from it. Obviously, as both Austin (1962, p. 147) and Davidson (2020, p. 100) are quick to point out, these dimensions are abstractions; it is not like one can have a locutionary act without an illocutionary force or perlocutionary effects.⁵ Nevertheless, this distinction is important because the conventional aspect of the utterance will be related merely to the locutionary, not the illocutionary dimension of the act.⁶ To show why the locutionary dimension is special in this way, Davidson will first try to get clear on what the relevant *intentions* behind the acts are. This analysis of the intentions behind the acts will prove to be decisive for the argument developed in this paper, so let us see it in some detail.⁷

⁴ For the distinction, cf. especially Austin (1962, Lecture VIII).

⁵ In a later article, Davidson explicitly claims that "linguistic utterances always have an ulterior purpose", in the sense that "one must always intend to produce some non-linguistic effect through having one's words interpreted" (Davidson 1984, p. 272). I will come back to the relations between this later article and the John Locke Lectures in Section 4.

⁶ He thus appears to be deviating from Austin, who took it that the illocutionary force of an act could be analyzed in terms of conventions. Cf. Sbisà (2002, 2020) for a nuanced discussion of this aspect of Austin's analysis.

⁷ Davidson says that his discussion of speech acts in terms of the speaker's intentions "more or less follow[s] Austin" (Davidson 2020, p. 100), but, in fact, I believe the discussion "more or less" follows Grice, not Austin. Compare Grice (1957) and Austin (1962). It is telling that, when discussing the John Locke Lectures in his intellectual autobiography, Davidson mentions Grice, and not Austin, as being a major influence: "Here Grice had been an enormous help in devising convincing ways for telling when common usage merely took advantage of an underlying meaning and when it revealed that meaning more directly" (Davidson 1999a, p. 49). Note that this is precisely what is at issue here, namely the characterization of linguistic meaning. Cf. Avramides (2001) for a very interesting comparison between Grice and Davidson that is particularly germane to this essay (do note that Avramides is there concerned with *late* Davidson, whereas I am here suggesting that the Gricean influence may have taken its roots far earlier, already in the John Locke Lectures.)

³ I will not go into the broader implications of this abandonment for Davidson's philosophy, in particular his defense of an axiomatic truth theory as a theory of meaning, though I do agree with Camp (2016, esp. pp. 120–1) that this abandonment does make this defense much less tenable (as will become clear in the course of the paper, however, I believe this should push us in the direction of what she calls "semantic skeptics" instead of in the direction of conventionalists). I hope to treat these issues in a companion paper.

Davidson's key idea can be gleaned from the following quotation:

Thirdly, we may describe his action in terms of his actual words and the intentions necessary to uttering a particular sentence with shared truth conditions as an act of communication. For example: he spoke the words "A train is approaching" knowing that Gertrude knew that this sentence was true if and only if a train was at that moment approaching; by doing this he informed her that a train was approaching; and in doing this he intended to get Gertrude to move out of the path of danger. (Davidson 2020, p. 101)

The last two sentences after the semi-colon are clearly meant to evoke the illocutionary force and the perlocutionary effect of the act, and what is interesting about them is that they are characterized by intentions that go beyond the linguistic realm. That is, I can only assess whether or not you intended to *inform* Gertrude and *get her out of danger* by considering the wider context of the act (as we will see, this means that I need to *interpret* you in Davidson's sense of the term). This is obvious with respect to the effect I'm trying to achieve (get her out of danger), but Davidson (2020, p. 102) argues that even the *force* with which I'm uttering my words should be so characterized, as this force can be characterized without mentioning the *specific* linguistic intentions of the speaker.

According to Davidson, precisely because we must attend to the context of the act, it is impossible for there to be a conventional device associated with its illocutionary dimension. To see why this is the case, consider the argument Davidson gives in the following paragraph:⁸

Suppose, contrary to fact, that there were a conventional device that represented the speaker as making an assertion: let us say, the speaker prefaces an utterance with the words "I hereby assert that"; he says this in a deep and sonorous tone of voice, and crosses his heart as he speaks. Now that we know the device, however, what on earth can prevent some clown from putting on the assertion-skit as a joke, as nothing but a skit? In which case we no longer have an assertion. It is no help to add that part of what is required is that the speaker must be *sincere*. That's just the point; sincerity can't be part of a convention. Convention

is a substitute for sincerity, but cannot depend on it. (Davidson 2020, p. 22)

The discussion in the quoted passage is about the specific act of assertion, but Davidson clearly thought the argument had general import, characterizing all illocutionary actions. As this argument will be very important for our purposes here, it will be useful to have it schematized. Assessing it will also help to get in focus the import of Davidson's claims in the lectures.

- P1. Suppose that there is a conventional device for indicating a specific illocutionary action;
- C1. Then, for any context, the presence of this device indicates that the utterance is an instance of that kind of action;
- P2. But we can always find a context in which the utterance is not an instance of that kind of action;
- C2. So, in those contexts, the presence of the device cannot indicate that the utterance is an instance of the kind of action in question;
- C. Therefore, there can be no such device.

It seems clear that the crux of this argument rests on P2, since P1 is the assumption being targeted for reduction, and C1 and C2 are merely intermediary conclusions that extract obvious consequences out of P1 and P2. Thus, the question is: can one really imagine, for any possible convention for expressing a determinate illocutionary speech act, a context that negates it?⁹

It is not difficult to see why Davidson believed there to be such contexts. Let us consider the particular example Davidson cited in the above passage, namely in which a purported sign for indicating assertion was used in a pretense context, say, as he suggests, in the middle of a skit. In such a context, it is not part of my goal to make an assertion. Rather, my goal is to *fake* an assertion, perhaps as part of a larger goal to entertain my audience—and the audience knows this. We can see that an assertion has not been made because of the inappropriateness for an audience member to actually take my utterances *as* assertions—for example, by challenging me on the existence of a person named "Godot", for whom I am waiting, or trying to argue with me on some point, etc. Davidson's point, although made more vivid by pretense contexts, is obviously not restricted to them: any supposed

⁸ The argument from this paragraph reappears in a late essay "Mood and Performances" (Davidson 1979, p. 114), which shows that he never abandoned this aspect of his analysis of illocutionary force. Interestingly, Davidson also raises there the question about whether there is a convention about speaking the truth, but declines to discuss it in that article (Davidson 1979, p. 112).

⁹ Note that the claim is: for any utterance and kind of speech act, there is a context in which the utterance is not a species of that act. It is not: there is a context such that, for any utterance and speech act, the context makes it such that the utterance is not an instance of the act. That would only be the case if Davidson were committed to the thesis that there were contexts in which utterances could be made that were not speech acts at all, and, as far as I can see, he is not here committed to that thesis.

conventional device to indicate a speech act could be *misused* in order to do something else. This is not an accident either. It is a consequence of the fact that part of what makes an action the action it is is the *intention* with which it was made, and that the determination of an intention must take into account the teleological structure of the action, the way it is interacting with the agent's goals, beliefs, and further actions.¹⁰ In other words, given the intrinsic connection between the determination of a type of action, the intention with which it was made, and the wider context in which it took place, there can be no conventional device whose presence guarantees that an action took place.¹¹ Such a device would obviate the need to look for the rational structure of the action, but this, as we have just seen, is impossible. So there can be no such device.¹²

The immediate consequence of this discussion is that, if there is a kind of conventional signaling device that *always* indicates that a type of action has been performed, this type of action must be so that its unity is not derived from an intention. Now, if the considerations above are sound, we should be suspicious of any such claim, as it would run afoul of the connection established between intentions and actions. Still, it will be worth to see why Davidson by then believed that there was a special action—*speaking the truth*—which satisfied this condition, since this will help us to illuminate, in the next section, the tight connections between the rational structure of a speech act and its locutionary dimension.

