CHARITY AND HUMANITY IN THE PHILOSOPHY OF LANGUAGE

GARETH FITZGERALD
UNIVERSITY COLLEGE LONDON

Abstract

There is a popular idea that a shift from a Principle of Charity to a Principle of Humanity, as famously advocated by Richard Grandy (1973), offers considerable advantages in constructing theories of meaning for natural languages. My claim is that Grandy’s case for the superiority of the Principle of Humanity does not tell against the Principle of Charity developed by Donald Davidson. The paper outlines important developments in Davidson’s Principle of Charity, and his refinement of the Principle of Charity that he found in Quine’s writings. I argue that Grandy’s criticisms, whilst sound against the Quinean principle he had in mind, do not extend to Davidson’s refined Principle of Charity. I then raise two further issues that face advocates of the Principle of Humanity as an improvement on Davidson’s Principle of Charity.

A prominent tradition in the philosophy of language has been concerned with the project of constructing theories of meaning for natural languages on the basis of observable evidence from the behaviour of speakers.¹ There is a popular idea that a shift from a Principle of Charity (POC) to a Principle of Humanity (POH), as famously advocated by Richard Grandy², offers considerable advantages in constructing such theories of meaning for natural languages.³ My claim is that Grandy’s case for the superiority of POH does not tell against the POC developed by Donald Davidson (from here on, “Davidson’s Principle of Charity” reads DPOC). Section 1 outlines some developments of DPOC from the POC found in Quine. Section 2 suggests that Grandy’s criticisms, whilst sound against

¹ There are a number of competitors to this framework in philosophy and linguistics. Prominent amongst these is the Gricean programme, and most contemporary work in linguistics including the Relevance-theoretic framework partly inspired by Grice - has a more cognitive orientation.
² Grandy (1973)
³ See, for example, David Papineau (1987, pp.31-9) or David Wiggins’ papers on the philosophy of language. Wiggins says: “The requirement that we diminish to the minimum the theoretical need to postulate inexplicable error or irrationality is a precondition of trying to project any interpretation at all upon alien speakers. It was phrased by Davidson in another way, and called the requirement of charity. The replacement given here is closer to what has been dubbed the requirement of humanity.” (Wiggins 1987, p.112n)
the Quinean principle he had in mind, do not extend to DPOC. Section 3 points out two issues that face advocates of POH as an improvement on DPOC.

1. The Principle of Charity: From Quine to Davidson

A translator faces the task of constructing a manual, or theory, that maps sentences of one language on to those of another that are alike in meaning. Grandy’s translator, like Quine’s, has been charged with the task of construing the utterances of an unfamiliar speaker in the language of the translator, a language that he understands. Quine’s POC, with which Grandy is concerned, instructs the theorist to “maximise agreement on obvious truths.” (Grandy 1973, p.440) The class of obvious truths, for Quine, consists of observation sentences – those sentences that would receive community-wide assent or dissent under uniform stimulation – and logical statements. Quine supported POC on the grounds that translation manuals that found absurdity in a speaker’s utterances or beliefs were less probable to be correct than to be instances of bad translation:

> Assertions startlingly false on the face of them are likely to turn on hidden differences of language.

Davidson originally saw himself as applying Quine’s principle, but applying it “across the board” (Davidson 1973, p.136n) rather than to a privileged class of obvious truths including observational statements and logic.

Both Quine and Davidson accepted that some such principle was required by the theorist of meaning in order to disentangle questions of what a speaker means from questions of what he believes. This is because the way we construe a speaker’s words will depend on what beliefs we take him to be expressing, and similarly, which beliefs we think he is expressing will depend on what we take his words to mean. Applying a POC to the speaker’s beliefs serves to hold one of these unknowns (his beliefs) fixed, to some degree, whilst the theorist solves for the other. In this way the theorist can get some purchase on the meanings of the speaker’s words. Davidson, however, dropped Quine’s talk of obvious truths. One reason for this was that he did not accept Quine’s distinction between

