Making Sense of Others: Donald Davidson on Interpretation

William Taschek

What follows is the text of a talk presented to the Fellows Seminar of the University of Michigan Institute for the Humanities during the fall term of 1988. It was aimed at an audience of non-philosophically trained humanists from various disciplines. Its intention was primarily to bring (without too much distortion or technicality) a few of the more central strands of Davidson's program into enough focus to generate interest in and discussion of his views. By way of background, I recommended that the participating Fellows read Davidson's "Radical Interpretation," "Thought and Talk," "On the Very Idea of a Conceptual Scheme," and "Rational Animals." This paper, while not critical, was meant to be provocative.

1. Over the last three decades, Donald Davidson has evolved a complex and highly original philosophical program—a program immensely rich in its implications for a wide variety of central issues in the philosophy of language, mind, action, and metaphysics. By way of bringing a few of the diverse elements of this program into sufficient focus to enable us profitably to discuss them, it may be useful at the start to locate Davidson's work within the context of a larger controversy that has concerned humanists since at least the work of Vico and Hegel—a controversy with which you will all be familiar. I have in mind what, in the late 19th century, emerged as the controversy over the nature of and the relationship between, on the one hand, the so-called natural sciences, die Naturwissenschaften, and, on the other hand, the so-called human sciences, die Geisteswissenschaften.

A central feature of that controversy was the question whether the aims and methodology of the human sciences—given the nature of their characteristic subject matter—was inherently distinctive from and irreducible to the aims and methodology of the natural sciences. On the one hand, there emerged what we might loosely call the "positivist" position, which suggested either that their methods and aims were not inherently distinct or that, if they were, so much the worse for any claimed epistemological legitimacy for the methodology and aims of the human sciences. On the other hand, there emerged a myriad of anti-positivist positions. Of particular interest here—since it is within this tradition that I want to place Davidson—is that strand of opposition to the positivist position that insisted that the aim of the human sciences is not simply to explain human action and its products by subsuming them under appropriate invariant laws, but is, rather, in some distinctive sense, to make the actions, products of actions, indeed, the human beings that perform them, intelligible to us, to make sense of them. On this view, the human sciences necessarily involve an interpretive dimension; the peculiar sort of understanding that is the aim of the human sciences requires interpretation.

Davidson's own concerns and views may, I believe, usefully be understood as belonging to this tradition. In particular, the overarching concern of Davidson's

William Taschek is an associate professor of philosophy at The Ohio State University. His current research interests include the philosophy of language, the philosophy of mind, psychology, metaphysics, and the philosophy of art.
program, as I understand it, is nothing less than the attempt to come to some sort of systematic understanding of the various factors that constitutively enter into and control interpretation: the particular and distinctive way in which we do make sense of human thought and intentional action—including, importantly, speech.

In the standard case, when we are puzzled by an agent’s actions and want to make sense of why she acted as she did, we attempt to determine her reasons for so acting. We want it to make sense to us that she acted as she did for the reasons she did. Reflection on these cases suggests that an action (including an utterance) will be intelligible to us—in the relevant sense—only if it makes sense to us that it made sense to the agent that she did what she did. It is important that there be no equivocation here on the relevant notion of making sense.

I want to suggest that this reflexive constraint is an important (and not fully appreciated) internal constraint on the notion of making sense relevant to the interpretive enterprise as Davidson sees it. If this is right, then the following question immediately presents itself: How can whether an action makes sense to an agent make sense to us unless we suppose that the agent shares in large measure precisely those (fundamental) norms of ours, say, norms of rationality, relative to which we evaluate whether our own actions make sense?

If this question is appropriate, it leads directly to a prima facie dilemma. If we are ever in a position to suppose that an agent fails to share our most fundamental standards of intelligibility, we would seem to be precluded in principle from making sense of her actions, thoughts, and words. Alternatively, if we are ever in no position to know whether or not an agent shares our norms, we would seem to be faced with the real possibility that we are illicitly projecting our standards onto our subjects, thus only seeming to make sense of them. Does the sort of interpretation whose aim it is to make sense of others’ actions, in the relevant sense, inevitably face this dilemma, or is the problem an illusion based on a misunderstanding of the nature and aims of interpretation?