Davidson's central claim about speaking the truth is that it is “essentially disengaged”:

Asserting, commanding, promising, and questioning are not good subjects of investigation if we want to study what is peculiarly linguistic, because

¹⁰ This is connected to Anscombe's (1963) well-known thesis that intentions give the *formal* cause of the action, that is, the distinct kind of unity that makes it the action it is, which determines *what kind* of action has taken place. Cf. Frey (2019) for an insightful commentary. Many thanks to Eduardo Marchesan for making me realize the importance of Anscombe's account of intention in this context and in the philosophy of language more generally.

¹¹ There are complexities here about whether my intention is enough to secure which speech act I made, or if there need be uptake on the part of an audience. Since the latter would only impose a stronger condition on the unity of the act, I need not consider this complexity here.

¹² Green (1997) argues that Davidson failed to establish his thesis, since, in pretense contexts, no speech act has been made, and the conventionality of linguistic devices applies only in cases where *some* speech act has taken place. I am unconvinced, for two reasons. First, what matters here is precisely whether or not the mere presence of a device can indicate that a speech act has taken place, and, if no speech act has been made, then, of course, the presence of the device cannot guarantee that a *specific* act took place. Second, it is not clear to me that in pretense contexts *no* speech act has been made—say, narrating or joking.

in performing such actions we necessarily operate beyond the reach of the conventions of language. Speaking the truth, on the other hand, is specially suited to systematic study just because the conditions for speaking the truth are a matter of linguistic convention. What explains this possibility in part is that speaking the truth is essentially disengaged: it is not done for its own sake, nor, typically, for the sake of any other end; we often do not know it has been done, and do not care. (Perhaps it is obvious on reflection that systematic theory of meaning must be based on a feature of speech behavior that is disengaged in this way.) (Davidson 2020, p. 23)

Let us start with the following definition:

Essentially disengaged: A linguistic act is “essentially disengaged” iff speakers do not *engage* in it, i.e. it is not generally a part of the speakers' projects to perform the act and they do not care whether the act is done.

This may seem like a peculiar definition, but the above discussion should help to put it into its proper context. The point seems to be this. Let us call *interpretation* the activity of determining how a given person's action falls into a pattern of rational behavior.¹³ In case of “essentially engaged” activities, in order to correctly assess whether or not one is performing the activity, one is required to assess the broader projects that these activities serve. In other words, in the case of “essentially engaged” activities, one must assess the *teleological structure of the context of the activity* in order to determine if one is performing it or not. Hence, these activities are essentially caught up with the *intentional, rational structure of the action*, and thus require *interpretation* in order for one to determine whether or not an agent is performing the activity in question.¹⁴ Notice how the connection between *interpreting a person's actions* and *assessing*

¹³ This is neutral on whether interpreting another requires merely the attribution of belief-desire pairs, or also intentions as *sui generis* states, or something more demanding, for instance something that makes me see what is it that you consider good about the desire. For the last option, see Brewer (2009).

¹⁴ This is independent of larger controversies surrounding the notion of intention, e.g. if it is reducible to other states, as the early Davidson (1963) would have it, or if it is a *sui generis* state, as Bratman (1987), and, apparently, late Davidson (1978) would have it (cf. Lepore and McLaughlin (1985) for a discussion of Davidson's apparent change of heart regarding this issue). It is also irrelevant whether we should interpret intentions in terms of planning or as basic; what is important is that any intentional action can only be interpreted against a larger teleological context which gives it its peculiar structure. Cf. Moran and Stone (2011, pp. 53ff) for more on the distinctive teleological structure of intentional action—Moran and Stone, are, of course, commenting on the pioneering work of Anscombe (1963).

its *rationality* is already implicit in this idea. In the next section, I will show how this connection was made explicit in the last of Davidson's lectures; for now, let us continue to examine the above argument.

In claiming that speaking the truth is essentially disengaged, Davidson is claiming that *there needs to be no interpretation to determine whether or not one has spoken the truth, because the act of speaking the truth is not generally tied up with the rational structure of one's actions*. What the above argument shows, then, is that if we are to take the expressions of a language as conventional devices for indicating truth conditions, in such a way that a person can speak the truth merely by uttering the relevant expressions, then we must take speaking the truth to be an essentially disengaged action, since, otherwise, one must look beyond the device to the wider context in order to assess which act has been performed. This is the heart of the matter. To emphasize its importance, I will dub this principle **Davidson's First Insight**:

Davidson's First Insight: In order for a conventional device to be sufficient to determine a type of linguistic action, this type of action must be essentially disengaged. Conversely, if a linguistic action requires interpretation in order to determine its type, then it is not merely conventional.

If Davidson's First Insight is correct, the only way for a conventional theory of truth to be feasible is if there is an act that is essentially disengaged. Let us thus see how he struggles to find such an act in the lectures, as this struggle will, I hope, provides us with broader lessons regarding the feasibility of conventionalism.

It will be useful, then, to go back to the locutionary dimension of the act and the intentions that characterize it. According to Davidson, a speaker utters a specific sentence *knowing that the audience knows its truth conditions* and that they will therefore get into the appropriate state (e.g. of being informed or alarmed) because of this knowledge.¹⁵ In other words, Davidson here characterizes the necessary intentions for performing a locutionary act in terms of *shared knowledge of truth conditions*. I know that you know the truth-conditions for a certain sentence, say "A train is approaching"; I therefore form the intention to inform you that a train is approaching by exploiting such shared knowledge. That this knowledge is shared among the participants in a conversation is important, since it implies that the association between the sentence and its truth-conditions takes hold *before* the speaker chooses to utter it, and thus *independently* of her communicative intentions. Putting it

picturesquely, the idea is that I have a communicative intention, for example to inform you of a fact, and then choose as means a sentence that is *already* endowed with the meaning I require.