---

4 The development of the principle undertaken by Davidson is not sufficiently appreciated. Wiggins, for one, seems to take Quine’s POC and DPOC to be equivalent. Critiquing a principle of utility maximisation in the context of discussing weakness of the will, he says: ”No doubt some theorists will propose that the statement is a regulative maxim or a principle of interpretation of behaviour and above the mêlée of falsifiable sentences. Such a theorist should study the fate of the so-called Principle of Charity, which was defended by Quine and Davidson as a principle of radical interpretation.” (Wiggins 1987, p.264n)

5 Quine (1960, p.59)
observation sentences – held true by everyone on the basis of their direct conditioning to sensory stimulation – and theoretical sentences – held true in virtue of their place within a wider body of beliefs.

But two further moves distance DPOC from the POC that Quine endorsed and that Grandy attacks. Davidson was unhappy with the idea of *maximisation* as applied to properties of beliefs. He could conceive of no useful way to count beliefs and so could give “no clear meaning to the idea that most of a person's beliefs are true.” (Davidson 1983, pp.138-9) Correlatively, the number of possible sentences to be construed is infinite and no clear sense can be given to an interpreter maximising over infinities. Neither should his early talk of maximisation (see, for example, Davidson 1967) be reconstructed in terms of a general presumption in favour of the truth of a speaker’s beliefs. This presumption would be compatible with each individual belief that we attribute to a speaker turning out to be false, which Davidson denied is a possible result of interpretation. So, Davidson suggested that rather than “maximise”, “a better word might be *optimize*” (Davidson 2001, p.xix) which connects to the second difference between POC and DPOC.

Davidson advocated a shift in focus from *agreement* to *intelligibility*. Davidson thought that finding some background agreement with a speaker is necessary to understanding him. But, nevertheless, he held that the target of a theory of meaning is always to understand the speaker’s words rather than agree with him:

> Charity prompts the interpreter to maximise the intelligibility of the speaker, not the sameness of belief. This entails ... that interpretation must take into account probable errors due to bad positioning, deficient sensory apparatus, and differences in background knowledge.⁶

These moves represent a shift from the old Quinean dictum of maximising agreement on obvious truth to the broader goal of optimising the intelligibility of a speaker – making as much sense of him (what he says and what he believes) as possible. For reasons I’ll turn to momentarily, Davidson thought that making a speaker intelligible involved finding the “right kind of agreement” (Davidson 2001, p.xix) with him. But Davidson thought there to be no more general way of saying what kind of agreement is necessary for making sense of someone's beliefs than there is of saying what counts as a good reason for holding a belief. Making someone intelligible involves rendering their beliefs and utterances explicable in light of respective background theories about that speaker and their environs.

---

⁶ Davidson (2001, p.xix)
The POC that Grandy attributed to Quine and wished to criticise is the following:

\[ \text{Choose that translation which maximizes agreement (at least of certain sorts) between ourselves and our translatee.} \]

It is worth noting that Grandy offers POC as a rule by which to choose between alternative available theories of a speaker’s language. Rightly or wrongly, this is not how Davidson saw DPOC. Davidson conceived of his principle as having the status of a necessary condition on obtaining a theory of a speaker’s language at all, not as a means by which to pick amongst several rival theories. The thought behind this necessity claim is clearest in light of later formulations of the principle. These later formulations of DPOC involve two constituent principles:

The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called Principles of Charity: one principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true belief about the world.\(^8\)

Here, Davidson is suggesting that to understand a speaker a theorist must take them to be engaging, with some consistency, with a shared environment. Repeated applications of these principles will ramify through the speaker’s holistically attributed belief-set, endowing the speaker with some beliefs that are true and consistent by the theorist’s lights.