Although Davidson nowhere sets up his problem in quite this way, nor is he explicit in recognizing what I have called the reflexive constraint on the relevant notion of making sense, his views nevertheless manifest an implicit sensitivity to the constraint and, while aiming to do it justice, provide the material for a compelling response to our dilemma. My hope is that seeing how Davidson might respond to this problem will bring sufficient portions of his overall program into enough focus to enable us usefully to discuss and evaluate them.

2. According to Davidson, the problem posed by our dilemma is, indeed, illusory. A correct understanding of the nature and aims of our interpretive practice will expose the illusion; it is precisely Davidson’s project to provide us with just such a correct understanding. Very roughly, Davidson’s proposal is something like this. Thought, language, and intentional action are inextricably interrelated; they mutually condition the possibility of one another—and are such that any attempt to understand one of them in isolation from the others is bound to fail. Indeed, it is only in the context of interpretation (of trying to make sense of other’s thoughts, words, and actions) that we have whatever grip we have on such concepts as belief, meaning, intention, and so forth. Hence, we cannot make sense of some creature as a potential object of interpretation unless we can make sense of our treating that
creature as a thinker, a communicator, and an actor—in other words, as a person.

If it could be argued that regarding a creature as a person—in the rather thin sense just outlined—required in some appropriately strong way the assumption of shared fundamental norms of intelligibility, our dilemma would vanish. If it were a conceptual fact about our concepts of belief, meaning, intention, and so forth that only a creature sharing in large measure our norms can even count as believers, language users, agents—that is, persons—then the purported possibilities raised by our dilemma would be exposed as unintelligible. The question of whether a creature shared our fundamental norms of intelligibility could not—as both horns of our dilemma require—arise independently of whether that creature is a potential object of interpretation, by which I mean a creature with beliefs, a language user, and an agent.

If something like this were defensible, then not only would it be permitted to interpret on the assumption of shared norms of intelligibility, but it would be required. How though might Davidson argue in defense of this position? In order to help sharpen Davidson’s position a bit, let me lay out what will, no doubt, already have struck some of you as the two principle alternative responses to our dilemma.

3. The first anti-Davidsonian alternative suggests that we can intelligibly attribute specific cognitive and conative attitudes to a subject, interpret her words, identify her intentional actions, and still find ourselves in principle baffled by her actions. The idea is that we can identify her action as intentional under a certain description, that we can identify those of her beliefs and desires that constitute her reasons for acting as she did, and that, nevertheless, since her norms of intelligibility differ from ours, we are precluded in principle from making any sense of her actions. They make sense to her; but precisely because we don’t share the relevant norms, it cannot make sense to us that they do so.

To take this line is, in effect, to embrace the first horn of our original dilemma. If Davidson is to defend his position, he must show that this purported possibility is unintelligible. And he will try to do precisely this by showing that it is in fact not even possible to attribute specific contentful attitudes to a creature nor meanings to its utterances, unless we share with it our most fundamental norms of intelligibility.

Suppose Davidson can show this. Would it suffice to establish the sort of strong connection between shared norms and “personhood” that he requires? The second anti-Davidsonian alternative proposes that this is not enough. According to this second suggestion, we should grant that if a subject failed to share our norms of intelligibility, then we would not in principle be able to interpret her—we would not be able to identify her beliefs, her desires, or even those descriptions under which her actions are intentional; we would certainly not be able to interpret her words. This would not, however, according to the suggestion at hand, suffice to preclude the subject from being a person—a thinker, a speaker, an agent. For after all, she may simply have a radically different conceptual scheme from ours, one incorporating fundamentally different norms of intelligibility. The point here is that someone might agree that interpretability, in the sense of being able to identify the contents of beliefs, desires, and the meanings of uttered words, may require shared norms of
intelligibility, but just because we are blocked—even in principle—from identifying those beliefs, desires, and so forth doesn't mean that there are no beliefs or desires, nor that his "utterances" have no meaning.

This second suggestion grants to Davidson something the first suggestion was unwilling to concede, namely, an essential connection between interpretability and shared norms of intelligibility. But it denies that just because a creature is even in principle uninterpretable by us that he isn't a thinker, language user, and agent—for, after all, he may have different norms of intelligibility. An adequate defense of Davidson's position will require that he also show that this second suggestion is unintelligible.

4. Let us begin with the first alternative. For Davidson, the coherent attributability of (propositional) thought in general, of meaning, and of intentional action, requires the attributability to subjects of beliefs (desires also, but we shall focus on beliefs). The attitude of belief plays a central role in Davidson's conception of a potential object of interpretation. For Davidson, to be a potential object of interpretation one must be a creature it makes sense to suppose has beliefs. Interpretation always involves (though is never exhausted by) the attribution of beliefs.