Since the association between a sentence and its truth-conditions obtains *independently* of any particular speaker's communicative intentions, there is a disconnect between the "direction", so to speak, toward which a sentence points—its truth—and the goals of the speaker. That is, according to this analysis, the sentence *intrinsically* aims at something that may not be the aim of the speaker. The sentence *intrinsically* aims at truth as a result of the conventions governing the language, and because the sentence *intrinsically* aims at truth, to utter is, *ipso facto*, to utter a linguistic item with those truth-conditions. So, by uttering a sentence, I'm automatically indicating that I had the intention to utter a linguistic item with those conditions, so that my locutionary intentions are *transparent* in a way that my other intentions are not.¹⁶ Unsurprisingly, given its connection with truth, Davidson calls this locutionary act the act of *speaking the truth*.¹⁷

Now, what endows the utterance with those truth-conditions are the conventions of the language, which constitute what *speaking the truth is*. So we have conventions that determine when speaking the truth obtains, but such conventions are, in a sense, normatively inert, as they do not classify such obtaining as a goal to be achieved. That is why Davidson (2020, pp. 22–3) is at pains, in the lectures, to emphasize how different speaking the truth is from winning a game.¹⁸ Whereas, like winning at a game, whether one speaks the truth is, according to early Davidson, constituted by certain conventional rules that define the activity in question, *unlike* winning a game, these rules do not point to a goal to be achieved by the participants in a linguistic interaction. Notice that this need not stem from any pessimism on Davidson's part on the widespread sincerity of our peers. It only involves the recognition that speech involves much more than merely informing others, that these other goals are generally what is immediately on the speaker's mind,

¹⁶ Many thanks to the two anonymous referees for helping me clarify what exactly Davidson was getting at here.

¹⁷ Importantly, speaking the truth here *should not be understood in its illocutionary sense*. Indeed, Davidson is pretty clear that he is using the phrase in an unorthodox, technical sense: "Up to this point I have been trying to ignore the plain fact that the expression 'Speaking the truth' is by no means unambiguous, and only one of its meanings is appropriate to these remarks and claims. In fact, without gloss the phrase will not take the weight I have been putting on it." (Davidson 2020, pp. 23–4).

¹⁸ Although he does not mention him in the lectures, it is clear that the target here is Dummett (1959). Indeed, in a later text that is closely connected to the lectures, Davidson (1984) explicitly mentions Dummett in the context of a discussion of the game analogy.

¹⁵ Again, shades of Grice (1957). Notice that this also means that the formal theory is thus meant to capture the *explicit knowledge* that the speaker has of the truth conditions of the sentences in question.

and that they generally do not require that what the speaker has said is true in order for them to be realized.

The same pretense context that we used to motivate Davidson's argument against the possibility of a conventional device for indicating illocutionary force can help us to understand the argument here. I can be part of a pretense and still, inadvertently, speak the truth (perhaps because I didn't know better). For example, I may be part of a stage play and say "Davidson was an American philosopher" as a line. Since this was part of the play, I do not undergo any commitment to the truth of this utterance, and my audience may not believe this utterance was true (perhaps they are all under the delusion that Davidson was British). Nevertheless, I have still spoken truly. This suggested to Davidson that this type of sentence—"Davidson was an American philosopher"—was such that it could realize its aim independently of whether anyone was trying to do so. Now, it is important to note that the actor had as his intention to utter a sentence with those *truth-conditions*, that is, a sentence that is true if, and only if, Davidson was an American philosopher, and that such truth conditions were part of a *shared knowledge*, in such a way that his intentions were transparent to the audience. But whether the sentence was in fact *true* is irrelevant to his goals, which could be simply to do his part in a play, or entertain the public, or achieve fame, or whatever. That is, again, it is essential to his intentions that the utterance possess those truth-conditions, but it is not essential to his intentions that the truth-conditions *obtain*.

In contrast, then, to the other dimensions of a speech act, Davidson believes that "speaking the truth" can be characterized in *isolation* from the broader context of the act of uttering a sentence. And the reason, then, that speaking the truth can be so characterized is because the association between a sentence and its truth conditions is forged before the speaker utters it and independently of his intentions, so that it is irrelevant for the assessment of the meaning of a sentence what this broader context is. Or so Davidson would argue in the first lecture. Unfortunately, for Davidson's project, his last lecture casts serious doubt on whether that is really the case, for the Principle of Charity inextricably binds truth to rationality. If that is right, then there is no escaping the need for interpretation when assessing the meaning of someone's utterance. But, then, the association between a sentence and its truth conditions is not really given *in advance* of its utterance, and, so, conventionalism cannot get off the ground. Or so Davidson would argue in his later works. Before getting to that, let us, in the next section, analyze the connections between meaning and rationality.

3 Interpretation and the Centrality of Truth

We saw in the last section how Davidson claimed that speaking the truth was "essentially disengaged", that is, whether or not a person was speaking the truth is something that could be ascertained without needing to *interpret* that person. In this section, I would like to show that, by his own lights in the John Locke Lectures, this cannot be so. As Davidson himself argues in the last of his lectures, to ascertain whether or not someone has spoken the truth *is* to fix on an interpretation of this person, that is, to locate his utterances in a rational pattern of action. Indeed, truth is so central to the act of interpretation that if I actually determine that most of the person's utterances are false according to my theory, then I have not interpreted the person at all. In other words, to consider most of what the person speaks to be false is precisely to *fail to make sense of the person*, to discern a rational pattern in her utterances. Let us see this in a bit more detail.

To bring his point home, Davidson asks us to consider the following fictional case. Suppose I am faced with a person who begins most of her utterances with a nod, which I interpret as expressing denial. The person nods and says that there is a cloud passing before the sun. Since I interpret the nod as denial, I interpret this utterance as saying that there is *no* cloud passing before the sun—even though we can both plainly see that there is, in fact, a cloud before the sun. Similarly, soon after, the person nods and says that there is a tree in front of us, which I interpret as meaning that there is *no* tree in front of us—again, though we can both see the tree. Not long after, I am interpreting this person as denying that there is, in fact, a sun, or trees, or clouds. As Davidson then says:

Somewhere long before this it must dawn on me that his words do not mean what I took them to mean, nor *can* he have all the beliefs my system of interpretation foisted on him. This is not because my theory of his meanings is *unlikely* to be true; at a point soon reached, it is an unintelligible theory. The pattern of sentences to which my companion assents is all my evidence for what he means by those sentences, *and* for what he believes. Each time I take him as not accepting a sentence of his to which I have assigned truth conditions that would, I think, make it a true sentence, I subtract something from the base of agreement on which I must depend in making sense of his utterances, and of him. I can have him wrong about the cloud, but only if I have him agree that clouds occupy the sky, can look like whales, can eclipse the sun. More disagreement irresistibly invites the question, disagreement *about what?* (Davidson 2020, p. 109)

The passage quoted above makes two points that are important enough to be highlighted:

Interconnection between meaning and belief: “The pattern of sentences to which my companion assents is all my evidence for what he means by those sentences, *and* for what he believes”. In other words, there is a tight connection between truth conditions and beliefs. The means by which I determine the truth conditions of a person’s utterances are the same means by which I determine what this person believes—namely, the person’s assertions.

The cost of falsity attributions: “Each time I take him as not accepting a sentence of his to which I have assigned truth conditions that would, I think, make it a true sentence, I subtract something from the base of agreement on which I must depend in making sense of his utterances, and of him”. That is, to attribute falsity to a person’s utterances comes with a *cost*: the person appears *less rational* than she would otherwise be, at least by my lights, since I’m attributing a *disagreement* to us. At the limit, the person appears as simply *unintelligible* to me, as not rational at all.