This does not require that the theorist cannot attribute the speaker beliefs that are incompatible with his own. But it does ensure that the theorist finds the speaker to have enough beliefs in accord with his own, on pain of undermining the assumption that the speaker is responding to any particular feature or features of their shared environment.\(^9\)

There must be some shared subject matter between theorist and speaker for the theorist to make sense of their disagreeing or even their agreeing. Finding too much initial error and unreason on the part of the speaker, so the argument goes, will thwart our attempts to see what the speaker could be in error or unreasonable about. It is not, as sometimes assumed, part of the content of DPOC that an intelligible speaker must have a preponderance of

---

\(^7\) Grandy (1973, p.440)
\(^8\) Davidson (1991, p.211)
\(^9\) The notion of “enough beliefs” employed here does not require that one be able to count a speaker’s beliefs, or maximise over the properties of such an infinite set. It implies only that the theorist can impute sufficient beliefs to the speaker that he can secure the assumption that they are responding to, and thereby thinking about, the same aspects of the environment.
true beliefs but rather a product of theories constructed according to its principles:

It is an artefact of the interpreter’s correct interpretation of a person’s speech and attitudes that there is a large degree of truth and consistency in the thought and speech of an agent.10

DPOC itself does not contain any general conclusions about the nature of belief or interpretable agents.

By the above kind of reasoning, Davidson defended DPOC as a necessary condition on constructing a theory of meaning for an unfamiliar speaker. Jumping forward slightly, it is worth noting that Grandy describes his own POH as a “pragmatic constraint” on translation. If we treat DPOC as a pragmatic rule of thumb for plausible translation then it seems we miss the bite of Davidson’s reasoning.

By way of summary, we can say that Davidson saw DPOC as a necessary condition on constructing a theory of meaning for a speaker’s language, serving to disentangle questions of meaning and belief, and provide an entering wedge into the speaker’s language. These are aims he shared with Quine. But DPOC differs from Quine’s POC insofar as (1) it is applied “across the board” rather than to a privileged class of obvious statements, (2) it does not appeal to the notion of maximisation as applied to agreement, but rather seeks to “optimise” such agreement as is required for making the speaker most intelligible, and (3) it consists of two constituent principles of Correspondence and Coherence, which make more perspicuous the requirement for sufficient agreement between theorist and speaker.

2. Grandy’s Critique of Charity

The POH which Grandy advocated instructs the theorist that:

If a translation tells us that the other person’s beliefs and desires are connected in a way that is too bizarre for us to make sense of, then the translation is useless for our purposes. So we have a pragmatic constraint on translation, the condition that the imputed pattern of beliefs, desires and the world be as similar to our own as possible. This principle I shall call the principle of humanity.11

Grandy has two arguments for its superiority over POC; one directed at its pronouncements on logical statements and a second, directed at observational descriptions. Both of them

11 Grandy (1973, p.443)
do in fact put POH one-up on POC but not, I shall argue, on DPOC.

Grandy’s first argument is much easier than the second for Davidson to accommodate. Grandy considers a complicated logical theorem that can be proven but that could, prior to one's seeing the proof, appear invalid. Grandy imagines showing the theorem to someone and asking them if it is valid. It is quite possible, likely even, that they will give the wrong answer. But Quine’s POC instructs the theorist to maximise agreement on logical truths and translate the speaker so that he takes the theorem to be valid, as the theorist does. Grandy says:

(T)he past history of a speaker is quite relevant to the question of what is obvious to him ... We should not go out of our way to find some complicated agreement on this logical truth, because the error is not only explainable but was predictable given some knowledge of his past history.\(^\text{12}\)

It strikes me that Grandy is accommodating the kind of considerations of overall intelligibility, and placing a theory of a speaker’s utterances against a background body of theory about that speaker’s beliefs, in a way that is very congenial to Davidson’s approach. For DPOC is not a commitment to preserving a class of obvious truths, and does not prescribe that agreement on such truths is preserved at a cost to overall intelligibility. DPOC does not counsel the theorist to find agreement on complex logical truths at the expense of the intelligibility of his overall interpretation. In optimising global intelligibility rather than agreement, DPOC guides the theorist to find some shared concerns and consistency in the speaker. But there is no reason to think this implies finding agreement over complicated logical theorems – to see the subject as holding such logical beliefs might require us to depart, in quite fundamental ways, from the background theory of beliefs that we have developed for him.