Of central importance, then, is the question: what are the minimal circumstances under which we can coherently ascribe and identify the beliefs of another person? Davidson will argue that it would be perverse to the point of undermining the very possibility of interpretation or communication to (try to) interpret another in such a way as systematically to attribute to her beliefs most of which are ones that one oneself would regard as absurd or manifestly false or irrationally held under the circumstances—beliefs, in other words, that in various ways violate fundamental norms that underwrite our cognitive practice.

Another philosopher, Sydney Shoemaker (in *Self Knowledge and Self Identity*, [Ithaca: Cornell University Press, 1963] 232), offers what I think is a telling illustration of the point Davidson wants to make. Shoemaker writes:

Suppose that we have discovered a new people whose language we do not as yet know, and that someone has proposed a way of translating this language that involves regarding a certain class of statements (or utterances) as perceptual statements and another class as memory statements. Suppose further that we find these statements to be most commonly uttered, confidently and assertively, in circumstances in which their proposed translations would be false. For example, the expression translated by the English sentence "I see a tree" is commonly uttered, confidently and assertively, when the speaker's eyes are not open or not directed toward a tree, and the expression translated by the English sentence "I ate meat last night" is frequently uttered by vegetarians but seldom by anyone who did eat meat on the previous evening. If this happened, surely there could be no reason for regarding the proposed way of translating their language as correct, and there would be every reason for regarding it as mistaken.
The point this passage makes is that we damage the intelligibility of our readings of the utterances of others when our method of reading them puts them into what we regard to be massive or inexplicable error and irrationality. In other words, we damage the intelligibility of our interpretations of the utterances of others if our method of interpretation has us usually and inexplicably disagreeing with them. But now plainly, this point, though made in the context of interpreting utterances, is inextricably connected with the correlative idea that we would likewise damage the intelligibility of our attributions of beliefs to others were our method to yield widespread and inexplicable error and irrationality. Davidson’s suggestion is that if our current best efforts systematically to interpret another creature would have us attribute to her massive error, incoherence, and irrationality (by our lights of course) across the board, that is, if they made her totally incomprehensible to us, then what better evidence could we have that we have failed accurately to identify the relevant beliefs?

One must be careful not to misunderstand Davidson here. He would allow that in a perfectly ordinary sense we can and often do interpret others in ways that makes them mistaken about a great many things. But the point is that there are limits in principle to how far we can go. The reason we miss this is because we tend primarily to notice and focus on our disagreements and find little occasion to reflect on the vast reservoir of mostly trivial beliefs that we share.

Indeed, as Davidson will argue, it is precisely our agreement with respect to some such vast body of unverbalized, usually trivial beliefs that gives content to our voiced (as well as unspoken) disagreements. This is an extremely important point. According to Davidson, without this fund of agreement, there would be no comprehending the differences—we would be in no position to know what we were disagreeing about, and so we couldn’t know that we were disagreeing. Indeed, according to Davidson, the very possibility of objective disagreement only makes sense against the background of a vast reserve of agreement. We can, perhaps, hear in Davidson’s view an echo of a sentiment expressed by Wittgenstein in his *Philosophical Investigations*, section 242: “If language is to be a means of communication, there must be agreement not only in definitions, but also (queer as this may sound) in judgments.”

The point, then, is not that we want to exclude all disagreements. Disagreements about some (perhaps many) beliefs seems inevitable. Rather, Davidson’s point is that the very intelligibility of such disagreement—as well, needless to say, of agreement itself—requires a vast common ground of belief. And this is because, according to Davidson, we cannot take ourselves to be talking about or thinking about the same things and thereby agreeing or disagreeing unless there is this background of shared beliefs.