The first point is important because it shows that the attribution of truth conditions to a person’s sentences is not a merely theoretical exercise. It is connected to *that specific person’s beliefs*. This already shows that there is some tension here between Davidson’s conventionalism and the interpretation task: if the meaning of a person’s utterances is inextricably bound up with her beliefs, then there is a trade-off between assigning the person deviant beliefs *or* deviant meanings. Hence, when interpreting another person, I may be authorized in assigning non-conventional meaning to her utterances if the end result is a theory that makes better sense of her behavior. But this means that conventions are not sufficient for the assignment of meaning anymore. Rationality considerations must also be taken into account.

This is made explicit in the second point. The reason why I am justified in attributing deviant meanings to a person if this minimizes deviant beliefs is that attributing deviant beliefs is *costly*. When I attribute a deviant belief to a person, I am at the same time attributing to the person a belief that is, by my lights, false. But this is not just an idle classification. As Davidson makes clear, this means that I *disagree* with the person, and the disagreement is substantial: the person is, by my lights, *wrong*. Truth is, therefore, a *norm* for our beliefs.¹⁹ If we can agree to call our epistemic norms

rational norms, then it is clear that, in attributing a false belief to a person, I am thereby attributing some measure of irrationality to her. That is why, if my goal is to make *rational sense* of her, I should attribute as few false beliefs to her as possible. This is precisely what Davidson calls the principle of charity:

In constructing a theory of truth for an alien tongue, the general policy, I have suggested, is *charity*: we should choose a theory that makes the speaker, or the speakers, of that tongue wrong as little as possible—by our own lights, of course, since we have no others to go by. “Wrong as little as possible” is pretty vague, and must be interpreted in terms of our remarks on plausible error. (Davidson 2020, p. 108)

Let us recap. In the last section, we saw how Davidson had mounted an argument to the effect that, if a certain action type needed to be placed in a rational pattern in order for one to determine whether or not it had been performed, then that action type could not be essentially conventional. Thus, Davidson concluded, if speaking the truth is conventional, that is because it is essentially disengaged, that is, one need not interpret another person’s actions to determine whether or not that person spoke the truth. What has emerged in this section, however, is that matters are not so simple. Since truth is a norm for rationality, it follows that I must try to minimize attributing false beliefs to a person. In this sense, Davidson’s remarks about “plausible error” are revealing; they show, precisely, that, in deciding whether or not the person has said something false, I must *place her utterance inside an overall pattern*, to see if there is a plausible, that is, a *rational* story to be told of how she came to that error. If there is no such story, the interconnected character of belief and meaning implies that I should look rather to *the meaning of her words* as a possible source of confusion: perhaps I have attributed to her utterance the wrong truth conditions—the wrong meaning—and she spoke truly after all. But then, the determination of whether or not she spoke truly is *not essentially disengaged*, since this determination must be carried out against the backdrop of the rest of her behavior. This is what I will call **Davidson’s Second Insight**:

Davidson’s Second Insight: The interconnected character of meaning and belief, together with the costly nature of falsity

for beliefs, a suggestion that is very close to what I’m arguing for here. Davidson’s denial, then, is particularly puzzling for my interpretation (I’m not the only one to be puzzled—see Glüer (2011, p. 139).) Cf. also Engel (2001; 2008) and Davidson’s (2001a, pp. 295-7) reply. Unfortunately I lack the space to solve this puzzle here, which would require an article of its own about the nature of epistemic norms for Davidson.

¹⁹ There is a complication. Davidson was particularly clear that “Truth is not, in my opinion, a norm” (Davidson 1999b, p. 461). In that context, Davidson was reacting against Engel’s (1999, p. 455-7) suggestion that the role played by truth in interpreting another is explained precisely by its close ties to belief, i.e. by the fact that truth is a norm

attributions, implies that the determination of the meaning of an utterance by an audience must take into account the whole rational pattern of the speaker's behavior. That is, the audience must *interpret* the speech act in order to arrive at its meaning.

I call this *Davidson's Second Insight* because, although he does not draw this conclusion in the John Locke Lectures—even though, as I emphasized, the elements for him to draw it are already there—he *does* draw it in later writings, particularly in “A Nice Derangement of Epitaphs” (Davidson 1986). Important for our immediate purposes, however, is the consequence that, if meaning is given, then, in terms of truth conditions, this implies directly that the truth conditions of an utterance can only be arrived at by way of interpretation. Hence the determination of whether or not a person is speaking the truth is also subject to interpretation. It is not, therefore, essentially disengaged, and thus cannot be determined solely on the basis of linguistic conventions, as Davidson thought in his first lecture.²⁰

4 Interpretation and the Principle of Charity

If the preceding analysis is correct, then we have a mystery in our hands: how could Davidson have missed the interconnection between interpretation and speaking the truth? Let us go back to Davidson's analysis of a communicative situation in the lectures. To simplify, let us consider a case in which I wish to inform you of something, say, to take Davidson's early example, that there is a train arriving at the platform. As we saw, Davidson claims that in such a case I must have the intention to utter a sentence such that I know that you know that it means that a train is arriving at the platform. But there is a subtlety here in the Gricean analysis that he employs. We can agree with Grice that I have the intention that you recognize my intention to utter a sentence with such and such truth conditions. And this *can* be taken to mean that you recognize my intention because we both know the truth-conditions of the sentence in question, so that, as Davidson's analysis in the lectures implies, the association between a sentence and its truth conditions must obtain *in advance* of the utterance and *independently*

of the particular speaker's intentions. Importantly, however, it *need* not be taken in this way. Perhaps you recognize my intention because of other factors, and *then* acquire the knowledge of the truth-conditions in question. *This* is, I think, the blind spot that prevented Davidson from recognizing how his account of interpretation made the appeal to conventions otiose: that the shared agreement on a sentence's truth-conditions could be the *result* of the interpretation process, and thus be *posterior* to the utterance itself. That this blind spot prevented him from reaping the fruits of his last lecture should be clear: if the participants in a dialogue must agree on the truth-conditions of an utterance *before* it is uttered, then, as we saw, nothing short of conventions would guarantee this outcome. And since, as we also saw, in order for these truth-conditions to be conventional, speaking the truth *had to be* essentially disengaged, we can see that there was here a stumbling block in Davidson's thought. It was only after he realized that the participants in a dialogue need not home in on the meaning of an utterance *in advance* that he was able to liberate himself from the conventionalist account.²¹

This interpretation is supported by a text that closely follows his discussion in the John Locke Lectures, but with the crucial difference that Davidson there does not argue anymore for the conventionality of speaking the truth, namely “Communication and Convention” (Davidson 1984). In that text, we find this interesting observation:

Is speaking the truth, in the sense of intentionally uttering a sentence that happens to be true, like winning? It is in this respect, that what it is to speak the truth is what a theory of truth seeks to define. *In so far; then*, as the truth conditions of utterances are known to speakers and interpreters in advance, and agreed upon as a condition of communication, speaking the truth has one of the features of winning. (I'll question how far this is true later.) (Davidson 1984, p. 268, original emphasis)

Note that Davidson appears to agree with his early account in the John Locke Lectures: speaking the truth is similar to winning the game, in that there are apparently norms or conventions that constitute this act—but with an important caveat: only in so far as the participants in a dialogue have to know these conditions in advance. So Davidson himself flags here the main difference between his early and his later positions. Unsurprisingly, when he returns to question this assumption later, we find that he considers it to be especially damning for the conventionalist approach:

²⁰ Notice that this point is conceded even by conventionalists like Armstrong (2016, p. 97). That is, according to Armstrong, although conventions can and do explain a range of features of our linguistic practice, they are not *essential* to such practices since successful communication can occur without them being in place. Rather, according to Armstrong, what is important is that both in the case of novel linguistic use and of conventional uses there is in place the same *mechanism* guiding our practice, namely “a suitably concordant system of mutual expectations” (Armstrong 2016, p. 97). I think Davidson would wholeheartedly agree with this latter claim.

