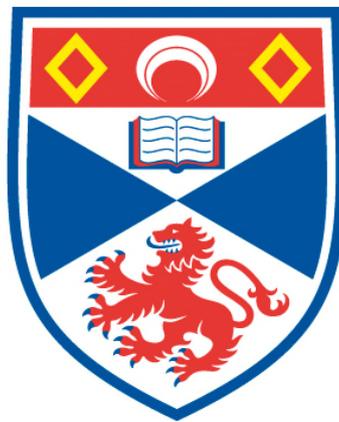


# DAVIDSON, INTERPRETATION AND VALUES

Simon Payne

A Thesis Submitted for the Degree of MPhil  
at the  
University of St Andrews



2014

Full metadata for this item is available in  
St Andrews Research Repository  
at:

<http://research-repository.st-andrews.ac.uk/>

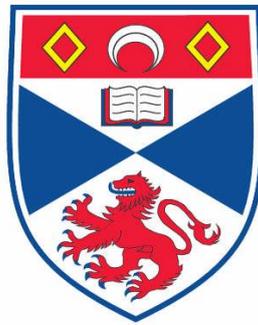
Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/9452>

This item is protected by original copyright

# Davidson, Interpretation and Values

Simon Payne



This thesis is submitted in partial fulfilment for the degree of MPhil  
at the  
University of St Andrews

02/06/2014



### 1. Candidate's declarations:

I, Simon Payne, hereby certify that this thesis, which is approximately 35000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in [month, year] and as a candidate for the degree of MPhil in September 2010; the higher study for which this is a record was carried out in the University of St Andrews between September 2011 and June 2014.

~~(If you received assistance in writing from anyone other than your supervisor(s):~~

~~I, ....., received assistance in the writing of this thesis in respect of [language, grammar, spelling or syntax], which was provided by .....~~

Date 2/6/2014    signature of candidate .....

### 2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of MPhil in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 29.05.14    signature of supervisor .....

### 3. Permission for publication: *(to be signed by both candidate and supervisor)*

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

#### PRINTED COPY

a) No embargo on print copy

~~b) Embargo on all or part of print copy for a period of ... years (maximum five) on the following ground(s):~~

- ~~• Publication would be commercially damaging to the researcher, or to the supervisor, or the University~~
- ~~• Publication would preclude future publication~~
- ~~• Publication would be in breach of laws or ethics~~

~~c) Permanent or longer term embargo on all or part of print copy for a period of ... years (the request will be referred to the Pro-  
Provost and permission will be granted only in exceptional circumstances).~~

#### Supporting statement for printed embargo request:

#### ELECTRONIC COPY

a) No embargo on electronic copy

~~b) Embargo on all or part of electronic copy for a period of ... years (maximum five) on the following ground(s):~~

- ~~• Publication would be commercially damaging to the researcher, or to the supervisor, or the University~~
- ~~• Publication would preclude future publication~~
- ~~• Publication would be in breach of law or ethics~~

~~c) Permanent or longer term embargo on all or part of electronic copy for a period of ... years (the request will be referred to the  
Pro-Provost and permission will be granted only in exceptional circumstances).~~

#### Supporting statement for electronic embargo request:

Date 2/6/2014    signature of candidate .....

signature of supervisor .....

*Please note initial embargos can be requested for a maximum of five years. An embargo on a thesis submitted to the Faculty of Science or Medicine is rarely granted for more than two years in the first instance, without good justification. The Library will not lift an embargo before confirming with the student and supervisor that they do not intend to request a continuation. In the absence of an agreed response from both student and supervisor, the Head of School will be consulted. Please note that the total period of an embargo, including any continuation, is not expected to exceed ten years. Where part of a thesis is to be embargoed, please specify the part and the reason.*

**Acknowledgments:**

I would like to thank my supervisor Adrian Haddock for helping me work through, develop and articulate my ideas. Adrian has always been generous with his time, insightful with his comments and patient with his criticism. I have learned a great deal from working with him.

I would also like to thank Peter Sullivan and Simon Hope for giving up their time at various stages over the last few years to read my work and provide me with feedback.

I have benefitted greatly from both formal and informal discussions with David Silverman, Pat McDevitt, Brian Ho and Alex Stathopoulos, who have always been happy to discuss philosophy with me. Also, and crucially, they have been equally happy not to discuss philosophy with me. I would like to thank Pat in particular for his efforts in reading drafts of my work and providing thoughtful feedback.

Above all, I would like to thank Jill McMaster for her love and support. Without Jill I would not have been able to start this project never mind finish it.

# **Davidson, Radical Interpretation and Values**

**Contents:**

**Abstract:**

**Introduction:**

**Chapter 1: Radical Interpretation**

**Chapter 2: Interpretation and Values**

**Chapter 3: Interpretation and Value Properties**

**Chapter 4: Shared Values**

**Conclusions:**

**Bibliography:**

## ***Abstract:***

The broad aim of this research is to outline, assess and elaborate on Davidson's work on radical interpretation and its connections to his understanding of values, and particularly on his understanding of the nature of ethical judgments and concepts.

The central idea that I consider is that the principle of charity must play much the same role in the attribution of values to a speaker as it does in the attribution of beliefs. This is taken to show that certain general claims about the content of propositional attitudes, which Davidson thinks follow from the consideration of radical interpretation, can be applied to values as well as beliefs.

## ***Introduction:***

Donald Davidson has maintained that, despite the manifest difficulties, no satisfactory understanding of ethics (or value generally) can fail to accommodate our lively conviction that ethical claims are objectively either true or false. Intriguingly, he has also suggested that the consideration of radical interpretation can show that this must be so. Davidson's thoughts in this area are set out primarily '*Problems of Rationality*' (Davidson 2004) which is the fourth volume of his collected essays. Many of these papers however have important connections to ideas most clearly set out in papers found in other volumes, particularly in '*Inquiries into Truth and Interpretation*' (Davidson 2001b).

Given the influence of Davidson's philosophy generally there is a perhaps surprising dearth of critical work on this aspect of his work. The broad aim of my work is to outline, assess and elaborate on Davidson's work on radical interpretation and its connections to his understanding of values, and particularly on his understanding of the nature of ethical judgments and concepts.

In Chapter 1 I introduce Davidson's work on radical interpretation, setting out the aims of the work and detailing the nature and scope of the constraints that emerge from reflection on the role of the principle of charity in interpretation. I emphasise two distinct components of the principle of charity, which Davidson has discussed under the titles of the principle of correspondence and the principle of coherence. I then go on to introduce certain issues that arise when we consider in detail the potential implications of Davidson's account of radical interpretation.

In Chapter 2 I will consider how Davidson's work on radical interpretation is connected to his remarks about the nature of values, and what its implications might be for ethics in particular. The central idea is that the principle of charity must play much the same role in the attribution of values to a speaker as it does in the attribution of beliefs. This is taken to show that a certain general presumption about the content of propositional attitudes, which Davidson thinks follows from the consideration of radical interpretation, can be applied to values as well as beliefs. I will argue that there is prima facie plausibility to the idea that the role of the principle of charity in the attribution of values to a speaker will be similar to the role of the principle of charity in the attribution of beliefs to the speaker. However I will suggest there are important limitations to Davidson's approach as it is set out, particularly with regards to certain claims that he wants to make about ethical judgements.

In Chapter 3 I develop the idea that in order to grant Davidson's general presumption we must accept that there are a class of 'basic beliefs' which must have some true instances because their content is determined by what causes them. Furthermore, in order to make substantial claims about ethical judgments on the basis of this general presumption, it looks as if it needs to be shown that ethical properties are such that they can feature amongst basic beliefs. I argue that nothing in Davidson's account of radical interpretation seems to secure this point outright. I then argue that the nature of ethical properties should make us wary of this idea more generally. Specifically, my case against Davidson here turns on worries about whether ethical properties can be akin to secondary qualities.

In Chapter 4 I will consider what I regard as a substantial elaboration of Davidson's general approach that can be found in the work of Susan Hurley. Hurley argues more specifically that any attribution by an interpreter of desires and preferences to a speaker is inextricable from the attribution of some ethical values. I then look at the more specific question of whether or not the argument from interpretation that Hurley presents could show that all speakers, as Bernard Williams has put it, must share in some more or less determinate form 'the same materials of an ethical life'. I conclude by suggesting that the sort of problems I raise for Hurley can be shown to be problems for the approach more generally.

## ***Chapter One: Radical Interpretation***

Radical interpretation is first mentioned by Davidson in his 1967 essay '*Truth and Meaning*'. The central aim of this work was to show how to provide a compositional theory of meaning for a natural language. Davidson's proposal was to employ a truth-theory, in the sense of a finite axiomatic theory characterising a truth predicate for a language in the style made famous by Tarski, to do the work of a theory of meaning. Davidson recognised that in order for it to be applied to any natural language such a theory must also be an empirical theory, which could be confirmed for particular speakers on the basis of their behaviour in an environment shared with the interpreter. Davidson's general strategy was to embed the formal structure of a theory of meaning (the structure he found in a Tarskian truth theory) within a more general theory of interpretation. Consideration of radical interpretation would serve to illustrate how a theory of meaning might pass the requirement of empirical verifiability. Davidson found the basic model for such a theory set out in chapter two of Quine's important book '*Word and Object*' (Quine 1960, 26-79).

In his 1973 essay '*Radical Interpretation*' Davidson posed two questions: firstly, what would it be sufficient for an interpreter to know in order to understand the speaker of a foreign language, and secondly how could an interpreter come to know it? He suggests that 'a theory of truth, constructed more or less along the lines of one of Tarski's truth definitions, would go a long way towards answering the first question' (Davidson 2001b, 179). His account of radical interpretation was to provide his answer to the second question. Elaborating on this account became a prominent focus of much of Davidson's work from this point onwards.

It is this account of radical interpretation that I am primarily interested in. Ernest Lepore and Kirk Ludwig have suggested, correctly it seems to me, that the fundamental idea which shapes and motivates much of Donald Davidson's later philosophy is that the stance of the radical interpreter is conceptually basic in understanding meaning and psychological attitudes (cf. Lepore and Ludwig 2005, 147-151). According to Davidson the content of any thought or propositional attitude is to be understood in part in terms of constraints that emerge from the consideration of the process of radical interpretation.

In what follows I will introduce Davidson's work on radical interpretation, setting out the aims of the work and detailing the nature and scope of the constraints that emerge from reflection on the role of the principle of charity in interpretation. I will emphasise two

distinct components of the principle of charity, which Davidson has discussed under the titles of the principle of correspondence and the principle of coherence. I will go on to introduce certain issues that arise when we consider in detail the potential implications of Davidson's account of radical interpretation.