Davidson’s argument for this point runs roughly like this: Plainly, we cannot intelligibly take ourselves to be disagreeing with someone unless we can take ourselves to be talking about or thinking about the same thing or things as that other individual. In other words, we cannot intelligibly take ourselves to be disagreeing with someone unless we can identify her beliefs and recognize that they concern the same subject matter about which we hold a differing opinion. But now, according to Davidson, we cannot intelligibly suppose that we and another are talking about or thinking about the same things—that the subject matter of our beliefs is relevant-
ly the same—unless there is a substantial background of shared beliefs involving the very concepts or notions in terms of which the dispute is couched. And why is this? The primary reason has to do with the holistic way in which, according to Davidson, the contents of thoughts are individuated and the centrality that beliefs in particular play in this. The idea is that any thought, including a belief, is identified and described only within a dense pattern of other thoughts—especially other beliefs. This is what Davidson has in mind when, in “Thought and Talk” (in Mind and Language, ed. S. Guttenplan [Oxford: Clarendon Press, 1975] 156), he writes:

It is doubtful whether the various sorts of thought can be reduced to one, or even to a few: desire, knowledge, belief, fear, interest, to name some important cases, are probably logically independent to the extent that none can be defined using the others, even along with such further notions as truth and cause. Nevertheless, belief is central to all kinds of thought. If someone is glad that, or notices that, or remembers that, or knows that, the gun is loaded, then he must believe that the gun is loaded. Even to wonder whether the gun is loaded requires the belief, for example, that a gun is a weapon, that it is a more or less enduring physical object, and so on. There are good reasons for not insisting on any particular list of beliefs that are needed if a creature is to wonder whether a gun is loaded. Nevertheless, it is necessary that there be endless interlocked beliefs. The system of such beliefs identifies a thought by locating it in a logical and epistemic space.

Davidson is suggesting here that any thought has the content it does in virtue of its location in a pattern of beliefs and thoughts—not the least important of which is its standing in various epistemic and logical relations (that is, norm-governed relations) with vast numbers of other beliefs that in relevant respects share features of the content of that thought. In other words, my thought that the gun is loaded could not plausibly count as just that thought unless I also believed, for example, that guns were weapons, that they were relatively enduring physical objects, that they can be used to shoot people, that pistols and rifles and plausibly cannons are guns, that their efficacy depends upon their being loaded, that loading involves adding bullets or something that counts as ammunition, that guns are not the sort of thing you swallow to get rid of a headache, that guns are not good to eat, that guns aren’t afraid of the dark, and so on, and so on. It is in virtue of having this, or some relevantly similar, open-ended, interlocking collection of beliefs and other thoughts that make it the case that the thought in question is plausibly a thought about guns anyway. Lots of these beliefs are trivial and would hardly bear mentioning, and yet if most of them were not held, if instead we supposed their contraries to be held, then the plausibility of counting the original thought a thought about guns would quickly evaporate.

This example helps to illustrate two points. First, it makes plausible the claim that the content of any particular belief is importantly a function of its place in, and essentially depends upon there being, a vast network of other thoughts—most importantly other beliefs—the contents of which themselves concern features of the
subject matter of the original thought. And second, it suggests that the intelligibility of attributing a particular thought to someone rests upon the possibility of attributing to him lots of other thoughts, lots of other beliefs—but beliefs, notice, most of which importantly must be true and intelligibly held by our lights. To assume that a person has only or mostly false beliefs about guns is to undermine the intelligibility of the assumption that any of his beliefs are about guns! It is just this set of ideas that, for Davidson, underwrites application of the so-called “principle of charity” in interpretation: interpret another’s utterances so as to minimize inexplicable error and irrationality.

To be sure, in practice we individuate and identify beliefs, as we do desires, intentions, and meanings, in a great variety of ways. But according to Davidson, consideration of the pattern of relations among beliefs (and among beliefs and the other attitudes) always plays a decisive constitutive role. And reflection makes it clear that the guiding principles of appropriate pattern identification derive ultimately from normative considerations—in particular, from our fundamental norms of intelligibility.

The conclusion, then, is that we can only understand the beliefs, thoughts, and words of another if we can suppose that her beliefs are incorporated in a pattern that is in essential respects like the pattern of our own beliefs, which is to say, a pattern essentially governed by our norms of intelligibility. If this is right, then the first anti-Davidsonian suggestion proves to be unintelligible. However, even if we accept this much, we are still left with the second anti-Davidsonian possibility suggested earlier: the possibility of radical conceptual difference grounded in radically different norms of intelligibility.

5. Before directly examining Davidson’s argument against this possibility, I want very roughly to outline those considerations that, according to Davidson, show that propositional thought requires language, that only interpreters can be thinkers. Though this is a bit of a detour, it will, I believe, eventually help clarify an otherwise obscure feature of Davidson’s “official” argument against the possibility of radical conceptual difference.