²¹ Thanks to both anonymous reviewers for forcing me to clarify the nature of Davidson's blind spot in the lectures.

I have doubts, however, both about the clarity of the claim [that all linguistic activity involves regularities] and its importance in explaining and describing communication. The clarity comes into question because it is very difficult to say exactly how speaker's and hearer's theories for interpreting the speaker's words must coincide. They must, of course, coincide *after* an utterance has been made, or communication is impaired. But unless they coincide in advance, the concepts of regularity and convention have no definite purchase. (Davidson 1984, p. 278, original emphasis)

That this was the blind spot should not be surprising, as the Gricean approach strongly suggests the conventionalist assumption. After all, if I must get my message across by making my interlocutor recognize my linguistic intention, how could I achieve this if not for the fact that we are both party to a convention? Davidson himself, right after that last quoted passage, recognizes the difficulty:

Yet agreement on what a speaker means by what he says can surely be achieved even though speaker and hearer have different advance theories as to how to interpret the speaker. The reason this can be is that the speaker may well provide adequate clues, in what he says, and how and where he says it, to allow a hearer to arrive at a correct interpretation. Of course the speaker must have *some* idea how the hearer is apt to make use of the relevant clues; and the hearer must know a great deal about what to expect. But such general knowledge is hard to reduce to rules, much less conventions or practices. (Davidson 1984, p. 278)

This is the crux of the matter. Davidson here agrees that the speaker and the hearer must have a great deal of knowledge about each other (and in common) in order for communication to take place. *But this knowledge need not be linguistic.* This is what will liberate him from the need for conventions and thus allow him to grant the Principle of Charity its pride of place in his interpretation theory, which now, as the above quotation implicitly makes clear, will also extend to the meaning of our utterances, and thus to the locutionary aspect of our speech acts—so that, in a way, this aspect will regain its status as an *intentional action*.²²

To understand this, let us consider the Principle of Charity again. According to Davidson in the lectures, as we saw, this principle says that, when interpreting another, “we should choose a theory that makes the speaker, or the speakers, of that tongue wrong as little as possible—by our own

lights, of course, since we have no others to go by” (Davidson 2020, p. 108). Note that, by “minimizing mistakes” or “maximizing agreement”, Davidson is not simply talking about assigning the truth-value true to the greatest possible number of the person's beliefs, if that is even a coherent thought. Rather, he is talking about *making rational sense of the person*, that is, placing the person's thoughts and actions in a pattern that allows me to understand them. That is, the point of charity is not simply to assign the maximum number of truths to the person's beliefs, but, rather, to assign such truths *in order to understand the person*.²³ Later in his career, Davidson is particularly explicit about this:

Charity is a matter of finding enough rationality in those we would understand to make sense of what they say and do, for unless we succeed in this, we cannot identify the contents of their words and thoughts. Seeing rationality in others is a matter of recognizing our own norms of rationality in their speech and behavior. These norms include the norms of logical consistency, of action in reasonable accord with essential or basic interests of the agent, and the acceptance of views that are sensible in the light of evidence. (Davidson 1999c, p. 600)

Notice that both in the John Locke Lectures and the above quotation, Davidson is also explicit that we make rational sense of a person *by projecting our own rationality onto the other person*. That is why I try to maximize agreement: to project my rationality onto the other person is also to project my beliefs onto her, which simply *is* to maximize agreement, allowing, of course, for her different perspective given her own particular career throughout the world. Importantly, however, this projection is not limited to beliefs, values, and norms of consistency, such as those enshrined in rational decision theory. *It also includes beliefs about how words function in our language.* That is, when interpreting another person, I also project my own language use onto the person, thus taking her words as *I* would take them—again, subject to correction given the evidence, as discussed in the last section. Of course, the more I get to know a person, the more I learn about her own idiosyncratic language use, and the more I get to appreciate how much our languages differ—just as, in learning more about the person, I learn more about her particular beliefs and particular values, and how much they differ from my own. And that is precisely why the interpretation of an utterance may have to be given *after* the utterance, for the utterance itself is a piece of behavior

²² Many thanks to Carlos Mario Márquez Sosa for making me realize that the Principle of Charity made the appeal to conventions unnecessary.

²³ Some, such as Grandy (1973), think that this amounts to a change from a Principle of Charity to a Principle of “Humanity”, but I’d argue that this is only if one takes a rather uncharitable view of Davidson’s early formulations of his principle.

that helps me in understanding the *person*. As we saw in the last section, these differences, including linguistic differences, can only be spotted against a larger background of agreement.

I will get to what justifies this assumption in a moment. For now, just note how this makes the appeal to conventions otiose. Knowledge of conventions was supposed, on Davidson's first account, to ground our communicative intentions: I take it that you share in the same conventions as I, and therefore form the intention to communicate something by uttering a sentence that conventionally signifies that something. Similarly, in interpreting you, I take it that you share in the same conventions as I, and therefore interpret your words according to those conventions. According to the new model, however, I don't need to appeal to conventions. Rather, I simply (and perhaps automatically) take it that you mean your words *in the same way as I*, and understand you accordingly—again, subject to correction, as Mrs. Malaprop teaches us. Indeed, the fact that I'm able to correct my initial assumption of a shared meaning is precisely an indication that I'm *not* relying on a convention, but rather, on interpretation. More than that, conventions presuppose that there is a *commonality* of meaning among those sharing in the conventions. Davidson's account, however, shows that there is no need for such a commonality: sufficient overlap in the known facts is enough to guarantee smooth communication.²⁴

This last point is important because, if Davidson's account is correct, there *can be no such commonality of meaning*. Given that there is a blurring of the boundaries between knowing a language and knowing “our way around in the world” (Davidson 1986, p. 107), due especially to the intermingling of beliefs about words and beliefs about the world that is characteristic of the interpretation process, it follows that there is no way to separate my uses of a word—what I mean by it—from my worldly knowledge, in such a way that these uses are highly idiosyncratic.²⁵ What gives the impression of a commonality is that I start by *projecting* my own uses of certain words onto you, therefore creating the *illusion* of a shared language.²⁶ This illusion is itself sustained by there being enough overlap in meaning as to allow for communication, but it is still an illusion. The fundamental role of this projection—one form of the principle of charity—for communication is, I take it, his fundamental

insight from “A Nice Derangement of Epitaphs” (Davidson 1986). I will call this **Davidson's Third Insight**:

Davidson's Third Insight: Interpretation of another person's utterances is just a particular case of interpretation in general and it proceeds in the same way, viz. by *projecting* our own norms and values onto the other person, subject to correction as the process goes on.