### **1. The task and the procedure:**

The task of the radical interpreter is to interpret the linguistic behaviour of the speaker without reliance on any prior knowledge of the speaker's beliefs or the meaning of the speaker's utterances, which is to say without having reference to a shared history or any knowledge of the speaker's language or culture. This is what I will refer to hereafter as the radical interpretation scenario. I will also talk simply about speakers, interpreters and thinkers (just any speaker or an interpreter) rather than natives and linguists or whatever else.<sup>1</sup>

The radical interpretation scenario presupposes the existence of at least two agents, each responding to the same or similar features of a shared environment. To begin with, the interpreter has nothing to go on but the behaviour of the speaker in the shared environment. As Davidson once put it, while the interpreter cannot 'directly perceive' the speaker's propositional attitudes, he can 'attend to the outward manifestations' of these attitudes, which is to say that he can directly perceive the utterances and general behaviour of the speaker in the shared environment (Davidson 2001c, 210). Assuming that the interpreter is able to discover what the speaker thinks and means on the basis of this evidence – that is, assuming that radical interpretation is possible – Davidson concludes that there must be an intelligible relation between the evidence available and the propositional attitudes in question.<sup>2</sup>

Davidson believed that an interpreter can often recognise (or is justified in assuming) that the speaker has certain types of attitudes towards certain objects (where this means entities

---

<sup>1</sup> For the sake of simplicity only the speakers and interpreters I discuss will tend to be men.

<sup>2</sup> It is important to note that Davidson is not endorsing a crude form of behaviourism. Propositional attitudes are not to be understood as being 'nothing but' the available evidence. The point is that they must be accessible to the interpreter on the basis of 'nothing but' the available evidence in the radical interpretation scenario. The thought is simply that the postulation of unobserved phenomena must ultimately be tied to the observable phenomena (cf. Glock 2003, 268-9). Davidson certainly thinks that the interpreter's theory of interpretation for the speaker's language will have to be suggested by and tested against the observable behaviour of the speaker in the shared environment. However, as should become clear, Davidson also thinks that other constraints come into the picture. It is important to avoid any suggestion that the radical interpreter simply reads off, or in any strict sense 'derives', meanings or beliefs from physical evidence.

or events) in the shared environment. These objects include the speaker's utterances. Specifically, Davidson emphasised (initially at least<sup>3</sup>) that the interpreter can pick out when the speaker holds true a given utterance. It is important to note that to some extent the whole method rests on this assumption. As Bjorn Ramberg has pointed out, the interpreter must assume that he is observing a creature who asserts, and that he is reasonably adept at telling when those observed are engaging in the particular linguistic activity even when he has no clue what is being asserted (Ramberg 1989, 68). If a speaker holds true an utterance, this is in part because of what the speaker believes and in part because of what the sentence means. In this sense, holding true an utterance is a 'vector of two forces', belief and meaning. Granting this much, the fundamental problem that the radical interpreter faces is that he cannot assign meanings to the speaker's utterances without knowing what the speaker believes, while he cannot identify the speaker's beliefs without knowing what the speaker's utterances mean.

The interpreter must then find a way to provide a theory of belief and a theory of meaning simultaneously. There is for Davison only one way that this can be achieved. The solution to the problem of how an interpreter could simultaneously identify the beliefs of the speaker and the meanings of his utterances rests on a general principle that Davidson, following Quine, has typically called the principle of charity. The principle of charity, very broadly, directs the interpreter to read his own standards of truth and coherence into the pattern of sentences held true by the speaker. Davidson says: 'From a formal point of view, the principle of charity helps solve the problem of the interaction of meaning and belief by restraining the degrees of freedom allowed to belief while determining how to interpret words' (Davidson 2001c, 149).

In fact, charity provides a multifaceted constraint on interpretation. There are various constraints that are subsumed under the general title of charity. These constraints are regarded as structural constraints on all interpretation. Interpretation is possible only because the interpreter is forced to interpret the speaker in accordance with these various constraints. These constraints are also meant to tell us something about the nature of propositional attitudes more generally. Essentially then, charity is regarded as both a constraint on and an enabling presupposition of any propositional attitude whatsoever.

---

<sup>3</sup> In recent attempts to provide a unified theory of mind and action Davidson has in fact altered the evidential basis of radical interpretation from holding a sentence true to preferring a sentence true (c.f. Davidson 2004, 19-38).

The principle of charity has received many formulations in Davidson's work, and many more in the work of his commentators and critics. Increasingly in later works Davidson has felt the need to emphasise two distinct components which fall under the broad title of the principle of charity. The content of the speaker's propositional attitudes is in part constrained by causal relations that exist between himself, the speaker and the shared environment, and in part constrained by holistic connections with other propositional attitudes. Davidson has discussed these components of charity under the headings of the principle of correspondence and the principle of coherence respectively (cf. Davidson 2001c 211, and Davidson 2005, 44-45). In what follows I will consider these two components of the broader principle of charity in turn. It is however important to emphasise from the start that while the principle of correspondence and the principle of coherence may seem to involve different types of constraints on interpretation they are understood to operate in tandem.

## **2. The Principle of Correspondence:**

The principle of charity directs the interpreter to read some of his own standards of truth and consistency into the pattern of sentences held true by the speaker. The principle of correspondence can be taken to emphasise the point that charity, as employed by the radical interpreter, must be applied (initially at least) specifically to the speaker's beliefs about his immediate environment – not simply general beliefs, but specific beliefs. That is to say that the interpreter must assume that the speaker's beliefs about his immediate environment are largely true. This point plays an essential role in the interpreter's initial solution for the problem of solving for the meanings of the speaker's utterances and for the speaker's beliefs in the radical interpretation scenario. It is what allows the interpreter to bridge the gap between noticing correlations between the speaker's attitudes, utterances and his environment, and assigning meanings to the speaker's sentences.

Which sentences a speaker holds true depends on what the speaker thinks the sentences mean and what he believes. If the interpreter knew either he would be in a good position to assign the other. The basic idea is this: it is only by assuming that what the speaker believes is true, in the light of the conditions in which he is in, that the interpreter is able to solve for the belief contents, and then assign the corresponding meaning to the utterance. This thus represents a methodological constraint on radical interpretation: the interpreter is constrained to start by assuming that the speaker's beliefs about his immediate surroundings are by and large true. It is only by restraining the degree of freedom allowed to the speaker's beliefs in this way that the interpreter can proceed.

Davidson has often discussed this aspect of the interpretive process in terms of ensuring that the interpreter respects the causal connections that link himself, the speaker and the shared environment. Here is a fairly crude, but not I think unfairly crude, outline of how this is supposed to work. When an interpreter interprets a speaker, he must assume that the objects of the speaker's most basic beliefs (where again this can be any entity or event) are what causes them. The only way for the interpreter to determine what those objects are is to identify objects common to both, and assume that what the speaker is caused to believe is basically similar to what the interpreter himself is caused to believe by the same objects. For this process to be effective (for it to lead to interpretation), when an interpreter interprets the beliefs of a speaker it is a necessary presumption (often overridden by other considerations) that similar causes are followed by similar responses for both the interpreter and the speaker.

There are two distinct assumptions in play here. Firstly, the interpreter is prompted to assume that the speaker is responding to the same features of the world as the interpreter himself is. Secondly, the interpreter is prompted to assume that the speaker responds to these features of the world *in more or less the same way* as the interpreter himself does. Both of these assumptions are essential. Davidson goes as far to say that unless we assume that we can identify and share reactions to common stimuli 'thought and speech would have no particular content – that is no content at all. It takes two points of view to give a location to the cause of a thought, and thus to define its content' (Davidson 2001c, 214).

As well as insisting of what I have called a methodological constraint, Davidson has argued for what looks like a much stronger claim. Not only is it the case that the interpreter must assume that most of the speaker's beliefs about his surroundings are largely true (this is the methodological claim) but it is also the case that most of the speaker's beliefs about his surroundings must in fact be true (this is the stronger claim). This claim seems to be about the nature of thought more generally, but Davidson thinks it follows from the consideration of radical interpretation. More precisely, the claim is about the nature of perceptual beliefs. According to Davidson, that nature of interpretation 'guarantees that a large number of our simplest perceptual beliefs are true, and that the nature of these beliefs is known to others' (Davidson 2001c, 204).

Davidson's view is that the content of any thought or utterance is partly determined by a history of causal interactions between speakers and the environment. He has remarked that is 'a commonplace of the empirical tradition' that any speaker will learn his first words and

basic sentences ('Red!', 'Mamma!', 'Fire!' etc.) through 'a conditioning of sounds or verbal behaviour to appropriate bits of matter in the public domain'. Crucially though, for Davidson this is not just a story about how a speaker learns to use his first and most basic words, but it is also 'an essential part of an adequate account of what words refer to, and what they mean'. This is to say that the contents of any speaker's earliest and most basic sentences must be determined by whatever it is in the world that typically causes him to hold them true. It is at this point that the fundamental ties between language and the world are established, and certain central constraints on meaning are fixed. For Davidson then what ensures that a speaker's view of the world is, in its plainest features, largely correct 'is that the stimuli that cause our most basic verbal responses also determine what those responses mean, and the content of the beliefs that accompany them' (Davidson 2001c, 204). Further, Davidson thinks this can be shown by 'appealing to obvious facts about language learning and facts about how we interpret words and languages with which we are unfamiliar' (Davidson 2001c, 43-45).

Of course, an interpreter may find a speaker who has lots of true beliefs about wolves say, but who has never actually encountered a real wolf before. The content of these beliefs about wolves has not been determined by direct exposure to wolves, but by the possession of other words and concepts, such as those of a mammal or a pack animal, and perhaps some exposure to pictures of wolves in books. The point however is that *somewhere along the line* we must come to direct exposures that anchor thought and language to the world. Davidson claims that 'such direct conditionings of words and objects *must* lie at the basis of interpretation: and if this is so, correct interpretation makes a speaker believe a lot of true things about what exists' (Davidson 2005, 45).

Davidson was well aware that this sort of view faces a number of obstacles. I will briefly set out two potential issues. Firstly, there is what Davidson has referred to as the problem of error. The problem of error arises due to the fact that even in the most straightforward of cases, it is obvious that the same cause (a rabbit scampers by) may engender very different beliefs in the speaker and the interpreter 'and so encourage assent to sentences which cannot bear the same interpretation' (Davidson 2001c, 152).

Secondly, Davidson seems to accept that claims about the direct conditioning of sentences and objects can only apply directly to some sentences. They seem to apply most directly to what Quine referred to as occasion sentences. Occasion sentences are sentences that all or most speakers assent to (or hold true) in response to some same sensory stimulation in the

environment. It would seem however that the interpreter must move from such sentences to understand sentences which are not like this – such as sentences about the past or about the future, about people or objects that are not present, or which seem to be about more abstract properties or concepts. There are two things to be said here. Firstly, it is far from straightforward how an interpreter is supposed to arrive at the attribution of such sentences from a starting point of occasion sentences. Secondly, we might think that there are reasons to acknowledge a fundamental distinction between the epistemological status of occasion sentences and other sorts of sentences. Quine for example certainly gave epistemic significance to the distinction between occasion sentences and others.