The present argument proceeds in three stages. First, we argue that propositional thought requires belief. This will have been established if we accept Davidson’s views, just rehearsed, concerning the centrality of belief in the holistic determination of thought content. Next, we argue that having beliefs requires, in some sense, possessing the concept of belief. And finally, we argue that possessing the concept of belief, in the relevant sense, requires being an interpreter, a language user.

Davidson very succinctly lays out this argument in two paragraphs toward the very end of “Thought and Talk” (170). The argument is explored in more detail in his “Rational Animals” (reprinted in Actions and Events: Perspectives on the Philosophy of Donald Davidson, ed. by E. LePore and B. McLaughlin [New York: Basil Blackwell, 1985] 473-480). A propos the second step of the argument, Davidson writes:

Can a creature have a belief if it does not have the concept of belief? It seems to me it cannot, and for this reason. Someone
cannot have a belief unless it understands the possibility of being mistaken, and this requires grasping the contrast between truth and error—true belief and false belief.

And apropos the final crucial stage, he writes:

We have the idea of belief only from the role of belief in the interpretation of language, for as a private attitude it is not intelligible except as an adjustment to the public norm provided by language. It follows that a creature must be a member of a speech community if it is to have the concept of belief. And given the dependence of other attitudes on belief we can say more generally that only a creature that can interpret speech can have the concept of a thought. Though there are many questions that could be raised concerning both the first and second steps of this argument, the principal interpretive difficulty, I believe, concerns just how we are to understand this last step. How exactly is having the concept of belief, or better, of objective truth, supposed to depend on the possibility of interpretation?

At the end of “Rational Animals” (372), after presenting an analogy concerning the apparent impossibility, for an individual bolted to the earth, of giving content to the concept of an object’s being located a certain distance from him, Davidson writes:

Not being bolted down, I am free to triangulate. Our sense of objectivity is the consequence of another sort of triangulation, one that requires two creatures. Each interacts with an object, but what gives each the concept of the way things are objectively is the base line formed between creatures by language. The fact that they share a concept of truth alone makes sense of the claim that they have beliefs [my emphasis], that they are able to assign objects a place in the public world.

The conclusion of these considerations is that rationality (by which, in this context, he means possessing propositional attitudes) is a social trait. Only communicators have it.

The gist of Davidson’s point here would seem to be something like this. Having beliefs requires understanding the possibility that what one takes to be the case may not in fact be the case. To understand this possibility just is (as far as Davidson is concerned) to have the concept of belief. Perhaps one need not admit this possibility with respect to all one’s beliefs—say, one’s beliefs in simple and obvious logical truths—but unless one admitted it with respect to some, if not most, of one’s beliefs, one could not properly be said to have the concept of belief, nor, indeed, attitudes we should be willing to call beliefs. Understanding the possibility that what one takes to be the case may not in fact be the case is tantamount to having a conception of objective truth. So according to Davidson, having a conception of objective truth is constitutively tied to being what we might call a self-conscious
believer: having the concept of belief or of objective truth involves the recognition that the way I take things to be may be mistaken.

As a self-conscious believer, in this minimal sense, I must understand the distinction between my merely taking something to be the case and something in fact being the case. One might suppose that all that I ever have at my disposal, so to speak, is what I take to be the case, the way things seem to me. But without an appreciation of the distinction between how I take things to be and how they are it seems impossible for me even to have a conception of things seeming to me to be any particular way, for evidently the notion of a thing’s seeming to be a particular way only makes sense in the context of a possible contrast to the way things actually are. Now what is required if we are to make sense of my appreciating that there may be a difference between the way things seem to me and the way they are?

Davidson’s suggestion seems to be that I cannot have a conception of the relevant contrast unless I can understand the possibility of another subject entertaining the same thought as my own, though significantly disagreeing with my attitude toward it—that is, where I believe it, she does not. In other words, Davidson is arguing that belief requires the conception of another and potentially contrary point of view on the same subject matter. This would seem to be what Davidson is driving at when in the “Rational Animals” passage (quoted earlier) he writes that it is the fact that creatures “share a concept of truth [that] alone makes sense of the claim that they have beliefs, that they are able to assign objects a place in a public world.” The claim, then, seems to be that this conception of another point of view on the same subject matter can only emerge in the context of interpretation.