This way of formulating the insight only makes more forceful the question: what justifies this projection? As is well-known, in his later writings Davidson developed an answer based on his “triangulation argument”.²⁷ In a nutshell, Davidson's point is that we do not develop our capacity to think in a vacuum; rather, we develop it in reaction to each other and to a shared environment. It is not, therefore, surprising that we end up with similar thoughts, as we are similarly endowed creatures reacting to a shared environment.²⁸

A Wittgensteinian way of putting the point is to say that what justifies this projection is the fact that we presume ourselves to share the same sensibility, the same form of life. Here, Davidson meets with Stanley Cavell, who is quite clear on these matters:

We learn and teach words in certain contexts, and then we are expected, and expect others, to be able to project them into further contexts. [Footnote suppressed] Nothing insures that this projection will take place (in particular, not the grasping of universals nor the grasping of books of rules), just as nothing insures that we will make, and understand, the same projections. That on the whole we do is a matter of our sharing routes of interest and fulfillment, of what is outrageous, of what is similar to what else, what a rebuke, what forgiveness, of when an utterance is an assertion, when an appeal, when an explanation—all the whirl of organism Wittgenstein calls “forms of life”. Human speech and activity, sanity and community, rest upon nothing more, but nothing less, than this. It is a vision as

²⁴ Hence why it is particularly well-suited to handle the phenomenon of homesign. For homesign, cf. the remarkable study of Goldin-Meadow (2003). Begby (2017) gives an interesting account of the phenomenon from a broadly Davidsonian perspective.

²⁵ This important point is emphasized by Begby (2016).

²⁶ I thus think that Davidson would be sympathetic to what Rey (2006, 2008, 2020) has called “folieism”, precisely the thesis that the notion of a shared language is a kind of sustained delusion (a “folie”) shared by those who take themselves to speak the same language.

²⁷ I have in mind especially Davidson (1991, 1992). As is equally well-known, Davidson's triangulation argument is extremely controversial. I will obviously not be able to touch on this controversy here. Cf. Myers and Verheggen (2016), as well as their contribution to this collection (forthcoming), for a compelling sympathetic reconstruction of the argument; I will comment on a specially relevant aspect of Verheggen's interpretation in the conclusion; Goldberg (2009) also offers an interesting perspective that I think is congenial to my own, especially with his emphasis on the reconciliation dimension of triangulation. Many thanks to the first anonymous referee for emphasizing the importance of triangulation in this context.

²⁸ Again, many thanks to the first anonymous referee for emphasizing the importance of a shared environment in constituting our thoughts as part of what justifies those projections.

simple as it is difficult, and as difficult as it is (and because it is) terrifying. (Cavell 1962, p. 48)

That is, what justifies our projecting our own uses of words (among much else) onto others is simply the fact that our lives are so similarly shaped—including by our living in the same environment—that we do share enough of a background both so as to make this projection a near enough approximation in most cases, and also as to allow us to correct these projections when things go awry.²⁹ This last bit is important because, as Cavell makes clear in the quotation above, nothing *guarantees* that our projections will bear out, that we *will* be able to interpret one another satisfactorily. That is, although our shared form of life—structured as it is by our triangulation with others—guarantees that we will often enough home in on the correct interpretation of each other, that fact alone is no guarantee that in a *particular, local* situation our interpretation will be correct. There is a kind of “generic” guarantee of extensive agreement, but this does not translate into *specific* agreement on a *specific* situation. This brings me to my next topic, for, in making such corrections one needs not—in most cases, will not—proceed alone: the other who we are interpreting will, often enough, help us in such a task.³⁰ Understanding this correction process, and the *responsibilities* it brings, will better help us to see the proper *normative* dimension of language, in so far as it is anchored in our *care* for each other.

5 Conclusion: The Foundations of Charity

In the course of analyzing Davidson’s first struggle with conventionalism, we have unearthed three important insights. The *first* insight pointed to how little the notion of convention explains when we are dealing with *rational, intentional action*. Of course, that is not to deny the importance of conventions when explaining actions; much of our behavior is indeed conditioned by, say, social conventions. But it *is* to deny that conventions *by themselves* can explain rational action; for that, we will *always* need to consider the general teleological structure of the action. The *second* insight points to how our *linguistic actions*, even in their locutionary dimension, are *also* rational, intentional actions, and therefore are subsumed under the considerations of the

first insight. This, in turn, forms a powerful argument against the role of linguistic conventions for the determination of meanings, and thus points to the conclusion that semantics cannot be founded on linguistic conventions. Finally, the *third* insight disarms the most obvious objection to this picture, namely that it makes communication impossible, by showing how we can reasonably enough approximate each others’ meanings by *projecting* from our own case and correcting such projections when they are shown to be mistaken. I would like now to suggest, by way of conclusion, what I take to be some broader implications of these ideas.³¹

One important theme emerging from the preceding considerations is that the determination of an utterance’s meaning requires interpreting against the overall pattern of the speaker’s behavior, since it involves consideration of the person’s *goals* in uttering whatever it is that she uttered. This may seem a daunting task, until we remember that, like all purposive behavior, our speech acts generally occur as part of large, *collaborative* endeavors—for example, sustaining a conversation. Communication is not, therefore, an all or nothing affair, but can be continually refined in the course of an interaction, as the parties in the conversation continually interact to clarify their meanings. Indeed, it may even be that this process of interaction leads the *speaker* herself to a greater realization of what she first meant, so that her meaning is actively construed by everyone involved.³² In this sense, the audience is not merely passive in the communication process, but plays an active part in the whole endeavor.³³

There has been too much emphasis in the philosophy of language on the supposed *conventional constraints* involved in communicating. This is only natural: if we want a *systematic* theory of meaning, we must hope that interpretation

²⁹ Notice that the appeal to forms of life, here, is not meant to ground the phenomenon of linguistic projection, in opposition to projection in general. Rather, it is meant to explain how we can *in general* project our beliefs into others as required by charity. That is, there is nothing specifically linguistic that requires appeal to forms of life—the worry applies to charity in general, so to speak. Thanks to the second anonymous reviewer for pressing me on that.

³⁰ This need for correction is exactly what Goldberg (2009) calls reconciliation, an apt word in this context.

³¹ These suggestions have important consequences for Davidson’s overall project of a truth-theoretic semantics, which I unfortunately lack the space to develop here.

³² This is one of the take home lessons, I believe, of Clark (1996). McDonald (2021) argues in this connection that the illocutionary force of the speech act can be retrospectively clarified or negotiated by the participants in a dialogue, but my point here is that the locutionary dimension should be considered in a similar light. In a remarkable study, Brewer (2009) argues that there are some activities in which we engage not because we have a clearly defined conception of their value, but precisely because we seek a greater understanding of this value. My argument here is parallel to this: sometimes, what we seek is precisely to achieve a greater understanding of our own words; perhaps psychoanalytic therapy could be conceived along those lines. See also Alshanetsky (2019) for the closely related notion of articulating a thought and Salje (2025, esp. chap. 6), in particular, for the idea of articulatory self-knowledge and how it can be achieved through conversation with others.