Davidson was very keen to avoid any such distinction. Despite his references to the importance of ‘direct conditioning’ between the world and the content of propositional attitudes,<sup>4</sup> Davidson, unlike Quine, has repeatedly played down the significance of some privileged range of sentences which relate directly to stimuli in the surrounding area.<sup>5</sup> But as well as raising a point about epistemology, these remarks relate to a basic distinction between Davidson’s approach to radical interpretation and Quine’s. While Quine restricted charity to the translation of occasion sentences and logical constants, Davidson advocated the application of charity ‘across the board’. That is to say that he has instead focussed on the way that the radical interpreter must work his way into a whole block of propositional attitudes by following a much more general policy of finding the speaker’s thoughts, utterances and behaviour to be generally rational and coherent.<sup>6</sup> I will say more about how the very nature of these propositional attitudes constrains interpretation in a moment.

Before I move onto this I will quickly address one rather odd but nevertheless persistent misconstrual of Davidson’s work. Davidson has argued that consideration of radical interpretation can secure a guarantee that most of any interpreter or speaker’s basic perceptual beliefs about the world must be correct. Some commentators have accused Davidson of claiming far too much here. Certainly, to suppose that a speaker must be right about *all* of his basic perceptual beliefs would be, to borrow a phrase from Ramberg, to engage in a strange form of epistemological occultism (Ramberg 1989, 69). But the principle

---

<sup>4</sup> In fact, Davidson’s argument that a large number of our basic perceptual beliefs must be true sometimes looks as if it is meant to be justified by his remarks about direct conditioning.

<sup>5</sup> Davidson does say however retain the distinction between ‘sentences whose causes to assent come and go with observable circumstances and those a speaker clings to through change’ and remarks that this distinction ‘offers the possibility of interpreting the words and sentences beyond the logical’ (Davidson 2001c, 149).

<sup>6</sup> Davidson’s change of emphasis does in fact seem to be epistemologically motivated. Cf. Davidson and Blackburn.

of charity does not and was never intended to rule out the thought that the speaker may make mistakes. The point, as with propositional attitudes considered more generally, is that the only incentive for attributing an error is contradictions with previously interpreted sentences. As interpretation develops, the interpreter may have good reasons to revise some of his initial attributions, including attributions of perceptual beliefs, and to maintain that some of the speaker's beliefs and utterances are false or unreasonable.

The principle of correspondence then can be seen to represent a methodological constraint on radical interpretation. The interpreter is constrained to start by assuming that the speaker's beliefs about his immediate surroundings are by and large true. It is also part of a bigger claim that most of any speaker's basic perceptual beliefs must be true. Davidson has always emphasised other constraints on the content of sentences, which I will now discuss. What a sentence means depends 'partly on the circumstances that cause it to win some degree of conviction'. It also depends 'partly on the relations, grammatical and logical, that the sentence has to other sentences held true with varying degrees of conviction'.

### **3. The Principle of Coherence:**

The principle of correspondence emphasises the need to ensure that the interpreter respects the causal connections that link himself, the speaker and the shared environment. It is taken to show that the interpreter must assume that the speaker's beliefs about his environment are generally true. It is also connected to the further claim that most of any speaker's perceptual beliefs must in fact be true. The principle of coherence emphasises the need for the interpreter to endow the speaker with a degree of rational consistency. It shows that the interpreter must assume that the speaker's propositional attitudes in general are largely consistent and rational.

The necessity of the requirement that the interpreter find the speaker to be largely rationally coherent is intimately connected to Davidson's holistic understanding of propositional attitudes more generally. Propositional attitudes are only identified/individuated in part according to their relations to other propositional attitudes.<sup>7</sup> For the interpreter, this means that the attribution of any belief, say, to a speaker requires the attribution of many related beliefs. This general holistic point is straightforward enough. Suppose the interpreter attributes to the speaker the belief that 'there is a black snake in the grass'. The interpreter must assume that the speaker has an idea of what a snake is.

---

<sup>7</sup> Cf. Davidson 2001b, 168: 'a belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about'.

They must assume that the speaker believes such things as: a snake is an animal, it has no feet, this particular snake is black and not green, it is in the grass and not somewhere else, and so on. The principle of coherence shows how the interpreter is constrained to attribute a considerable number of beliefs with related content to the speaker if he is to attribute any at all.

Davidson has used considerations of coherence to support the claim that most of the beliefs a speaker holds true must be true, at least in the opinion of the interpreter. This is not simply a version of the claim that was sketched in the previous section – that was the claim that most of the speaker's *basic perceptual* beliefs must be assumed true by the interpreter. This point is meant to apply to the beliefs of the speaker more generally. Davidson has often made this claim by highlighting the limits of any incoherence that can intelligibly be attributed to a speaker. Suppose that the interpreter found that many of the speaker's beliefs about what he (the interpreter) called snakes were false. (There is evidence that the speaker thinks that snakes have feet). In such a case, it would be hard to see in what sense the interpreter could hold that the initial attribution of the belief that there was a black snake in the grass could be correctly described as being about a snake at all. Thus the belief that there is a black snake in the grass, whether true or false, depends on 'a background of true beliefs, true beliefs about the nature of snakes, of animals, of physical objects in the world' (Davidson 2004, 16). The interpreter could not attribute a true belief that there was a snake in the grass, while at the same time attributing wholly false beliefs about snakes. This can be generalised to the claim that the interpreter could not attribute any belief to a speaker without attributing lots of other true beliefs (true in the interpreter's opinion) with related content.

There are a number of important caveats to these general claims about the holistic nature of propositional attitudes. The failure to appreciate these points has resulted in a significant amount of misunderstanding. Firstly, Davidson does not endorse what might be called 'rigid holism'. By rigid holism I have in mind the view that for a speaker to be regarded as having any thought with a particular content, there must be some precise list of other thoughts with particular contents that the speaker also has to have. Davidson has in fact always been very careful to deny that there must be a precise list of things that a speaker must believe in order for them to believe for example, that there is a black snake in the grass. Any list of connected beliefs would no doubt be very large, but its makeup would be indefinite. The point applies quite generally. The interpreter would be unable to attribute the belief that there is a black snake in the grass to a speaker without also attributing many of the other

beliefs I mentioned above (that a snake is an animal etc.). But, crucially, most of these associated beliefs *taken individually* could turn out to be held false by the speaker, or not held at all. Plausibly, a speaker could have lots of true beliefs about black snakes without actually having the conceptual resources to differentiate between green and black.

Some connections among beliefs are more significant than others. Davidson illustrates this with the following example: my belief that it is raining today probably contributes essentially nothing to the content of my other beliefs about rain except those that are logically related, while my belief that rain is caused by the condensation of drops in water-saturated air contributes a great deal (cf. Davidson 2004, 15). But this is not to say that the interpreter must assume that the speaker is aware of all of the logical consequents of any given belief. It is possible for a person to believe something and fail to believe a logical consequence. Evidence might even conceivably push the interpreter to attribute a direct contradiction to the speaker.<sup>8</sup>

Neither, it should go without saying, does Davidson endorse what Lepore and Ludwig have called 'extreme holism'. Extreme holism is the view that a speaker can have any particular attitude content if and only if he has exactly all the other related attitudes he in fact has, so that any change in the content of one attitude would entail a change in the contents of every other attitude. Again, Davidson's expressed views in fact very clearly reject any such suggestion.<sup>9</sup> He has in various places stressed that while a change in one belief may necessitate changes in others (beliefs that are recognised as being directly tied by logic to the altered belief may well change) in general these changes would leave most of our overall belief structure untouched.

There is an important point about interpretation here. The prudent interpreter must 'strive to conserve as much of the existing beliefs as he could when adjusting for new evidence' (Davidson 2004, 14). There is also an important point about Davidson's holistic understanding of the mind more generally. Holism should not be thought to entail that everything a speaker believes is in a constant flux due to the input of new information or the

---

<sup>8</sup> Davidson could be criticised on the grounds that, having given up the analytic/synthetic distinction, he has no way of distinguishing between the relations which define a state of mind (or the meaning of the utterance) and those that are 'merely' contingent, and so do not touch content. Davidson has remarked that 'it is well to remember that giving up the idea of a firm line between the analytic and the synthetic does not mean giving up the idea of a continuum in which some connections among thoughts are far more important to characterising a state of mind than are others' (Davidson 2004, 15).

<sup>9</sup> For an example of Davidson's explicit denials here, see Davison 2004, 11-18. Lepore and Ludwig make this point at Lepore and Ludwig 2005, 211-3.

as a result of recent reflections. Davidson acknowledges that much may change from moment to moment, but maintains that 'serious changes in our world outlook, ambitions, and tastes are for the most part glacially slow'. So while any given change in a belief may call for change in some other beliefs, it may have very little influence on any others, and particularly on what matters most to a speaker. Should I change my mind about whether a friend of mine is wearing a scarf (she seemed to be, but now I see that it is just an extension of her dress) this will not alter any of my most basic beliefs about my friend or about scarfs, never mind anything else. As Davidson has put it: 'The importance of holism rests only in small part on its flow. Its real importance rests on the fact that the content of any given attitude depends on its place in the whole network' (Davidson 2004, 15).

Davidson distinguishes between two types of holistic constraints that restrict the interpreter's attempts to individuate the propositional attitudes of the speaker. The first Davidson has called 'intra attitudinal holism' (Davidson 2004, 13). This category concerns the relations among beliefs within the category of belief for example, or desires within the category of desires. The basic norms of intra-attitudinal holism are those formulated in classical predicate logic or first order logic. One of the ways by which beliefs are individuated is by their relations to certain other beliefs (beliefs with related content). When these relations are limited to obvious logical relations, it is fairly clear what Davidson has in mind. The idea is that the interpreter must look for the best ways to 'fit' his own standards of logic onto the utterances and beliefs of the speaker. An example may be useful here. Suppose that the interpreter identifies a connective that creates a sentence out of two sentences, such that the interpreter always (or perhaps almost always) assents to the compound sentence when and only when he assents to each sentence alone. In this case, the interpreter can do no better than to treat that connective as the sign for a conjunction in the speaker's language.

The interpreter then must read the logical structures of 'first order quantification theory (plus identity)' into the language of the speaker. Further, this will not mean simply taking the logical constants one by one, but '*treating this much logic as a grid to be fitted on to the language in one fell swoop*' (Davidson 2001a, 136). Davidson says: 'The point is that by interpreting by the only standards of interpretation available to me, I have, on a primitive level, made the speaker I am interpreting a good logician (by my own norms of reasoning, it should go without saying; I have no others)...With respect to the simplest and plainest logical matters, a sharing of norms of rationality is an inescapable artefact of interpretation' (Davidson 2004, 49-50).

This feature of Davidson's account of interpretation follows directly from the requirement that interpreter understand the thoughts and utterances of the speaker in accordance with something like a Tarski-style truth theory. This leads to various formal constraints on the interpretation of the speaker. It is a necessary constraint on any interpretation that the interpreter find, for example, a recursive structure in the speaker's language<sup>10</sup>. The exact nature of such constraints is more difficult to make out. We can easily imagine that a speaker's language could be more impoverished in terms of both semantics and syntax than the interpreter's own. Davidson in fact suggests at one point that we could imagine a language entirely without imperative or interrogative moods (cf. Davidson 2004, 54). In such cases, should the interpreter simply read his own sophisticated semantics and syntax into the utterances of the speaker (that is, treat it 'as a grid to be fitted onto the language in one fell swoop') this could lead to gross distortions. Certainly however Davidson regards it as a fundamental constraint on interpretation that the interpreter must find the norms of elementary logic have application in the beliefs of those he is interpreting.