Grandy’s second argument concerns the translation of definite descriptions in observational statements. Grandy claims that applications of POC lead to unnatural translations whereby a speaker is attributed true but inexplicable beliefs. Grandy’s point, as we shall see, is solid against the old maximising agreement POC but does not carry over in any straightforward way to DPOC. Crucially, according to DPOC there will be cases, even early on in the theoretical endeavour, where the best interpretation may involve the attribution of false belief. Grandy recognises that Quine’s POC can accommodate the attribution of some false beliefs to a speaker. But the interesting point about the cases Grandy has in mind is that they are widespread, and can easily be generated systematically so that the maximising

\(^{12}\) Grandy (1973, pp.443-4)
POC cannot accommodate them. Grandy claims that these cases are best accommodated by adopting POH. The reason he offers for the superiority of POH in these cases is that it is sensitive to causal evidence afforded only a peripheral role by POC. But I think that DPOC gets these cases right without committing to the further claims about reference and belief that Grandy does.

The sorts of cases Grandy focuses on are the following.\textsuperscript{13} Paul arrives at a busy party. Paul asserts “The man with a martini is a philosopher”. There is a man in plain view who is drinking water from a martini glass and who is not a philosopher. There is in fact only one man who is drinking martini at the party. This man, as it happens, is a philosopher and he is standing out of sight in the garden. Who should the theorist take Paul to be talking about in his statement? Grandy points out that, according to Quine’s POC, which advises the theorist to maximise agreement, we should take the remark at face value and count Paul’s utterance as true. This implies that “the man” refers to the man in the garden. We thereby maximise the truths that Paul believes. Moreover, we maximise agreement on Quine’s obvious truths because we do not attribute to a speaker a false observation statement. But the more natural thing to do is to take the utterance to be false because we cannot explain how Paul could have the true belief about the man outside who he has not seen. The falsity of the utterance is:

\begin{quote}
Predicted by the principle of humanity, of course, for that constraint instructs us to prefer the interpretation that makes the utterance explainable.\textsuperscript{14}
\end{quote}

Where Paul has had no causal contact with the man in the garden but we can see how he could be causally interacting with the man in front of him, then it seems explicable how he could have the relevant beliefs and be speaking about the latter but not the former. We could, of course, supplement the details of this story in further ways such as the converse is true but as it stands it makes more sense of Paul’s utterance in this everyday circumstance to take him to be referring to the man in front of him.

Grandy draws a general moral from such cases: POH fits best with a causal theory of belief. As Grandy thinks a causal theory of belief is integral to our understanding of others, he

\textsuperscript{13} This case is owed to Donnellan (1966). Important here is the distinction between attributive and referential uses of definite descriptions. The case I describe involves taking a speaker to make a false utterance on a referential interpretation of a definite description rather than a true statement on an attributive interpretation. For the utterance in question to come out false, the speaker must succeed in referring to the individual picked out by the definite description regardless of the fact that he does not satisfy the condition specified by the descriptive material.

\textsuperscript{14} Grandy (1973, p.445)
thinks this speaks strongly in favour of POH. Grandy’s principle directs us “to bear in mind that the speaker is a person and has certain basic similarities to ourselves when we are choosing between translations.” (Grandy, 1973, p.445) Grandy suggests that part of our own view of ourselves is “what could be loosely described as a causal theory of belief”:

In the example just given, the operative principle that makes the case forceful is that the object referred to according to the charitable interpretation did not interact causally with the speaker. Thus the causal theory of belief accords much better with the principle of humanity than with the principle of charity.  

I’m not sure whether the causal theory of belief is “part of our own view of ourselves” but I think that DPOC can give the right answers in these cases, and do justice to the relevant causal factors, in the case that Grandy describes.

Grandy’s notion of sharing “certain basic similarities” needs elaboration. If Paul’s patterns of belief-formation are similar to ours then without his either being observationally positioned to observe to the man in the garden or his holding some relevant background beliefs then we cannot make intelligible his belief about which drink the man in the garden has. It is entirely plausible to reason in such a way but it would be wrong to think that DPOC prescribes against such reasoning, quite the contrary. Rather the Principle of Correspondence can do for Davidson what POH can do for Grandy. Remember the Principle of Correspondence instructs the theorist to take his speaker to be responding to the same features of the world as he would in the circumstances. In this case neither Paul nor his interpreter would be responding to the man in the garden.