Davidson admits that these considerations (of which I have given only the roughest outline) do not amount to an airtight argument for the claim that only communicators can be self-conscious believers and hence thinkers. Nevertheless, if one finds them persuasive at all, it should now seem that the burden of proof will be on Davidson’s opponent, who will need to show exactly how something short of, or other than and not presupposing, communication suffices to underwrite a conception of objective truth, of the contrast between the way things seem to me and the way they are.

For our present purposes, the important consequence of Davidson’s argument, if we accept it, is this: if it is indeed the case that it is only in the context of interpretation that the relevant contrast emerges, then it will follow that the only concepts of truth and objectivity that we have are precisely those that emerge in the context of interpretation. It is just this consequence that we shall want to keep in mind when we consider Davidson’s argument against the possibility of radically alternative conceptual schemes—to which we now return.

6. The second anti-Davidsonian possibility that we noted earlier depended upon accepting the possibility of radical conceptual difference. Davidson has argued, in “On the Very Idea of a Conceptual Scheme,” that reflection on the conceptual demands of interpretation will suffice to demonstrate the impossibility of radically different conceptual schemes—indeed, if Davidson is right, such reflection will suffice to undermine much of the temptation we might have to speak of conceptual schemes at all.

In order adequately to understand Davidson’s argument, we need first to be
clear about what his target is. What exactly is Davidson denying the possibility of? What would it be for there to be a radically different conceptual scheme in any interesting sense—in particular, in a sense capable of subserving the aims of the second anti-Davidsonian proposal? Plainly, to suppose that some other person or culture has a radically different conceptual scheme from our own, it won't do merely to suppose that they have concepts or words for which we haven't got any direct word-for-word translation. (Also compare Davidson's remarks on Quine, the Sapir-Whorf hypothesis, and Kuhn's "paradigm shifts," in "On the Very Idea of a Conceptual Scheme," [henceforth referred to as OVICS] as reprinted in his Inquiries into Truth and Interpretation, [Oxford: Clarendon Press, 1984], 184.) Rather, a radically different conceptual scheme of any interesting sort would seem to require a more fundamental incompatibility, incommensurability, or incomparability. If a conceptual scheme is to be radically different from our own, in the relevant way, it seems that it must in some sense be in principle inaccessible to us and incommensurable with our own.

That someone has a radically different conceptual scheme from our own (in the relevant sense) would seem then to require either that he is in massive disagreement with us about beliefs—that is, relative to our view of the world most of his beliefs are mistaken; or, alternatively, that most of his beliefs are such that not only don't we share them—in the sense that our beliefs don't overlap with his—but somehow in principle we couldn't come to share them while preserving most of our current beliefs.

Let's identify these two conceptions respectively as the "conflict" conception of radical difference and the "non-overlapping in principle" conception of radical difference. Two individuals would count as having radically different conceptual schemes in the "conflict" sense just in case they are in massive disagreement about their beliefs; that is, relative to one, the other's beliefs would be counted mostly in error. Two individuals would count as having radically different conceptual schemes in the "non-overlapping in principle" sense just in case the one not only failed to share most of the other's beliefs, but also could not in principle come to share those beliefs without giving up most of her current beliefs.

In OVICS, Davidson is primarily concerned with the latter conception of radical conceptual difference, though it should be clear that he would also dismiss the conflict conception as unintelligible. On what basis? Well, if we suppose that Davidson has already successfully argued that the possibility of disagreement only makes sense against a background of massive agreement, it will follow that the situation described in the conflict conception could never arise.

Assuming that Davidson has thus efficiently dispatched the conflict conception of radical conceptual difference, what about the more interesting non-overlapping in principle conception? The suggestion here is that two individuals would differ radically in their conceptual schemes just in case their beliefs failed in principle mostly to overlap. The idea is not that their beliefs are false, by our lights; rather, we are encouraged to suppose that their beliefs are likely to be mostly true—as true as ours are. But even so, we can't, even in principle, share those beliefs without giving up most of our own. Unlike with the conflict conception, our inability simultaneously to entertain both sets of beliefs does not result from their being in conflict—after all, we're supposing that both sets of beliefs, ours and our radical alien's,
are mostly true.

We can bring our problem into line with Davidson’s discussion by supposing that associated with a conceptual scheme will be a language, or a set of inter-translatable languages. (This suggestion will be justified by considerations in support of the interdependence of thought and talk of the sort previously discussed.) We can then understand the claim that there can be radically different conceptual schemes as amounting to the claim that there can be natural languages supporting and supported by beliefs that are mostly true but in principle untranslatable by us. Plainly, it will follow from Davidson’s own principles that, were it intelligible to suppose that there could be a creature or culture most of whose beliefs were both true and in principle not sharable by us, then the language of that creature or culture could not be translated by us.