The second form of holistic constraint on interpretation Davidson has called 'inter-attitudinal holism'. This category concerns the relations between the different types of propositional attitudes, between for example beliefs and desires, or between beliefs, desires and intentions. While attention to intra-attitudinal holism might be thought to emphasise the idea that attributing any belief to a speaker will also require the attribution of a cluster of other beliefs with related contents, attention to inter-attitudinal holism emphasises that, for example, attributing any desire to a speaker will also require the attribution of various beliefs and intentions with related contents. This emphasis on inter-attitudinal holism plays a crucial role in the development of the idea that not only must the interpreter find the speaker's basic perceptual beliefs to be largely true, but that the interpreter must find the speaker to be rational and intelligible overall. It also what enables Davidson to articulate his view that, to a large extent, the propositional attitudes come as a set.

The general idea can be brought out by considering certain relations amongst the various evaluative attitudes, and belief and action. Few I think would deny that the interpreter must understand the speaker as having certain evaluative attitudes; wants, desires, preferences etc. Furthermore, it seems highly plausible that the interpreter must assume that these evaluative attitudes will interact with the speaker's other evaluative attitudes and with the

---

<sup>10</sup> For Davidson's arguments that this is a necessary condition of any learnable language, see Davidson 2001b, particularly essays 1 and 2.

speaker's beliefs and actions in a fairly structured way – otherwise *they would not be* wants, desires, preferences etc. An interpreter would struggle, for example, to understand a speaker who sincerely claimed to prefer A to B, but who whenever faced with a free choice between A and B consistently chose B.<sup>11</sup> This is in part because preferences are understood to relate to actions in a certain way.

It is important to notice the full scope of this constraint when it comes to radical interpretation. The propositional attitudes are intimately connected, and come as a set.<sup>12</sup> Most (if not all) of a speaker's desires will depend on his beliefs. The interpreter could not attribute the desire to eat the contents of the dish to the speaker without attributing a variety of beliefs about food generally and the particular food in question. The point can also be applied to propositional attitudes generally. There are many other kinds of attitudes, such as hope, intention, despair, expectation etc. which depend on beliefs to give their content substance. Davidson says: 'Many of the attitudes, like being pleased, proud, or angry that something has happened, depend on the true belief that it has occurred. We cannot be worried lest something will befall us unless we think it may, or hopeful that we will win a prize unless we know, or at least believe, we may' (Davidson 2004, 16).

The most basic norms of inter-attitudinal holism are closely linked in Davidson's work to subjective decision theory, as developed by Frank Ramsey and Richard Jeffery among others (Davidson 2004, p151-166) and to his understanding of rationality more generally. Ramsey and Jeffery both invoke structural rationality constraints built into the very notions of preference and degrees of belief. Davidson has sketched how such constraints can be employed in radical interpretation<sup>13</sup>. Piers Rawling highlights two parallels between the projects which make the link attractive. 'Firstly, degrees of belief necessarily fit a rational

---

<sup>11</sup> Assuming of course that there is no further preference C which dictates the choice between A and B.

<sup>12</sup> Davidson says: 'The propositional attitudes provide an interesting criterion of rationality because they come only as a matched set. It may sound trivial to say that a rich pattern of beliefs, desires and intentions suffices for rationality; and it may seem far too stringent to make this a necessary condition. But in fact the stringency lies in the nature of the propositional attitude, since to have one is to have a full compliment. One belief demands many beliefs, and beliefs demand other basic attitudes such as intentions, desires...This does not mean that there are not borderline cases. Nevertheless, the intrinsically holistic character of propositional attitudes makes the distinction between having any and having none dramatic' (Davidson 2001b, 96).

<sup>13</sup> In more recent attempts to integrate the theory of meaning into a unified theory of mind and action, Davidson has in fact altered the evidential basis of radical interpretation from holding a sentence true to preferring a sentence true (c.f. Davidson 2004, 19-38). Davidson insists that the unified theory includes rather than replaces the old project of radical interpretation. But he has increasingly stressed that a theory of interpretation and decision theory must be combined (cf. Glock 2003, 177).

pattern. The interpreter need not insist on perfect coherence in order for a speaker to qualify as having degrees of belief, but massive incoherence would indicate that he was tracking the wrong feature. Secondly, this rational pattern ensures that degrees of belief are accessible from without: we can determine them from observable behaviour, which is of course the raw data of radical interpretation' (Rawling, in Ludwig ed. 2003a, 87).<sup>14</sup>

Taken together, these holistic considerations regarding propositional attitudes are supposed to show that it will be a necessary condition of any interpretation that the interpreter finds a considerable degree of rational consistency in the speaker's utterances and beliefs.

Attributions of propositional attitudes must be consistent with each other and with an overall theory of behaviour. There are limits as to how much logical inconsistency or irrationality an interpreter can intelligibly ascribe to a speaker. The rationality constraint forces the interpreter to endow the speaker with a whole range of propositional attitudes when making sense of their utterances and behaviour. Davidson has summed up this point by remarking that if an interpreter cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs and attitudes largely consistent and true by *his own* standards, he will 'have no reason to count that creature as rational, as having beliefs, or as saying anything' (Davidson 2001b, 137). Charity, on Davidson's account, does then not simply constrain the interpreter to find the speaker correct with regards to his basic perceptual beliefs and with regards to logical constants – charity is applied across the board, which means that the interpreter must read his standards of truth and coherence into the propositional attitudes of the speaker quite generally.

#### **4. The Right Sort of Agreement:**

The principle of charity is taken to show how the propositional attitudes that the interpreter attributes to the speaker are constrained in part by an assumption of causal relatedness between those propositional attitudes and the objects of those attitudes, and in part by an assumption of rational coherence amongst the speaker's propositional attitudes. These constraints make the propositional attitudes of the speaker the propositional attitudes they are. In his early formulations of the principle of charity, Davidson emphasises the idea that the principle requires the interpreter to 'maximise agreement' with the speaker, with regards to both truth and overall coherence. However, Davidson has openly acknowledged that this way of putting things can be problematic. He has said: 'I apply the Principle of

---

<sup>14</sup> I will say more about the connection between radical interpretation and decision theory, and look in detail at what might be involved in attributing desires and preferences at all in Chapter 4.

Charity across the board. So applied, it counsels quite generally to prefer theories of interpretation that minimise disagreement. So I tended to put things in earlier essays, wanting to stress the inevitability of the appeal to charity. But minimising disagreement, or maximising agreement, is a confused ideal' (Davidson 2001b, xix).

It is a confused ideal for two fairly straightforward reasons. Firstly, as the number of sentences in any natural language that a speaker might utter is infinite, there is in one sense no real way in which there can be a maximum of agreed sentences (cf. Davidson 2001b, xix). It certainly makes little sense to speak of counting thoughts or propositional attitudes due to our ability to combine a limited repertoire of concepts in a potentially infinite number of ways. Consider the following example. Suppose I can have at least one belief, say that my chair rotates, and also that I have the concept of a negation. Then it seems that I already have an infinity of possible (though of course massively trivial) beliefs: my chair rotates, it is not the case that my chair does not rotate, etc. Other connectives, such as conjunction and alteration, will 'add to the infinities; the possibility of predicating redness or solidity to any of an endless number of items swells the list, and so on' (Davidson 2004, 12).

Secondly, there are good reasons to suppose that a high number of false beliefs which accompany the attribution of a particular belief need not necessarily put a strain on that attribution. Once the interpreter ascribes to the speaker a belief in a certain superstition, say about the luck bringing qualities of severed rabbit feet, it makes sense to ascribe a whole raft of false beliefs (false in the interpreter's opinion) about severed rabbit feet to the speaker – that a severed rabbit foot will provide luck in situation  $X$ , in situation  $X_1$ , in situation  $X_{101}$  etc. There are of course limits as to how much logical inconsistency an interpreter can ascribe to a speaker while at the same time making them intelligible. This is a substantial point, and one that is I think well illustrated by consideration of the constraints on interpretation set out above. But I take the above remarks to suggest that the idea that the principle of charity recommends that the interpreter must strive to 'maximise agreement' between himself and the speaker on matters of truth and overall coherence is not as clear as it might initially have seemed.

There is a further, and I think more significant issue that emerges when Davidson talks about the agreement between the interpreter and the speaker that is required for successful interpretation. As David Lewis has famously argued (Lewis 1974), it would be wrong to ascribe a speaker with beliefs that the interpreter takes to be correct even in cases where there is no plausible explanation of how the speaker could have acquired the beliefs. This is

not the obvious point that the interpreter can attribute false beliefs to the speaker (false in the interpreter's opinion) about rabbit feet or whatever. Certainly, Davidson has always acknowledged that once the interpreter's theory of the speaker's language begins to take shape, it makes sense to accept intelligible error and to make allowances for the relative likelihood of various kinds of mistake (cf. Davidson 2001b, 196). The point is that, for example, the interpreter should not start by ascribing to a speaker beliefs about quantum physics or the rules of cricket unless it is plausible that the speaker could have such beliefs. When an interpreter applies the principle of charity across the board, he must avoid attributing unintelligible insights to the speaker as well as attributing unintelligible error. That is to say that the interpreter must maximise agreement of a certain sort.

What Lewis says here seems to be highly plausible. It would also be very uncharitable to assume that Davidson would have disagreed (in fact he didn't). But perhaps the fact that Lewis felt the need to make the point in a response to Davidson shows that some of Davidson's formulations of his position are not as clear as they might be. Davidson is essentially saying that there are limits to how much inconsistency (and falsehood) an interpreter can ascribe to a speaker. Or, to put this another way, what a speaker believes must be consistent and true for the most part, by the lights of the interpreter. Perhaps we might think that the talk of 'maximising agreement' is simply an infelicitous way of making this point.

However I think that there is more to this issue than this. Davidson continued to suggest that an interpreter must 'optimise' agreement between himself and the speaker, that the interpreter should make the speaker right when plausibly possible. He did so while at the same time recognising that it is difficult to say what exactly the right sort of agreement is. In one place he remarks that his point has always been 'that understanding can be secured only by interpreting in a way that makes for the right sort of agreement. The 'right sort', however, is no easier to specify than to say what constitutes a good reason for holding a particular belief' (Davidson 2001b, xvii). In fact, Davidson often suggests that charity urges the interpreter to maximise agreement on 'basic' or 'structurally important' propositional attitudes. But this simply defers the problem, leaving him with the task of specifying the 'basic' or 'structurally important' propositional attitudes on which we must maximise agreement. In any case, without some clear specification of the 'right sort of agreement', our understanding of the principle of charity seems to be incomplete.