³³ This is also emphasized by Moran (2011) in reflecting upon Cavell’s work on skepticism. As he puts it, other people are not only other, but also people who “*do* have a say in the matter as to what shall count as being known and being understood” (Moran 2011, p. 135).

is sufficiently constrained to allow for the theory; and, in any case, interpretation itself would have a hard time if it were *wholly* unconstrained. Nevertheless, I believe that this emphasis has made us ignore how much such constraints are *set up by us* in the process of communicating, how much of it is not only creative, but the result of a *joint creation*.³⁴ As emphasized by Verheggen (2017, p. 154), this is part of Davidson's answer as to how our words—our thoughts—can have content at all. For Verheggen, what distinguishes the purported solitary thinker and those who are in a triangulation situation is precisely that, in the latter situation, the triangulators hold each other accountable, and thus must negotiate what they mean in a way that is answerable to themselves and to the world. Importantly, though this is not emphasized by Verheggen, the triangulating dimension introduces the notion of *responsibility*. For to be accountable to one another means that we are *responsible* one to the other, that we take up responsibility for what we say and what we intend to say,³⁵ and, more importantly in this context, that we also take responsibility for how we *interpret* one another (to interpret the other as a rational person is part of this responsibility). This is, I hold, the real constraint on interpretation: not supposed conventions, but the other and her intentions.

Here, again, we see how Davidson's thought can be complemented by Cavell.³⁶ For Cavell reminds us of “the threat, or the truth, of skepticism: that it names our wish (and the possibility of our wishing) to strip ourselves of the responsibility we have in meaning (or in failing to mean) one thing, or one way, rather than another” (Cavell 1984, p. 135). Richard Eldridge glosses this as follows:

Underlying both the skeptic's insinuating questions and the antiskeptical replies to them, then, is the wish not to have this responsibility, the wish for one's words instead to be wedded to the world apart from having to make them make sense to others, apart from having to face charges from them of not making sense. It is a wish that one should be whole and sufficient to oneself, not open to the readings and criticisms of others of one's efforts to make sense to them. (Eldridge 1989, pp. 195–6)

³⁴ In this sense, I echo here Rhee's (2005) complaint that the language game metaphor may encourage an undue focus on *rules*, thus making us forget that there are *people* communicating.

³⁵ Hence the importance of restoring the locutionary dimension to an *intentional act*.

³⁶ The two anonymous reviewers worried about the relevance of Cavell for my argument. I hope the following may help to clarify that. For a related discussion of Davidson, Wittgenstein, and Cavell, see Avramides' contribution to this collection (Forthcoming).

The conventionalist approach, according to this reading, is a convenient fantasy. It is the fantasy that the associations between sentences and their meanings are always already given, that there are rules such that we only need to follow them in order for our words to mean what we want them to mean, as if there were no need, precisely, for interpreting the other—the mere existence of the convention would suffice to make us comprehensible one to the other.³⁷ If the argument developed here is sound, then this is, as Davidson would surely agree, just a fantasy. In conversing with another, there is always a need not just for our *sense of occasion*, but also, and perhaps more importantly, for our *sense for the other*, so that we may, together, arrive at truth—and whatever other goals we may also have in communicating. This is, I believe, the real sense in which language is *social*: not in that it involves conventional straight-jackets, but in that it involves a collective endeavor in which each party is *responsible and responsive* to the other. This, then, can perhaps explain the true foundation of charity, the true foundation of our projection practices and our shared form of life: that we own up to the *responsibility* of understanding another person.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 88887.084535/2024-00. I'm extremely grateful for the participants in an informal study group on the philosophy of language in which I first presented this material, in particular Eduardo Marchesan, Carlos Mario Márquez Sosa, João Lucas Pimenta da Silva Pinto, Euclides Torres Ometto Stolf, Fernanda Birolli Abrahão, and Tales Rodrigues Noronha Silva. Thanks also to Anderson Luis Nakano and Herivelto Pereira de Souza for their careful reading and valuable feedback. I also had the good fortune of receiving two excellent anonymous referee reports that significantly improved the paper (or so I hope!). Finally, while it is customary to end these acknowledgements by excusing those thanked from any responsibility for the author's mistakes, I feel that Eduardo's and Carlos's roles were so fundamental for the ideas here developed that I cannot wholly excuse them. I hope they will find this a light burden to bear.

References

- Alshanetsky E (2019) *Articulating a Thought*. Oxford University Press, Oxford
- Anscombe GEM (1963). *Intention*, 2nd edn. Harvard University Press, Cambridge, Massachusetts
- Armstrong J (2016) Coordination, triangulation, and language use. *Inquiry* 59:80–112. <https://doi.org/10.1080/0020174X.2015.1115270>
- Austin JL (1962) *How To Do Things With Words*. Oxford University Press, Oxford

³⁷ John McDowell (1979, p. 61) states in a similar vein that “We recoil from this vertigo [spoken of in Cavell's passage quoted in the last section] into the idea that we are kept on the rails by our grasp of rules”. The metaphor of rails is particularly apt, since it induces the idea that there is something—the rails—*forcing* us in a particular direction, and thus exempting us of our responsibility.