Davidson’s central claim is that we cannot make sense of such a total (or massive) failure of translation. But how does he argue for this? As Davidson says (OVICS, 185-186):

> It is tempting to take a very short line indeed: nothing, it may be said, could count as evidence that some form of activity could not be interpreted in our language that was not at the same time evidence that that form of activity was not speech behaviour. If this were right, we probably ought to hold that a form of activity that cannot be interpreted as language in our language is not speech behaviour. Putting matters this way is unsatisfactory, however, for it comes to little more than making translatability into a familiar tongue a criterion of languagehood. As fiat, the thesis lacks the appeal of self-evidence; if it is a truth, as I think that it is, it should emerge as the conclusion of an argument.

What is this argument?

According to Davidson, the question of whether the notion of a radically alternative conceptual scheme is intelligible is just the question of whether or not the idea of an in principle untranslatable language, supporting and supported by mostly true sets of beliefs, is intelligible. And the question of whether this is intelligible depends, according to Davidson (OVICS, 194) on:

> how well we understand the notion of truth, as applied to language, independent of the notion of translation [interpretation].

And Davidson’s claim is that we don’t understand it independently at all [cf. previous discussion]. The sense in which this is true is, evidently, supposed to provide the basis for the claim that the possibility of an in principle untranslatable language is unintelligible.

In an especially enigmatic passage (OVICS 194-95), Davidson sets out his argument in this way:

> We recognize sentences like “‘Snow is white’ is true iff snow is white” to be trivially true. Yet the totality of such English sen-
tences uniquely determines the extension of the concept of truth for English. Tarski generalized this observation and made it a test for theories of truth: according to Tarski's Convention T, a satisfactory theory of truth for a language L must entail, for every sentence s of L, a theorem of the form "s is true iff p" where "s" is replaced by a description of s and "p" by s itself if L is English, and by a translation of s into English if L is not English. This isn't, of course, a definition of truth, and it doesn't hint that there is a single definition or theory that applies to languages generally. Nevertheless, Convention T suggests, though it cannot state, an important feature common to all the specialized concepts of truth. It succeeds in doing this by making essential use of the notion of translation into a language we know [my emphasis]. Since Convention T embodies our best intuition as to how the concept of truth is used, there does not seem to be much hope for a test that a conceptual scheme is radically different from ours if that test depends on the assumption that we can divorce the notion of truth from that of translation.

What is the argument here? It is obviously important that, according to Davidson, Convention T "embodies our best intuition as to how the concept of truth is used," and that Convention T makes essential or ineliminable use of the idea of translation (interpretation). But it's not just that it makes ineliminable use of the idea of translation, but rather, according to Davidson, that it makes "essential use of the notion of translation into a language we know (interpretation by us)" [my emphasis]. Where does this latter claim come from, and how is it used against the radical conceptual relativist?

Before addressing this claim directly, let me first clarify the larger strategy of the argument, at least as I understand it. Davidson wants to undermine the claimed intelligibility for us of an untranslatable language. He wants to do this first by revealing certain conceptual commitments one has just insofar as one is willing to countenance something as a language irrespective of the question whether it can be translated or interpreted by us or not.

He then wants to argue that the very conditions that make it possible to regard something as a possible natural language at all preclude its being wholly untranslatable/uninterpretable. The idea is just this. If I am going to suppose that an untranslatable language is possible, then I am supposing two things: I am supposing that something is both a language and untranslatable. Davidson's point will be that the very conditions for my coherently supposing that something is a language at all are inconsistent with the possibility that it might be in principle untranslatable by me. How are we to understand Davidson's remarks in the quoted passage as instantiating this general strategy?

Clearly, the notion of truth plays a central role in Davidson's implementation of the strategy. Recall that our picture of this radically alternative conceptual scheme is one relative to which someone is capable of entertaining mostly true beliefs that we ourselves are precluded from entertaining. The language coordinated with this scheme will, consequently, have to be one in which a multitude of non-translatable
truths can be expressed. This way of understanding the situation is in effect forced on us by the fact that nothing would even count as a conceptual scheme in the relevant sense—a point of view on the world—unless it were a point of view relative to which many true beliefs could be entertained. And correlatively, nothing would count as a language unless it were such that speakers of that language would be capable of expressing truths in it.