There are two reasons why I think we should regard this is a problem. The first is that, if this is so, then any vagueness on this point might be expected to spread to whatever implications we want to draw from the consideration of the principle of charity. The principle of charity is deeply implicated in Davidson's arguments against the possibility of our being fundamentally mistaken about how things are (that is, his epistemological claims) and his arguments against conceptual relativism (that is, his rejection of the possibility of radically alternative conceptual schemes). When a key premise of these arguments is formulated in terms of a rough and unspecified notion of 'maximising the right sort of agreement' this is likely to result in suspicion, and justifiably so.

The second is that there seem to be a variety of different kinds of agreement and disagreement that can exist between speakers. As Bernard Williams has emphasised, there is no reason why an interpreter should not find that a great deal of disagreement exists between himself and the speaker, or that this should be alarming or even surprising at all. What seems to be important is that in different contexts disagreement will require different sorts of explanation, as will agreement. Williams says: 'The way we understand a given kind of disagreement and explain it has important practical effects. It can modify our outlook or our attitudes to others. In relation to other people, we need a view of what is to be opposed, and in what spirit. In relation to ourselves, disagreement can raise awareness that we can be wrong' (Williams 1985, 132-3).

Indeed, Williams has occasionally drawn attention to what strikes me as an extremely important form of ethical disagreement - namely the sort of disagreement that can occur between those who do use a certain ethical concept and those who don't. Whether the consideration of radical interpretation can tell us anything about how an interpreter could hope to maximise the 'right sort of agreement' in this context is yet to be shown. Given Davidson's claims about ethical objectivity this is a significant issue. What we can say about this type of disagreement is important because, as Williams has pointed out, one of the things that people want when they talk about objectivity in ethics is this: 'They want there to be one canonical, homogeneous ethical language. They want it to be conceptually homogeneous across cultures, and across disagreements within our culture' (Williams 1995a, 240).

##### **5. Necessary and Contingent Constraints:**

There is another general issue with Davidson's work on radical interpretation which I wish to raise at this stage. In response in part to the sort of consideration raised by Lewis, some of

those interested in the radical interpretation method (that is to say, some of those interested in how the consideration of radical interpretation might yield insights into the nature of belief, meaning and understanding more generally) have attempted to modify or refine the principle of charity from its earlier formulations. As Glock has remarked, the majority of these modifications are based around the thought that the interpreter should ascribe to the speaker beliefs that it is psychologically or anthropologically plausible for them to have, whether or not these beliefs coincide with the interpreter's own (cf. Glock 2003, 185-6). Such modifications have often been signalled by redrawing the principle of charity as the 'principle of humanity'.<sup>15</sup> Other advocates of the radical interpretation method have dismissed these modifications out of hand as distortions of Davidson's original project, and tended to deride the psychological or anthropological assumptions invoked as 'labour saving heuristic aides' which may guide actual interpretation but are irrelevant to the process of radical interpretation (cf. Ramberg 1989, 64). This issue marks an important divide in the literature on radical interpretation that is often overlooked.

While Davidson was always mindful to emphasise the distinction between the question of how an interpreter would actually go about understanding a speaker and the question of what would be necessary and sufficient for such understanding (cf. Davidson 2005a, 111) I am of the opinion that the precise connections between these questions are somewhat ambiguous. There is certainly some evidence that Davidson was open to the sort of concerns that have been expressed by advocates of the principle of humanity. As early as his 1973 essay '*Radical Interpretation*' Davidson was emphasising that charity was subject to 'considerations of simplicity, hunches about the effects of social conditioning, and of course our common-sense, or scientific, knowledge of explicable error' (Davidson 2001b 196). Also, in his reply to Lewis and Quine, Davidson appears to endorse something like the principle of humanity: 'The improved principle of charity (Lewis's, that is), insofar as it says there are cases where you can make exceptions *right from the beginning*, is what I espouse...We imagine that the speaker does what he did and I do what I do, and we construct his resulting beliefs, not by making his beliefs like mine, but rather by imagining what I would have believed if I had done what he did and been where he was' (Davidson 2001b, 282). Davidson here is acknowledging that the radical interpreter will be making assumptions about the psychology of the speaker, though he does go on to add the potentially important caveat

---

<sup>15</sup> According to Steven Lukes, for example, the principle of charity counsels 'Count them right in most matters' while the principle of humanity counsels 'Count them intelligible or perhaps count them right unless we can't explain their being right or can better explain their being wrong' (Lukes 1982, 262).

that 'as a roughly stated principle I feel that there is a lot to this; as a sharp principle I'm not sure that it doesn't assume too much about what must be going on, or should be going on in the speaker's head'.

How far these remarks represent a move away from Davidson's original construal of the principle of charity is an intriguing question that I do not take up here. I am more interested in the following question: What, if any, psychological or anthropological assumptions are required in order for a radical interpreter to be able to interpret a speaker? That is, what, on Davidson's account (the account that I have set out, the account that emerges from his later papers), must the interpreter assume about what is or should be going on in the speaker's head? In order to consider this issue I will borrow a distinction introduced by Simon Blackburn in his book *'Spreading the Word'* (Blackburn 1984, 277). Here Blackburn distinguishes between a 'homely' radical interpreter and a 'bleak' radical interpreter.<sup>16</sup>

According to Blackburn, a 'homely' interpreter will set out to interpret the speaker in such a way as to make him generally rational and intelligible. This will involve the interpreter seeing the speaker as having certain intelligible aims, and performing certain actions as means of furthering those aims. The general constraint to find the speaker intelligible 'may include quite specific constraints on how to interpret them – for instance, if they appear to have their attention focussed on one place or thing, whose changes are exciting them, not to regard them as talking about things they cannot see' (Blackburn 1984, 277). The homely interpreter, by exercising the principle of humanity, enters not only into the likely beliefs of the speaker, but also into the likely desires, needs and wants of the speaker. He also seems to make assumptions about what the speaker finds salient in his environment. To a significant extent then the homely interpreter seems to enter straight into the psychology of the speaker, making a variety of psychological and anthropological assumptions from the start.

The 'bleak' interpreter, by contrast, 'dislikes armchair psychologising'. The bleak interpreter attributes meanings to the speaker's utterances on the basis of a 'bare description of what is said, and of what purely natural, non-psychological features of the speaker's situation surrounded the saying of it' (Blackburn 1984, 278). The bleak interpreter can look to the natural (where this seems to mean physical or causal) relations that speakers have to things in the environment. But his evidence is presented in non-psychological terms for the bleak

---

<sup>16</sup> For my purposes here I can avoid the implication that the 'homely' or 'bleak' interpreters might represent the approaches of any actual philosophers.

interpreter will use this evidence in his attempt 'to construct a scientific semantics of the speaker's language'.<sup>17</sup> The bleak interpreter is constrained by a general policy of charity. That is he is constrained to assume that the speaker's utterances are true as far as possible. But any attempt to cash this out in terms of assumptions about the general psychology or anthropology of the speaker are out of keeping with the procedure of the bleak interpreter.

Now regardless of what we might think about the bleak interpreter, about his project or his prospects, the key point is that his approach does not seem to be completely in line with the approach to radical interpretation that Davidson sets out in his later work. This is suggested by the distinction between Davidson's work on radical interpretation and Quine's work on radical translation which was introduced earlier. Where Quine restricted the application of the principle of charity to the translation of logical constants and occasion sentences, Davidson has advocated the application of the application of the principle of charity across the board. What does he mean by this? Firstly, as was mentioned above, Davidson has increasingly played down the significance of some privileged range of sentences which relate directly to stimuli in the surrounding area. Secondly, as was also mentioned above, Davidson claims that the interpreter must find the speaker to be largely similar to himself, and that this similarity involves not just beliefs but also desires, intentions and other propositional attitudes (remember the remarks about the interpreter working his way into a whole block of propositional attitudes). Finding the speaker to be rational and intelligible seems to involve attributing a range of propositional attitudes (cf. note 11 above). The point I take it, is that if the speaker has any propositional attitudes, they have many.

It appears then that it is fair to say that the radical interpreter, on Davidson's account, must to some extent enter into the psychology of the speaker. If the radical interpreter can understand the speaker at all, he has to understand his utterances and actions in terms of a whole range of propositional attitudes – beliefs, wants, needs, preferences, desires etc. Perhaps I can put the point like this: the attribution of general rationality necessarily involves a variety of what look like, on any description, psychological assumptions about the speaker.

This is not to say however that Davidson's radical interpreter should be regarded as a homely interpreter. This is for the following reason. As Blackburn puts it: 'If we think in terms of the homely radical interpreter, we are asking how we determine the facts about

---

<sup>17</sup> As Blackburn puts it, the bleak interpreter is to be regarded as 'embodying the scientific or physicalistic truth about speakers' (Blackburn 1984, 288).

meaning, reference, and the satisfaction conditions of sentences' (Blackburn 1984, 279).<sup>18</sup> I take it that the 'we' signalled in this remark is a 'we' of which there are or may be many others (we 'western liberals' perhaps). Davidson certainly wanted the conclusions from his investigations into the nature of interpretation to be more general than this. As Lepore and Ludwig have emphasised, Davidson's account of radical interpretation 'aims to be a completely general account of interpretation, and hence of meaning, communication etc.' (Lepore and Ludwig 2005, 219). They go on: 'To take on board as fundamental constraints contingent assumptions about the psychology of particular speakers, or groups of speakers, would undermine the generality of the results we would hope to obtain from reflection on the nature of radical interpretation'. Davidson clearly thinks that consideration of radical interpretation can lead us to very general conclusions about the nature of thought.

It might look then that the claim must be that any psychological assumptions required for radical interpretation cannot be contingent, but must be necessary assumptions, constitutive of any interpretation whatsoever. That is, whatever psychological or anthropological assumptions play a role in interpretation, they must be perfectly general, and apply to any speaker or group of speakers. This is in fact sometimes suggested by what Davidson says.

Davidson's account of interpretation repeatedly emphasises that the successful interpreter must match his own norms of rationality to those of the person he is interpreting. Significantly, Davidson thinks that every speaker subscribes to the basic principles of rationality – at least the principles of decision theory and basic logic – whether explicitly or not. Davidson's claim is that if someone does go against those principles, he 'goes against his own principles'. He says: 'These are principles shared by all creatures that have propositional attitudes or act intentionally; and since I am (I hope) one of those creatures, I can put it this way: all thinking creatures subscribe to my basic standards or norms of rationality' (Davidson 2004, 195-6). This does, he admits, sound somewhat sweeping. However, he insists that it *comes to no more* than the claim that 'it is a condition of having thoughts, judgments, and intentions that the basic standards of rationality have application'. The standards of rational

---

<sup>18</sup> Compare: 'If we think in terms of the bleak interpreter, we are asking, as David Lewis put it, how the facts determine the facts – how semantic interpretations sit on top of the physical facts about things'.

consistency are taken to be constitutive of propositional attitudes and thought in the sense that unless one is mostly rational, one is not an agent at all.<sup>19</sup>

Davidson then is certainly committed to the idea that reflection on radical interpretation can yield what might be regarded as perfectly general constraints on mindedness as such. However, it is one thing to be committed to this and quite another to be committed to the idea that the interpreter of a particular speaker cannot make use of constraints that pertain only to minds of a particular local class (how they are brought up, what their biology is like, etc.). In other words, it is quite another thing to think that the only constraints that a (radical?) interpreter can appeal to in the interpretation of a speaker are perfectly general constraints that hold of mindedness as such. Indeed, one of the perfectly general constraints on mindedness as such could be that the interpretation of a particular thinker *must* involve appeal to constraints that pertain to the local class of thinkers to which that particular thinker belongs.<sup>20</sup>

So, while it is one thing to hold that there are general constraints on mindedness, e.g., that all minds must meet a standard of logical consistency, it is quite another to think that the interpretation of a particular mind cannot appeal to any constraints that are not perfectly general. The issue then becomes whether, and to what extent, we can tell the difference between constraints on mindedness as such, and constraints that pertain only to minds of a local class. I will try and say a little about why I think this issue is important.