By DPOC, attributing Paul a belief about the man in the garden will violate the Principle of Correspondence (unless we have a background theory of Paul’s beliefs which make such an attribution intelligible, selecting the man in the garden as an aspect of the environment that the theorist might be responding to in Paul’s circumstances). The theorist would not in Paul’s circumstances be responding to a man in the garden, for he is not in a position to respond to this man. The Principle of Correspondence affords a central role to Paul’s interaction with his surrounding environment, and in doing so, picks up on the causal factors relevant in belief attribution to which Quine’s principle was insensitive.

Moreover, according to Davidson’s holism, attributing Paul the belief that the man in the garden is a philosopher and is drinking a martini would require attributing him a surrounding network of beliefs for which one would have to find some evidence. So it is not peculiar to

15 Grandy (1973, p.445)
POH to attribute subjects such explicable falsehoods as Paul’s, rather than mysterious true beliefs. Given the features of the world the theorist would be responding to in the speaker’s circumstances (Principle of Correspondence), and the other beliefs we have grounds to attribute that speaker with which this belief must cohere (Principle of Coherence), the erroneous belief on the speaker’s part is the more intelligible according to DPOC.

Unlike Quine’s POC, DPOC falls prey to neither of the problems that Grandy raised. The first of Grandy’s arguments against POC presses on its commitment to maximising the logical truths within a speaker’s belief set. But DPOC is not committed to maximising agreement with a speaker over a class of obvious logical truths. As DPOC is committed to optimising the intelligibility of a speaker against a background theory of their beliefs it would not prescribe that agreement on such logical truths is preserved at a cost to overall intelligibility. The second of Grandy’s arguments against POC presses on its commitment to maximising the truth amongst a speaker’s beliefs, including in particular their observational beliefs. The truth maximising principle dictates that the theorist attribute a speaker true beliefs about objects that we cannot explain them having beliefs about, rather than false beliefs about objects we can explain them having beliefs about. The Principle of Correspondence that is part of DPOC, interacts with the Principle of Coherence, to block this attribution.

3. Some Issues with the Principle of Humanity

In this final section, I want to raise two problems facing advocates of Grandy’s POH. The ultimate goal of such theories of meaning is to make as much intelligible sense of a speaker’s words as he can on the basis of the observable evidence. But Grandy’s statement of POH actually instructs the theorist to impute a higher degree of similarity in the beliefs and desires of speaker and theorist than is strictly required by the condition that he makes as much sense of the speaker as he can. Recall Grandy’s principle:

If a translation tells us that the other person’s beliefs and desires are connected in a way that is too bizarre for us to make sense of, then the translation is useless for our purposes. So we have a pragmatic constraint on translation, the condition that the imputed pattern of beliefs, desires and the world be as similar to our own as possible. This principle I shall call the principle of humanity.\(^\text{16}\)

It is true that a translation that is too bizarre for the translator to make any sense of will be useless. But this does not prescribe that the theorist impute “the pattern of beliefs, desires

\(^{16}\) Grandy (1973, p.443)
and the world as similar to our own as possible”17. It only prescribes that the theorist discard translations that are unintelligible. There are degrees of similarity in the patterns that a theorist can impute to a speaker, all compatible with his finding the speaker intelligible. What degree of similarity he finds in such patterns should be sensitive to the evidence he collects. It is unwarranted to suggest that only a theory that makes the subject as similar as possible along these dimensions could further our understanding of him. It is a plus point of DPOC that it does not assume this degree of similarity. When viewed as a necessary condition on the disentanglement of meaning and belief, POH’s imposition of maximum similarity in patterns of belief and desire looks heavy-handed relative to DPOC.

There is also an issue about the way in which POH is to be applied and its breadth of application. DPOC is a means of breaking into the speaker’s language by fixing some of his beliefs independently of determining the meanings of his words. Grandy certainly introduces POH in the context of translation as a rival to Quine’s principle which plays such a role. But Grandy also envisages POH as a means by which, once we already have a belief-desire theory for a speaker and a translation manual for his language, to go about predicting behaviour. So Grandy’s concerns seem a lot broader than the disentanglement of meaning and belief for which DPOC is employed, extending to determinate predictions of a speaker’s behaviour.

