

Knowing Another's Thoughts

Essay for Mindreading and Joint Action course

Sara Jellinek

In the present essay I am going to discuss Davidson's theory on interpretation. First, I will introduce his theory and then discuss some potentially problematic points related to his idea.

Davidson in his 1973 paper *Radical Interpretation* presents – and later refine – a theory on how human hearers are able to understand speakers' utterances. More precisely he develops a theory on what we could know that enable us to interpret what others say.

The first step is that he formulates a claim about a theory of radical interpretation that would enable us to assign meaning to other's sentences. Such a theory has to be – among other requirements that are not discussed here – supported by evidence available to the interpreter. The hearer can not rely on the meaning or interpretation of the sentence as an evidence for the soundness of the theory. If she did so that would result a circular proof. Instead, one possibility is taking *translation* as a starting point. To know how translation works she first has to know how to interpret utterances of her own language – so that how she formulates her own utterances. (Davidson, 1973)

Davidson suggests that we best rely here on Alfred Tarski's theory of truth. (Davidson, 1973, 1974, 1980, 1990) However, Tarski worked out his theory for formal languages, Davidson holds that with some changes and additions it will fit natural languages. According to the theory for every sentence s of the object language " s is true (in the object language) if and only if p " (Davidson, 1973:131) From these we can obtain so called T-sentences by replacing " s " by a canonical description of s and " p " by a translation of s . These define the truth conditions for a sentence. In addition they do so for all the sentences in a language because of the Convention T that requires a corresponding T form for every viable sentence in the object language. These T-sentences give us a manual so that we know how our sentences correspond to states of the world if they are true. At this point we know how to interpret our own language.

We now have to have evidence that others also behave similarly so that we could use our

translation manual for the interpretation of their sentences as well. Truth seems to be an accessible element of the theory – because of T-sentences. Making a small step forward we suppose that a speaker holds a sentence true because what the sentence means and what that person believes. So to verify our theory that others also use theory of truth as a guide to translate and define the meaning of their sentences we only have to see if the meaning of their sentences and their beliefs correctly correspond to truth. The problem is that we have no access to all these before an interpretation. This is because according to Davidson "there is [...] a principled obstacle to verifying the existence of belief and intentions while being unable to tell what a speaker's words mean." (1974:143-144.) Which means that we only have an access to the speaker's belief via the meaning of her sentences. We have to find another way for finding out what a speaker holds true getting around *meaning* and *belief*.

The issue now under discussion is how can we tell that a speaker accepts a sentence to be true? One supporting assumptions can be that the circumstances the speaker utters her sentences are present to us. Also we suppose that the speaker behaves similarly to us. So even if we do not have a direct access to the speaker's beliefs we can legitimately hold that her beliefs are incorporated in a pattern that in their essential respects similar to ours. As a consequence we can project our own logic to her beliefs. (Davidson, 1980, 1990)

We are now close to the solution. There is, however, one more idea that has to be discussed. This is about how we choose our sentences we utter. If we find out this can enable us to infer about why and how the speaker chose her sentences that she thinks corresponds to truth about the world. Knowing about what she finds to be true we can in return infer about her beliefs and the meaning of her sentence so this gives us the possibility of interpretation. (Davidson, 1990)

To do all this we first have to start from the assumption that an agent always prefers – and have the possibility to prefer – one sentence true over another. (1990). But what defines her preference? To answer this question, Davidson turns to Bayesian decision theory. (1980, 1990) Summarizing the idea simply and briefly we can say that a person always try after maximizing utility. This is in this case to choose a sentence that best corresponds to truth among all the possible alternatives. Because of the *principle of charity*¹ and also considering the speaker as a rational agent we trust that she would choose the best possible option available to her. (1990) Knowing now which sentence she chose and that she thinks this corresponds best to truth comparing her choice to the circumstances within she uttered her sentence we can infer her beliefs

¹This means to consider the speaker's statement to be rational and interpret her utterance in the strongest possible way.

and also the meaning of her sentence.

To sum up briefly to be able to infer the meaning of a speaker's sentence – according to Davidson – one has to have in mind many assumptions and presuppositions. We have to have access to the circumstance the utterance made in. We suppose that the speaking agent is rational and has a similar structure of logic and reasoning to ours. Also the speaker has analogous belief patterns to ours. All these will make the agent to act just like we would do and choose the best possible sentence to express her ideas. Under *best* we mean the one that the speaker holds *true*. Truth in turn allows us to infer how things are in the world. What is more, since meaning and belief defines truth for an agent based on what she thinks to be true we can have access to her beliefs and what she means by the sentence. What is the most important is that we did all these without relying on the concept or belief that we wanted to find out at the end, so our reasoning is not circular.