It seems that in order to apply charity across the board, the interpreter must make a variety of psychological and anthropological assumptions about the speaker. It follows that an account of radical interpretation can either attempt to make these psychological assumptions explicit, or will simply end up smuggling them in with concepts such as 'natural' or 'normal'. In a later chapter I discuss Susan Hurley's claim that an interpreter would not be in a position to recognise a speaker's basic desires or preferences, and so the formal values of coherence and consistency that they exhibit, unless they shared with the speaker certain substantive evaluative and, according to Hurley, recognisably ethical interests.

At this stage however I just want to make the following point. Once we allow that psychological or anthropological assumptions are being made by the interpreter, we seem to

---

<sup>19</sup> This is of course not to imply that no one ever acts contrary to those principles. Charity, or the rationality constraint, is not undermined by the fact that people can be highly confused and can reason badly (this point is made by Millar 2004, 7).

<sup>20</sup> This point was put to me in conversation by Adrian Haddock.

invite the accusation that contingent psychological or anthropological assumptions (constraints that pertain only to minds of a particular local class) are being smuggled in amongst what are being passed off as necessary assumptions (constraints on mindedness as such). This, it seems, will cast doubt on the objective nature of any interpretation of the speaker, of the interpreter's understanding of what the speaker is about. Indeed, it is a long standing criticism of Davidson that he restricts the notion of rationality to speakers who are to a large extent 'like us'.

The worry can be put like this. The radical interpreter is a cognitive imperialist of sorts, who imports his own categories and concepts, beliefs and principles, and applies them to the utterances and behaviour of the speaker. This however seems to neglect the possibility that he may encounter a speaker who has categories and concepts, beliefs and principles that are somehow radically different from his own. The extent to which Davidson is embroiled in this sort of issue is somewhat unclear. This is in part due to the general issue regarding the vagueness of some of Davidson's formulations of the principle of charity that I raised at the end of the preceding section. This issue is also I think linked back to the questions of how much and what sort of disagreement with the speaker that the interpreter can tolerate. I will come back to these points in subsequent chapters.

Davidson of course has famously argued that we can make no sense of the idea that there might be fundamentally different conceptual schemes. The idea is that we could only tell when a speaker was expressing ideas or concepts which were very different to our own once we had already managed to interpret them more generally. Davidson wants this point, as Blackburn puts it, 'to sustain the conclusion that we can have no conception of what it might be for a speaker to have a language which expresses concepts and beliefs radically different from ours – too different to provide the backdrop of shared thoughts that make mutual interpretation possible' (Blackburn 1984, 277). My response to this would be that until we have a more substantial account of what this means (what the 'backdrop of shared thoughts that make mutual interpretation possible' might involve) it might not mean very much.

## **6. Summary:**

Most philosophers now acknowledge that there are considerable constraints on interpretation of the form that Davidson has set out. In the first part of Chapter 1 I have outlined some of Davidson's claims about radical interpretation. He has argued extensively that most of the speaker's basic beliefs about his environment must be assumed to be true by the interpreter. This is a methodological claim about radical interpretation. He has also

argued that most of the speaker's basic perceptual beliefs must in fact be true in order for them to be known to the interpreter. This is a more general point about how propositional attitudes come to have the content they have. Davidson has also emphasised that there is an extent to which the interpreter must take the speaker's beliefs, attitudes and actions to be generally rational and coherent. This constraint is supposed to show that most of the beliefs that the interpreter attributes must be true by the interpreter's lights. There is also the point that the propositional attitudes come as a set.

In the second part of this chapter I have suggested that we should be uneasy about the vagueness of Davidson's suggestion that the interpreter must seek to maximise 'the right sort of agreement' with the speaker. I have also somewhat more tentatively suggested that interpretation necessarily involves entering into the psychology of the speaker in ways that may be more problematic than Davidson allows. This follows from Davidson's insistence that charity be applied across the board. I will pick up on both of these issues in subsequent chapters.

In Chapter 2 I will consider how Davidson's work on radical interpretation is connected to his remarks about the nature of values, and what its implications might be for ethics in particular. The central idea is that the principle of charity must play much the same role in the attribution of values to a speaker as it does in the attribution of beliefs. This is taken to show that certain general claims about the content of propositional attitudes, which Davidson thinks follow from the consideration of radical interpretation, can be applied to values as well as beliefs.

## ***Chapter Two: Interpretation and Values***

In Chapter 1 I set out some of the key features of Davidson's work on radical interpretation. In Chapter 2 I will consider in more detail how Davidson's work on radical interpretation is connected to his remarks about values. In a series of later papers Davidson developed the idea that the attributions of desires, preferences and values to any speaker must be governed by the principle of charity in much the same way that the principle must govern the attribution of beliefs to any speaker. In fact, even in his earlier work Davidson would often speak of the idea that charity must be applied to propositional attitudes 'across the board'. I take it that this is the central idea that underwrites Davidson's remarks about values, and ethics more specifically.<sup>21</sup>

This aspect of Davidson's work has in fact not been the subject of a great deal of secondary literature. This is somewhat surprising as there seems to be a degree of promise to the idea that we might try to understand something of what people say and think about values in particular by looking at how we come to attribute values to them. I will consider how this central idea might be understood in terms of the principle of correspondence and the principle of coherence that were set out in Chapter 1. I will argue that there is prima facie plausibility to the idea that the role of the principle of charity in the attribution of values to a speaker will be similar to the role of the principle of charity in the attribution of beliefs to the speaker. I will then go on to discuss what might follow from this idea. One thing that might be thought to follow is that certain claims about propositional attitudes quite generally, which Davidson thinks follow from the consideration of radical interpretation, will apply specifically to evaluative and ethical judgments. I will suggest there is an important limitation to Davidson's approach as it is set out.

### **1. The Argument from Interpretation:**

In this section I will look at Davidson's claim that the principle of charity must play much the same role in the attribution of values to a speaker as it does in the attribution of beliefs. This claim is connected to the idea that the consideration of radical interpretation can yield important insights into how the content of propositional attitudes is determined. I will consider the potential implications of this claim with respect to the key aspects of charity

---

<sup>21</sup> I will refer to Davidson's arguments for this idea under the broad title of the 'argument from interpretation' hereafter.

which were separated out in the previous chapter under the titles of correspondence and coherence.

The principle of correspondence emphasises the significance of the causal connections that link the interpreter, the speaker, and the shared world. In the case of basic perceptual beliefs, there seem to be two separable claims. Firstly, the interpreter must assume that the speaker is largely correct in his beliefs about his immediate environment (by the interpreter's lights). This is what I called in Chapter 1 the methodological claim. Secondly, any speaker must in fact be largely correct about his beliefs about his immediate environment, or about his most basic perceptual beliefs generally. This is what I called in Chapter 1 the epistemological claim. Both of these claims are supported by causal considerations.

The argument behind the first claim is this. In the radical interpretation scenario, the interpreter must initially identify objects (where this is inclusive of entities or events) which are common to himself and to the speaker, and assume that what the speaker is caused to believe by those objects (in most cases) is what the interpreter would be caused to believe by those same objects. For example, in a straightforward case, the interpreter 'can do no better than to interpret a sentence that a person is selectively caused to hold true by the presence of rain as meaning that it is raining' (Davidson 2004, 36). It is only because the interpreter 'consciously correlates the responses of another creature with objects and events in the observer's world' that there is any basis for saying that the speaker is responding to one object in his surroundings rather than another (Davidson 2001, 212-3). Davidson acknowledges of course that this rule can accommodate numerous exceptions, but maintains that to ignore it is simply to abandon interpretation altogether (cf. Davidson 2004, 36).

Davidson has maintained that the same rule must apply when the interpreter attributes evaluative attitudes or judgments to the speaker. Again, the interpreter must identify objects (entities or events) common to himself and the speaker, and assume that the speaker's response is (in most cases) similar to the interpreter's own response to these objects. It is worth emphasising (as was pointed out in Chapter 1) that there are two distinct assumptions in play here. Firstly, the interpreter is prompted to take the speaker to be responding to the same object as he is. Secondly, the interpreter is prompted to assume that the speaker responds to those features of the world in more or less the same way the interpreter himself responds. The idea is that the interpreter must make both of these

assumptions in order to make a start at interpreting the speaker. Remember, Davidson says that 'unless we assume that we can identify and share reactions to common stimuli 'thought and speech would have no particular content – that is no content at all'. (Davidson 2001c, 214).

The argument behind the second claim is this. Davidson claims that the content of any speaker's earliest and most basic sentences must be determined by what it is in the world that causes him to hold them true. Consideration of radical interpretation suggests this, as it shows the observation of the circumstances in which the speaker utters a sentence (or applies a predicate) is basic to the understanding of the utterance. The idea is that a speaker must learn his most basic words, words like 'apple', 'man', 'dog' etc., which at the start function as sentences, through a conditioning verbal behaviour to specific objects in the environment. According to Davidson, the speaker learns to understand evaluative attitudes and judgements in much the same way. Consider for example the following simple sentences: 'Good!', 'Bad!', 'Evil!', 'Brave!' 'Cruel!' etc. The interpreter learns that these utterances 'apply to actions and objects of the sorts of things to which we find them applied' (Davidson 2004, 48). Davidson describes this view as the view that values are 'rooted in things'. The view is that the content of a thinker's values or value judgements (as with our thoughts or sayings generally) is partly determined by the history of his causal interactions with the environment. In the simplest and most basic cases, words and sentences derive their meanings from the objects and circumstances in which they are learned, and this applies to basic evaluative words and sentences too. The objects and circumstances in which basic words and sentences are learned are thus content fixing.