- Avramides A (2001) Davidson, Grice, and the social aspects of language. In: Costenza G (ed) *Paul Grice's Heritage*. Brepols Publisher, Turnhout, pp 99–115
- Avramides A (forthcoming) Charity Goes Deep. *Topoi*
- Begby E (2017) Language from the ground up: a study of homesign communication. *Erkenntnis* 82:693–714. <https://doi.org/10.1007/s10670-016-9839-1>
- Begby E (2016) Deranging the Mental Lexicon. *Inquiry* 59.1:33–55. <https://doi.org/10.1080/0020174X.2015.1115276>
- Bratman M (1987) *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, Massachusetts
- Brewer T (2009) *The Retrieval of Ethics*. Oxford University Press, Oxford
- Camp E (2016) Conventions' revenge: Davidson, derangement, and dormativity. *Inquiry* 59(1):113–138. <https://doi.org/10.1080/0020174X.2015.1115277>
- Cavell SL (1962) *The Availability of Wittgenstein's Later Philosophy*. Reprinted in: Cavell SL (2002), pp. 41–67
- Cavell SL (1984) *Being Odd, Getting Even (Descartes, Emerson, Poe)*. Reprinted in: Cavell SL (1988) pp. 105–149
- Cavell SL (1988) *In Quest of the Ordinary: Lines of Skepticism and Romanticism*. University of Chicago Press, Chicago
- Cavell SL (2002) *Must We Mean What We Say? A Book of Essays*, Updated Edition. Cambridge University Press, Cambridge
- Clark HH (1996) *Using Language*. Cambridge University Press, Cambridge
- Davidson D (1963) *Actions, Reasons, and Causes*. Reprinted in: Davidson D, 2001b, pp. 3–19
- Davidson D (1978) *Intending*. Reprinted in: Davidson, D (2001b), pp. 83–102
- Davidson D (1979) *Moods and Performances*. Reprinted in: Davidson D (2001c), pp. 109–121
- Davidson D (1984) *Communication and Convention*. Reprinted in: Davidson D (2001c), pp. 265–280
- Davidson D (1986) *A Nice Derangement of Epitaphs*. Reprinted in: Davidson D (2005), pp. 89–107
- Davidson D (1991) *Three Varieties of Knowledge*. Reprinted in: Davidson D (2001d), pp. 205–220
- Davidson D (1992) *The Second Person*. Reprinted in: Davidson D (2001d) pp. 107–121
- Davidson D (1999a) *Intellectual Autobiography of Donald Davidson*. In: Hahn LE (ed) *The Philosophy of Donald Davidson*. Open Court, Chicago, pp 3–70
- Davidson D (1999b) *Reply to Pascal Engel*. In: Hahn LE (ed) *The Philosophy of Donald Davidson*. Open Court, Chicago, pp 460–461
- Davidson D (1999c) *Reply to Richard Rorty*. In: Hahn LE (ed) *The Philosophy of Donald Davidson*. Open Court, Chicago, pp 595–600
- Davidson D (2001a) *Comments on the Karlovy Vary papers*. In: Kotatko P, Pagin P, Segal G (eds) *Interpreting Davidson*. CSLI Publications, Stanford, pp 285–306
- Davidson D (2001b) *Essays on Actions and Events*, 2nd ed. Oxford University Press, Oxford
- Davidson D (2001c) *Inquiries into Truth and Interpretation*, 2nd ed. Oxford University Press, Oxford
- Davidson D (2001d) *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford
- Davidson D (2005) *Truth, Language, and History*. Oxford University Press, Oxford
- Davidson D (2020) *The Structure of Truth: The 1970 John Locke Lectures*. Ed. by Cameron Domenico Kirk-Giannini and Ernest Lepore. Oxford University Press, Oxford
- Dummett M (1959) *Truth*. Reprinted in: Dummett M (1978), pp. 1–24
- Dummett M (1978) *Truth and Other Enigmas*. Harvard University Press, Cambridge, Massachusetts
- Eldridge R (1989) *A Continuing Task: Cavell and the Truth of Skepticism*. Reprinted in: Eldridge R (2001), pp. 189–204
- Eldridge R (2001) *The Persistence of Romanticism: Essays in Philosophy and Literature*. Cambridge University Press, Cambridge
- Engel P (1999) *The Norms of the Mental*. In: Hahn LE (ed) *The Philosophy of Donald Davidson*. Open Court, Chicago, pp 447–459
- Engel P (2001) *Is truth a norm?* In: Kotatko P, Pagin P, Segal G (eds) *Interpreting Davidson*. CSLI Publications, Stanford, pp 37–51
- Engel P (2008) *Davidson on Epistemic Norms*. In: Amoretti MC, Vassallo N (eds) *Knowledge, Language, and Interpretation: On the Philosophy of Donald Davidson*. Ontos Verlag, Frankfurt, pp 123–145
- Frey JA (2019) *Anscombe on practical knowledge and the good*. *Ergo*, an Open Access Journal of Philosophy 6:1121–1151. <https://doi.org/10.3998/ergo.12405314.0006.039>
- Glüer K (2011) *Donald Davidson: A Short Introduction*. Oxford University Press, Oxford
- Goldberg N (2009) *Triangulation, untranslatability, and reconciliation*. *Philosophia* 37:261–280. <https://doi.org/10.1007/s11406-008-9171-3>
- Goldin-Meadow S (2003) *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language*. Psychology Press, New York
- Grandy R (1973) *Reference, meaning, and belief*. *J Philos* 70(14):439–452. <https://doi.org/10.2307/2025108>
- Green MS (1997) *On the autonomy of linguistic meaning*. *Mind* 106(422):217–243. <https://doi.org/10.1093/mind/106.422.217>
- Grice HP (1957) *Meaning*. Reprinted in Grice HP (1989), pp. 213–223
- Grice HP (1989) *Studies in the Way of Words*. Harvard University Press, Cambridge, Massachusetts
- Lepore E, McLaughlin BP (1985) *Actions, reasons, causes, and intentions*. In: Lepore E, McLaughlin BP (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Basil Blackwell, Oxford, pp 3–13
- McDonald L (2021) *Reimagining Illocutionary Force*. In: *The Philosophical Quarterly*. Advance Article. <https://doi.org/10.1093/pq/pqab065>
- McDowell JH (1979) *Virtue and Reason*. Reprinted in: McDowell JH (1998), pp 50–73
- McDowell JH (1998) *Mind, Value, and Reality*. Harvard University Press, Cambridge, Massachusetts
- Moran R (2011) *Cavell on Outsiders and Others*. Reprinted in: Moran R (2017), pp. 123–135
- Moran R (2017) *The Philosophical Imagination: Selected Essays*. Oxford University Press, Oxford
- Moran R, Stone MJ (2011) *Anscombe on expression of intention: an exegesis*. In: Ford A, Hornsby J, Stoutland F (eds) *Essays on Anscombe's Intention*. Harvard University Press, Cambridge, Massachusetts, pp 33–75
- Myers RH, Verheggen C (2016) *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*. Routledge, London
- Rey G (2006) *The Intentional Inexistence of Language—But Not Cars*. In: Stainton R (ed), *Contemporary Debates in Cognitive Science*. Blackwell, Oxford, pp. 237–255
- Rey G (2008) *In defense of folieism: replies to critics*. *Croat J Philos* 8:177–202. <https://doi.org/10.5840/croatjphil20088212>
- Rey G (2020) *Representation of Language: Philosophical Issues in a Chomskyan Linguistics*. Oxford University Press, Oxford
- Rhees R (2005) *Wittgenstein and the Possibility of Discourse*. Ed. by DZ Phillips. 2nd ed. Blackwell Publishing, Oxford
- Rumfitt I (2022) *Review of The Structure of Truth: The 1970 John Locke Lectures*, by Donald Davidson, *Mind* 131(523): 1027–1038. <https://doi.org/10.1093/mind/fzab019>
- Salje L (2025) *Saying What One Thinks*. Oxford University Press, Oxford
- Sbisà M (2002) *Speech acts in context*. *Language & Communication* 22(4):421–436. [https://doi.org/10.1016/S0271-5309\(02\)00018-6](https://doi.org/10.1016/S0271-5309(02)00018-6)

- Sbisà M (2020) Assertion among the speech acts. In: Goldberg S (ed) *The Oxford Handbook of Assertion*. Oxford University Press, Oxford, pp 159–178
- Verheggen C (2017) Davidson's Semantic Externalism: From Radical Interpretation to Triangulation. *Argumenta* 3.1:145–161. <https://doi.org/10.14275/2465-2334/20175.VER>
- Verheggen C (2022) Review of *The Structure of Truth: The 1970 John Locke Lectures*, by Donald Davidson. *Rev Metaphys* 75(3):590–592. <https://dx.doi.org/10.1353/rvm.2022.0003>
- Verheggen C, Myers R (Forthcoming) The Status and the Scope of the Principle and Charity. *Topoi*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.