Unfortunately, there is a major problem with Davidson's theory: we can attribute and have an idea about others' beliefs without language. Take as an example preverbal infants. We can reason about their actions based on our idea about what they might know what beliefs they might have. If we had doubts that infants would have beliefs, another example can be silent movies. There is no sentence uttered, still we can understand what is happening and why is happening.

So then what is that we know that enables us to attribute beliefs, to know another's thoughts even without language?

I believe that there are many important assumptions Davidson made in his theory. I would now focus on two of them. One is that we suppose the speaker to be a rational agent. It has been proved that even infants and very young children act rationally. (Gergely et al. 2002) So trying to interpret others' actions relying on the presupposition that the agent acts rationally is a valid thing to do. Also, from this follow that we use knowledge of our own behaviour while reasoning about others', since it is plausible that the same considerations drive their thoughts and beliefs as ours. This is important for the reason that we can never have direct access to others' minds, but the best thing we can do is that trying to draw the strongest possible parallel between others' and our minds. We can do this by searching for a possibly shared feature of our behaviour that actually drives actions. And this is – we can legitimately suppose – rationality. Finding proof for our assumption of rationality is a matter of observation, which is great, since we do not have to rely on beliefs or contents of others' minds which we are searching for.

However, rationality is not a substantive property of behaviour. It always depends on and defined by past experiences, knowledge and actual circumstances. This is why observation can be a tool for the assessment of rationality. We have access to the same circumstances as within the other agent acts.

This would be the other assumption made by Davidson that I find important. If we have access to the same circumstances as some other agent then we can use these environmental factors that drives behaviour to form our own beliefs to reason about others' actions. In addition to present environment we might also rely and have access to general facts in the world, history of actions of the agent, past experiences, shared and distributed knowledge. These all form beliefs and drives reasoning about rational acts.

Probably a 'theory' that uses only rationality, shared knowledge and circumstances as a basis for explanation of behaviour seems to be pretty simplistic. Still, I believe, that the advantage is that this way we do not have to use vague or problematic concepts like *truth* or *meaning* that can be a matter of debate. These allow us to infer the content of most important constituents of human behaviour like goals, desires, intentions, actions. Also, since such a theory would not exploit metaphysical concepts it does not expel that the interpreter might be wrong about her interpretation or have false beliefs. Such a flexibility I think should be a requirement for all theories attempting to explain human behaviour.

This more simplistic account can be used even as a theory of meaning attribution in language. Davidson was not looking for a theory of meaning in the Fregean sense so that how do individual words get their meaning, rather how greater structures like sentences acquire meaning. However, he only considered affirmative sentences, the present preliminary theory' could attribute meaning also to questions, imperatives, requests, etc. By supposing that we can know others' beliefs we can make sense of utterances of the form other than affirmative sentences. If Davidson gave up his idea that beliefs are anchored in language, he could have a stronger theory. This seems to be the most crucial mistake in his account.

I believe, even if his theory is problematic in many respects, Davidson called the attention to some important aspects of human cognition that is useful not only in understanding language.

Appendix

There is one more issue left that is I think worth to mention related to language and meaning. It is the matter of figurative language. I believe most of the instances of figurative speech can be explained (and assigned meaning) by our present theory'. Keeping the discussion short, I would only present one simple example. To understand irony or sarcasm we can to rely on the supposed content of others' minds. By knowing that the literal meaning of the sentence does not fit with the assumed beliefs and thoughts of the speaker we can arrive to a right interpretation of her utterance.

I would claim on the other hand that the case is not this simple with metaphors. In metaphors – especially in novel ones – the speaker relates two distinct domains in a way that expresses in turn something about what she thinks, believes, how she sees the world. I think in this case we can not profit that much about what we hold about the speaker's beliefs, since here it is exactly what is expressed. If someone utters *Sam is a pig*. and we do not know anything about Sam or the relationship of Sam and the speaker, then our assumptions about the speaker's beliefs will mean no help for us.

However, neither Davidson's theory nor the present essay aimed to explain the process of understanding metaphors, still I believe, such problematic and interesting cases are the ones that trigger improvement of theories in science.

References

- Davidson, D. ([1984] 1973). *Radical interpretation*. In *Inquiries into Truth and Interpretation* (pp. 125–139). Oxford: Oxford University Press.
- Davidson, D. (1974 [1984]). *Belief and the basis of meaning*. In *Inquiries into Truth and Interpretation* (pp. 155–170). Oxford: Oxford University Press.
- Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, 11, 1–12.
- Davidson, D. (1990). The structure and content of truth. *The Journal of Philosophy*, 87(6), 279–328.
- Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, 415, 755.