For Davidson then, what ensures that a speaker's view of the world, in its plainest features, is largely correct 'is that the stimuli that cause our most basic verbal responses also determine what those responses mean, and the content of the beliefs that accompany them' (Davidson 2001c, 204). Further, Davidson thinks this can be shown by 'appealing to obvious facts about language learning and facts about how we interpret words and languages with which we are unfamiliar' (Davidson 2001c, 43-45).<sup>22</sup> Davidson seems to acknowledge that the interpretation of evaluations can be subjected to more additional considerations than simple beliefs when this general rule is applied. Nevertheless, he

---

<sup>22</sup> Of course, not all words and sentences are learned in this way. Often words and sentences are learned via their relation to other words and sentences. Further, any given belief or utterance may be caused by misleading sensations. But, in the simplest and most basic cases words and sentences derive their meaning from the objects and circumstances in which they are learned, and 'it is those that are that anchor language to the world' (Davidson 2001c, 44-45).

maintains that with evaluations as well as beliefs there must be ‘the general presumption (he adds ‘often overridden by other considerations’) that similar causes beget similar responses in the interpreter and the interpreted’ (Davidson 2004, 71). It is in fact difficult to overstate the importance of this point when we consider Davidson’s account of values. In one place Davidson says: ‘What makes our judgments of the "descriptive" properties of things true or false is the fact that the same properties tend to cause the same beliefs in different observers, and when observers differ, we assume there is an explanation. This is not just a platitude, it's a tautology, one whose truth is ensured by how we interpret people's beliefs. My thesis is that the same holds for moral values’ (Davidson 2004, 47).

The principle of coherence concerns more general holistic constraints on the attributions of propositional attitudes. Charity dictates that the interpreter’s attributions of belief to a speaker must be consistent with each other and with certain general principles of rationality. Some of the implications of this were discussed in Chapter 1. We saw for instance that Davidson has used these considerations to argue that most of the propositional attitudes that an interpreter attributes to a speaker must be true by the interpreter’s lights. Davidson has also suggested that the interpreter must find a similar degree of coherence and consistency (both intra-attitudinal and inter-attitudinal) in the evaluative attitudes and judgments of the speaker. This general remark could be thought to contain a number of elements which relate to how the content of evaluative attitudes and judgments is constrained by the principle of charity.

There is, one might think, a general holistic point about the content of values which relates to what could be called evaluative concepts<sup>23</sup>. This is just to say that in order to attribute a belief about justice, say, to a speaker, the interpreter would have to at the same time attribute lots of beliefs about justice to the speaker. (Remember the remarks about snakes in Chapter 1). So, where the interpreter is prompted to attribute a belief that contains an evaluative concept, he must also be prompted to attribute lots of generally true beliefs which involve that evaluative concept, and possibly other related concepts. Davidson does say in one place that any speaker (any creature capable of thought and action) ‘must have the concepts of the evaluative properties, and must employ these concepts in making judgement’ (Davidson 2005, 55). It remains to be seen whether the interpreter must endow the speaker with some minimal set of specifically ethical or moral concepts and interests.

---

<sup>23</sup> For now we can think in terms of a very broad class of evaluative concepts inclusive of ethical and aesthetic concepts, though distinctions in this class will become relevant further on.

Quite a lot seems to hinge on this issue. Davidson's list includes ethical concepts, but it is very important to keep in mind that this needs arguing.<sup>24</sup>

Davidson has stressed that the interpreter must find certain formal constraints among the desires and preferences of the speaker at a structural level. For example, he has often highlighted the fact that the interpreter will find it hard to credit even the most blatant evidence for the intransitivity of preference, noting that the interpreter would tend to explain such cases as being the result of changes in preference over time. In fact, in one place he claims that 'the constraints that a Bayesian theory of preference places on the pattern of beliefs and evaluations exert a prima facie claim on interpretation; consistency of preferences with one another and with beliefs is a constitutive pressure on interpretation simply because we cannot rationalise (i.e., explain or understand) deviations from it' (Davidson 2004, 71). The idea, as Lillehammer has put it, is that 'rational agents must exhibit a basic level of instrumental consistency in order for propositional attitudes to be intelligible as reasons explanatory of action. In particular, rational agents must be interpretable as pursuing their desires in accordance with their beliefs' (Lillehammer 2007, 206).

There is also sometimes the suggestion that the interpreter could not recognise a speaker's desires or preferences at all unless he shared certain specific and substantive moral or ethical interests with the speaker. The point might be put in the following way. The Davidsonian conception of rationality (the 'constitutive idea of rationality') does not simply involve the sort of formal constraints mentioned in the preceding paragraph. It is also substantive, and contains at a fundamental and structural level certain substantive shared ethical interests and concepts. What might be shown to follow from this thought, or how far these evaluative attitudes or interests can ever be made explicit, is never entirely clear in Davidson's work. In work that leans heavily on Davidson's own, Susan Hurley has developed the argument that any interpreter must attribute certain substantive and recognisably ethical evaluative interests to a speaker even in the course of crediting them with basic desires and preferences. This is because, Hurley claims, it is only by reference to some sense of the speaker's substantive and ethical values, to what the speaker thinks important and worthwhile, that an interpreter could give content to those desires and preferences. Must the interpreter find that any speaker must share specific ethical concerns, or as Williams has

---

<sup>24</sup> A big difference with the case of ethical concepts (as opposed to, say, snakes) is that speakers disagree about the nature of the concepts themselves. That is, there are disagreements about what justice, say, is as well as what it demands and whether or not it is instantiated. So the interpreter has to attribute to the speaker not only beliefs which involve the concept, but a belief about the concept itself.

put it 'a sense of what we think important, laudable, hateful, to be condemned or despised and so forth, in the actions and reactions of human beings' (Williams 1995, 139)? I will discuss Hurley's claim, and the question posed by Williams, in detail in Chapter 4.

However, for the time being we might remain neutral about the precise implications that these holistic constraints have on the interpreter's interpretations of the evaluative attitudes and judgements of the speaker, while agreeing that it seems difficult to deny that the interpreter must find the speaker to have certain basic evaluative attitudes: wants, desires, preferences etc. Furthermore, it seems plausible that the interpreter must assume that these evaluative attitudes interact with the speaker's other attitudes and with the speaker's beliefs and intentions (and actions) in an instrumental way – otherwise they would not be wants, desires, preferences etc.

We should accept I think that there is some prima facie plausibility to the idea that the role of the principle of charity in the attribution of beliefs to a speaker will be similar to the role of the principle of charity in the attribution of values to the speaker. Indeed, it is perhaps difficult to imagine that an interpreter could interpret the beliefs of a speaker in isolation from the evaluative attitudes and judgements of a speaker, due to the intimate connections that exist between the different types of propositional attitudes.<sup>25</sup>

Initially, the speaker must be regarded as forming evaluative attitudes and judgements with the same contents as those of the interpreter when confronted with the same or similar features of the world. The interpreter must assume commonality as a starting point, this is a methodological necessity for any understanding of the values of the speaker. Perhaps, also, the interpreter must expect that *at least some* these sentences and concepts are 'rooted' in the objects and circumstances which lead to their utterance or application, as is the case with many basic sentences more generally. Further, these evaluative attitudes and judgements will be further understood in terms of their place in an intricate overall pattern of thought and behaviour. This pattern provides certain constraints on the attitudes that an interpreter can intelligibly attribute to a speaker. The interpreter would for example struggle to understand a speaker who announced that they preferred A to B and B to A, because this is not how preferences work. Equally, the interpreter would struggle to understand a

---

<sup>25</sup> This idea does not seem to be out of keeping with Davidson's earliest work on radical interpretation. Davidson's later remarks on the connections between interpretation and our understanding of value have been presented as an 'argument by analogy' (cf. Ludwig Reference) or as an extension of the principle of charity to values. My view is that these remarks are perhaps best understood as a more detailed articulation of ideas which were always to some extent implicit.

speaker who sincerely claimed to prefer A to B but who, whenever faced with a free choice between A and B consistently chose B, because preferences are understood to relate to actions in a certain way.

Reflection on the nature of radical interpretation could then shed light on certain features of our evaluative attitudes and concepts. The view is that the content of evaluative attitudes or judgements (as with thoughts or sayings generally) is partly constrained by the history of the thinker's causal interactions with the environment and partly constrained by holistic and rational considerations. The precise implications of these general points are somewhat less clear. The idea that the principle of charity plays a role in the attribution of values to a speaker is in some ways more problematic than the above remarks suggest. As Davidson was of course well aware, there are basic differences between evaluative attitudes and judgements and beliefs, and between evaluative concepts and other sorts of concepts, and we might expect that these differences will have to be reflected in our account of interpretation.

## **2. The General Presumption of Truth and Coherence:**

One thing that might be thought to follow from the above remarks is that certain claims about propositional attitudes quite generally, which Davidson thinks follow from the consideration of radical interpretation, will apply to evaluative judgments. Indeed, Davidson suggests that the reflection on the nature of radical interpretation, and particularly on the role of the principle of charity, can support the conclusion that such judgements are true or false in much the same way that descriptive judgments are, and that there is more agreement on moral and other values than it may often seem (Cf. Davidson 2004, 44).

This is in part because reflection on radical interpretation sheds light on how such judgments get their content. Davidson has often emphasised how considerations of coherence and consistency constrain the content of propositional attitudes more generally. The emphasis on coherence is used to support the claim that the interpreter cannot attribute any belief to a speaker without attributing lots of other beliefs with related content. This can be generalised to the claim that most of the beliefs attributed to the speaker must be true in the interpreter's opinion. There is then, as Davidson has occasionally claimed, a 'general presumption' in favour of the truth and consistency of a speaker's beliefs, judgments and attitudes which emerges from the consideration of the principle of coherence (cf. Davidson

2001c, 153). According to Davidson, all of a thinker's beliefs and attitudes are justified in so far as they are supported by numerous other beliefs and attitudes.<sup>26</sup>

Davidson has strengthened this claim by arguing that the presumption in favour of any given belief is increased the larger the body of other beliefs with which it coheres. The idea is that the more central or significant a given belief is within the overall network of beliefs, the more difficult it becomes to coherently raise doubts about the belief. It would be difficult for example to continue working in certain areas of physics while casting aside  $E=MC^2$ . Of course some seemingly 'central' or 'significant' beliefs, which might appear to cohere with a great many other beliefs that the speaker appears to hold, could still be held false by the interpreter: that the world is flat, that a woman's place is in the home, that God is Great etc. Nevertheless, Davidson has in fact gone as far as to claim that the consideration of the interconnected nature of belief (and propositional attitudes more generally) can rescue us from a certain form of general scepticism by showing that it is unintelligible that all or most of the beliefs of any thinker could be false or inconsistent.<sup>27</sup>

Some philosophers have been suspicious of Davidson's move from the claim that the interpreter must find most of the speaker's beliefs to be true to the claim that most of these beliefs must in fact be true. It has been argued that the sort of holistic constraints on interpretation that Davidson has emphasised might permit the conclusion that no interpreter could correctly interpret a speaker in such a way that the speaker's beliefs came out as massively false in the interpreter's opinion. What follows from this is more controversial. It is often suggested that these holistic constraints on mental content which Davidson has emphasised do not permit the conclusion that no thinker's beliefs could turn out to be massively false.