In other words, to have a conceptual scheme in the relevant sense is at least to be in a position to have beliefs about the world—attitudes with respect to which the question of truth or falsity arises; and to be a language in the relevant sense is at least to be something in which claims about the world can be made—something some of the sentences of which are evaluable as true or false. Consequently to have a language in the relevant sense is at least to be able to express and interpret one’s own and others’ beliefs about the world. This should make it relatively clear then that there is an essential connection between being a language and the notion of truth. Nothing will count as a language in the relevant sense unless it is capable of expressing truths.

Needless to say, the concept that we express with the word “language” is our concept, as is the concept we express with the word “truth.” The question is whether it is coherent to suppose that there could be a language capable of expressing truths—where, recall, the concept expressed by that term is ours (whose else would it be?)—which, however, cannot in principle be expressed in our language or some suitable expansion of it. (Hereafter I will leave off making this important qualification explicit.)

Davidson wants to argue that it is not possible both to conceive of a language capable of expressing truths (which is just to say, to conceive of a language) and to conceive of that language as incapable in principle of being translated/interpreted by us. To imagine a language as capable of expressing true sentences is to imagine our truth predicate significantly applying to its sentences—which, importantly, is different from imagining our actually applying our truth predicate to its sentences. In order to imagine this latter possibility, we would have to imagine having satisfied ourselves that, at the very least, what we were predicating truth of was indeed an indicative sentence in the other language. But in the case of a radically untranslatable language, it is difficult to see how we could satisfy ourselves of even this much. So let me repeat: to imagine a language as capable of expressing true sentences is to imagine our truth predicate significantly applying to its sentences. But now what is required in order coherently to imagine our truth predicate significantly applying to a sentence?

It is at this point that Davidson’s appeal to Convention T becomes relevant. Davidson writes that “Convention T embodies our best intuition as to how the concept of truth is used.” Clearly, Davidson’s argument is going to be that whatever that best intuition is, it will be a consequence of it that we will be able coherently to imagine our truth predicate significantly applying to a sentence of a language only if we can coherently imagine the possibility of translating that sentence into our own language, that is, interpreting it. What, then, is the content of this “best intuition,” and how does it follow that it makes no sense to imagine our truth predicate applying to a radically untranslatable sentence? The relevant passage is this:
according to Tarski’s Convention T, a satisfactory theory of truth for a language L must entail, for every sentence s of L, a theorem of the form “s is true iff p” where “s” is replaced by a description of s and “p” by s itself if L is English, and by a translation of s into English if L is not English.

Davidson’s claim here seems to be that our understanding of the truth predicate and our grip on its possible extension—that is, our grip on what is to count as an item to which it might genuinely apply—is essentially captured by Convention T. In other words, it seems to be Davidson’s view that it is a necessary condition on the possibility of some sentence s belonging to the extension of the truth predicate that it be an s for which there could be in English (or an appropriate extension of English) a T-sentence of the form s is true iff p, where s is a sentence in some language and “p” is replaced by a translation of that sentence into English. The suggestion is that it is an essential part of our understanding of the truth predicate that its extension is uniquely fixed by the totality of true instances (in our language) of the T-schema.

If this is right, then plainly we cannot make sense of anything belonging to the extension of the truth predicate unless it is a sentence in principle translatable into our language. But then we cannot make sense of the truth predicate’s applying to the sentences of a language radically different from our own. But then we cannot make sense of the possibility of such a language at all.

But is this right? Well, the sense in which it is right (if it is) is revealed, I would suggest, by reflection on the important consequence, noted previously, of Davidson’s “Rational Animals” argument. I earlier said that if we accept that argument, then it will follow that the only concepts of truth and objectivity we have are ones that emerge in the context of interpretation. Consequently, where and only where interpretation is conceivable will the concepts of truth and objectivity find significant application. If this is right, it is not possible to make sense of any seriously radical conceptual relativity. Consequently, the second anti-Davidsonian proposal is unintelligible.

If we can assume that Davidson has indeed successfully countered both of the original anti-Davidsonian suggestions, then he will, it seems, have successfully established a strong sense in which in interpretation we must presume some shared fundamental norms of intelligibility. And, it will follow from the peculiar strength of this that how we must interpret someone (viz., as in large measure sharing our fundamental norms of intelligibility) is how in fact they are.