Evidence of Grandy’s broader concerns is his worry that a belief-desire theory, a translation manual and a collection of the facts about nonverbal behaviour are not sufficient evidence for determinate prediction of a speaker’s future behaviour by any heuristic that could be available to normal communicators. He takes it that this insufficiency calls for a further model of the speaker and moves for POH:

> In theory we could (perhaps) elicit the total belief-desire structure and use mathematical decision theory to arrive at the prediction, but this is not what we do in practice. And since it was the actual process of communication that was our original concern, we must look for an alternative model. The most obvious alternative is that we use ourselves in order to arrive at the prediction: we consider what we should do if we had the relevant beliefs and desires.18

So we can at least say that as Grandy intended it POH is not just a method for disentangling questions of meaning and belief in construing a speaker’s words.

But we might also note, by way of developing this point, that Grandy seems to have a

18 Grandy (1973, p.442)
different role in mind for POH than Davidson had for DPOC. Grandy says that:

Whether our simulation \textit{of a speaker} is successful will depend heavily on the similarity of his belief-desire network to our own.\textsuperscript{19}

If POH were supposed to be a necessary condition on understanding a subject then we would have no independent means by which to assess whether a belief-desire network is similar or dissimilar to our own. The similarity would be \textit{constitutive} of our speaker being interpretable as having belief-desire patterns. It wouldn’t make much sense to ask the prior question about whether or not he \textit{really} had a network of beliefs and desires similar to ours which our constitutive principles might match or fail to match up with. So either Grandy is thinking of POH a “pragmatic constraint” which means that it is not a necessary condition on interpreting a speaker, or, alternatively, there is no independent question about whether such simulation has been successful.

Grandy also sees POH as having a step-by-step application, first yielding a translation manual for the speaker’s language, then determining a belief-desire theory and finally predicting his actions:

\textit{The actual use of a translation in this prediction is only one of the intermediate steps. We can translate verbal behaviour into our own language and use this to determine what the person’s beliefs and desires are, and then use that information to predict actions.}\textsuperscript{20}

What struck Davidson was that none of these tasks can be completed independently of the others. They come as part of a complementary package. On Davidson’s view, if one could translate a speaker prior to discriminating amongst his intentions and beliefs then one would not need to apply DPOC in the first instance. Grandy does not directly challenge Quine and Davidson’s suggestion that the translator’s task cannot be discharged without applying some rationalising interpretive principle.

I have suggested that Grandy’s POH does not enjoy the advantages over DPOC that have been assumed. Moreover, arguments for alternative principles of translation, like POH, which assume we can translate a speaker’s words before going on to work out what he believes, desires and so on, do not get to the heart of the problem with which Quine and Davidson were concerned. Maybe Grandy did not think that the tasks of translation and theorising about beliefs and desires are really independent as he seems to suggest, but

\begin{itemize}
\item \textsuperscript{19} Grandy, (1973, p.443)
\item \textsuperscript{20} Grandy, (1973, p.442)
\end{itemize}
then there is the question of in what sense this could be a genuinely step-by step process with a POH applied at each stage, which is what he describes.

4. Conclusions
I have argued that DPOC is a substantively different principle to Quine’s POC. At the very least, it requires its own careful consideration because it is not vulnerable to the arguments typically raised against Quine’s POC and taken to motivate POH. Moreover, the role that Grandy foresaw for POH, and for interpretive principles more generally, requires further consideration. It is clear that Grandy saw his POH as playing a wider role in a predictive theory of human behaviour, though it is unclear that Grandy saw POH as a constitutive principle of meaning and belief in the way that Davidson argued DPOC is.

Acknowledgements
My thanks for helpful discussion and comments go to Guy Longworth, Will McNeill, Simon Riches, Barry Smith and two anonymous referees for this journal.

References
-----, 2001a, Subjective, Intersubjective, Objective, Oxford University Press, Oxford