Davidson was certainly aware that coherence on its own is not enough to justify any presumption about the truth of beliefs about what exists, for example. If the interpreter were to correctly attribute a belief about a snake to a speaker, they would have to attribute many other true beliefs about snakes to the speaker (that it would have no arms or legs etc.). However, as Davidson has put it, such truths are 'general truths' and 'general truths

---

<sup>26</sup> From the interpreter's point of view there must be a general presumption of truth for the body of beliefs as a whole, though the interpreter does not need to presume that each particular belief of a speaker is true. From the vantage point of each thinker 'there must be a graded presumption in favour of each of his own beliefs' (cf. Davidson 2001c, 153).

<sup>27</sup> He once put the point about consistency thus: 'There is no chance that a person's beliefs will not tend to be consistent, since beliefs are individuated in part by their logical properties: what is not largely consistent with many other beliefs cannot be identified as a belief' (Davidson 2001b, 155).

like these do not imply that snakes exist, but only that if there were a snake, then it would have no arms or legs' (Davidson 2004, 17). The coherence of a belief with a substantial body of beliefs then only enhances its chance of being true *provided that there is a reason to suppose that the body of beliefs is true* (or at least largely so). The point here however, as he once put it, is that coherence on its own, no matter how strongly defined, cannot guarantee that what is believed is largely the case (Cf. Davidson 2001c, 138).

What then does guarantee that what is believed is largely the case? The answer Davidson offers should not be at all surprising. The central idea was sketched in Chapter 1 and again briefly above in the remarks about correspondence and causation. To repeat a claim I highlighted earlier, what ensures that a speaker's view of the world is largely correct in its plainest features is that the stimuli that cause his most basic verbal responses also determine what those responses mean, and the content of the beliefs that accompany them. The idea then is that any coherent set of beliefs that a speaker might have must be suitably anchored to the world, and that this anchoring is content fixing. For Davidson, the nature of interpretation guarantees that a large number of our simplest perceptual beliefs are true by giving us an account of how such beliefs get their content.

We might then re-state Davidson's general presumption in the following way. Firstly, all of a thinker's beliefs are justified to some extent in the sense that they are supported by numerous other beliefs and attitudes. Secondly, most of any thinker's basic beliefs are anchored to the world, because the contents of these beliefs must in part be determined by the history of the thinker's causal interactions with the world. Again, in the simplest and most basic cases, words and sentences derive their meaning from the objects and circumstances in which they are learned.

It is fair to say that there remains a considerable degree of scepticism about the significance of Davidson's epistemological claims, or about the significance of the 'general presumption' more particularly. Simon Blackburn, for example, has appeared to suggest that Davidson struggles to avoid 'pure coherentism' and, by implication, idealism about mental content (cf. Blackburn, in McDonald, C. and McDonald, G. ed. 2006).

I feel that Blackburn and others are being uncharitable in this respect. Davidson has repeatedly tried to distance himself from the sort of 'pure coherentism' that seems to be at issue here.<sup>28</sup> Further, as Davidson himself has remarked, it is hard to believe that the sort of

---

<sup>28</sup> Cf. Davidson 2001c, xvi: 'I would...like to correct the impression that I think experience and perception play no role in our beliefs about the world...I made it sound as if I were repudiating all

direct interaction between language users and their surroundings is not part of any proper account of how words are related to things. In the simplest and most basic cases words and sentences appear to derive their meaning from the objects and circumstances in which they are learned, and thinking about radical interpretation gives us a good account of why and how this is so. As Davidson has stressed repeatedly, a sentence which a speaker has been conditioned to hold true by the presence of fires, for example, will be true when there is a fire present: a word which a speaker has been conditioned to be caused to hold applicable by the presence of rain will refer to rain.<sup>29</sup> Again, many words and sentences are not learned this way, but 'it is those that are that anchor language to the world' (Davidson 2001c, 44-45).

It does not seem to me like Davidson's account of the method of radical interpretation leaves him vulnerable to the accusation of 'pure coherentism' or idealism about mental content. Indeed, the accusation seems to neglect a fundamental component of Davidson's account. It is a crucial part of Davidson's understanding of radical interpretation that the contents of any mind will depend on the causal relations, whatever they may be, between the mind and the world.<sup>30</sup> Whether or not we can be completely satisfied with Davidson's attempts to articulate word-world relations, and with the conclusions that he draws from these attempts is another matter.

I can put this point another way. For Davidson, providing a proper account of the semantics of a speaker's language forces an interpreter to engage with the ontology of the language. Indeed, Davidson moves from claims about semantics to substantial claims about ontology, about the sort of things that exist. Primarily, his arguments suggest that the correct interpretation of a speaker's language by an interpreter must result in a large degree of agreement about what we might call basic ontology. However I have my doubts about the moves that Davidson appears to make from this claim to certain other claims, particularly his claims about ethical judgements.

These doubts can I think be expressed in slightly more general terms. The principle of charity appears to provide a certain 'blanketing' quality that is thought to render all substantive

---

serious commerce between the world and mind...My thesis is that the connection is causal, and in the case of perception, direct'.

<sup>29</sup> Of course, it is more problematic to say that 'a word which one has been conditioned to be caused to hold applicable by the presence of a Kraut' will refer to Krauts.

<sup>30</sup> The world here is taken to include other thinker's. Further, for Davidson causality is not restricted to the physical. Mental objects can interact causally with each other and are constrained by rationality. They can also interact causally with physical objects. So when Davidson is talking about properties causing attitudes, this is not necessarily understood in purely physical terms.

beliefs or propositional attitudes presumptively true, or perhaps that is thought to render all substantive beliefs *about what exists* presumptively true. However, any general presumption in favour of the truth of beliefs or propositional attitudes will simply shift the focus to the truth of specific beliefs or propositional attitudes.

That is to say, we can accept the general connection between interpretation and basic ontology that Davidson has provided. However, given the worries I raised in the previous chapter about the vagueness of 'the right sort of agreement' we can cast doubt on some of his more substantial claims which seem to follow from this. How much and what sort of agreement there must be is crucial. Specifically, we can't get from any general claim about agreement on basic ontology to a specific claim about a core of agreement on what exists when we consider ethical properties. It seems to be the latter point that is important for what Davidson says about ethics.

Davidson claims that as long as we 'adhere to the basic intuition that in the simplest and most methodologically basic cases words refer to what causes them, it is clear that it cannot happen that most of our plainest beliefs about what exists in the world can be false.' But he immediately acknowledges that many beliefs are given content by their relation to further beliefs, or are caused by misleading sensations, and remarks that 'any particular belief or set of beliefs about the world around us may be false' (Davidson 2001c, 214). So, I want to suggest, granting Davidson's general presumption in favour of the truth of our beliefs or attitudes does not seem to prevent the possibility that some specific set of beliefs or attitudes, for example beliefs or attitudes about ethics, about aesthetics or about religious matters, may generally be false or misguided.

Perhaps we can say that this general presumption applies unless we have specific reasons to doubt some particular area of discourse. The general presumption can give us that at least. (My beliefs about tables or trees are assumed in general to be true because in part because I have no reason to doubt them.) However, many people have argued that there are good reasons for doubting the truth or objectivity of judgements and attitudes in certain particular areas of discourse. More specifically, there are people who think that there are good reasons for regarding ethical or moral judgments as largely false or indeterminate, or as falling short of some standard of objectivity. I will say more about this in the following chapter.

The key idea at this stage is that, according to Davidson, there is a general presumption that the beliefs of any speaker are largely true, because there is a general presumption that the

basic beliefs of any speaker are largely true. But that point alone does not fix it which beliefs are the true ones, and so seems to leave it open that there might be a philosophical significant class of beliefs none of which are true, for example the ethical ones.

Davidson clearly wants to insist not only that ethical or moral judgements are properly understood as being objective (that is, true or false), but also that a majority of them must be true (that 'there is more agreement about them than there may seem'). My claim at this stage is that whatever the significance of Davidson's argument from interpretation, and the general presumption that follows from it, it does not seem to secure the objectivity of ethical judgments or the claim that many or most ethical judgements must be true. Even if we accept the general presumption (about thought generally and perceptual beliefs in particular) we can still raise certain worries about the status or nature of evaluative attitudes and judgments, or more specifically ethical attitudes and judgements.

Davidson has suggested that once we are clear about how value judgements acquire content we should lose interest in the ontological issues that are traditionally associated with questions regarding objectivity in ethics. I believe that this is an exaggeration. In one place he says: 'What cannot be the case is that our general picture of the world and our place in it is mistaken, for it is this picture which informs the rest of our beliefs and makes them intelligible, whether they be true or false' (Davidson 2001c, 214). The key issue, it seems to me, is whether or not our 'general picture of the world and our place in it' necessarily includes certain ethical interests or concerns, properties or concepts.

Above I suggested that it remains to be seen whether the interpreter must endow the speaker with some minimal set of specifically ethical or moral concepts and interests. (Davidson's list includes ethical concepts, but it is very important to keep in mind that this needs arguing.) A great deal seems to hinge on this issue. For it seems that to grant Davidson's general presumption is to accept that there are a broad class of 'basic beliefs' which must have some true instances (i.e., it cannot be that on every occasion on which someone holds a belief of this sort, the belief is false), because their content is determined by what causes them. The question then is whether or not an interpreter must include amongst these basic beliefs not simply beliefs about the colours and shapes of things, but beliefs about the 'ethical properties' of things (i.e. as given by thick ethical concepts). I take it that while nothing in Davidson's account of radical interpretation as I have presented it rules this out, nothing has established it either.

I want to suggest however that it is doubtful that Davidson could derive substantial conclusions about ethics from an a priori argument from interpretation *without this point*, that is without including beliefs about ethical properties amongst 'basic beliefs'. To some extent the whole ethical import of Davidson's account of radical interpretation hangs on this claim, which I am associating with the idea that a degree of concern for certain ethical properties is a condition of mindedness as such. My view, which I aim to develop in the subsequent chapters, is that if Davidson attempts to include concern for certain ethical properties amongst the constraints on mindedness as such that are general requirements of all interpretation, he is likely to leave himself open to the suspicions about smuggling that I raised briefly in Chapter 1.

## **Chapter 3: Interpretation and Value Properties**

In Chapter 2 I have argued that there is prima facie plausibility to the idea that the principle of charity must play a role in the interpretation of a speaker's values. Indeed it is perhaps difficult to imagine that an interpreter could interpret the beliefs of a speaker in isolation from the values of a speaker. Davidson however at times moves from this general point to make a number of more substantial claims about values, and about ethical judgements in particular.

These claims appear to draw some support from Davidson's argument that there must be a 'general presumption' in favour of the truth and coherence of the propositional attitudes of any thinker. I have summarised the key components of the argument for this general presumption in the following way: Firstly, most of a thinker's beliefs and attitudes must be justified to the extent that they are supported by numerous other beliefs and attitudes which make up a largely coherent and rational set. Secondly, most of a thinker's basic propositional attitudes are anchored to the world, because the contents of those propositional attitudes must in part be fixed by the history of the thinker's causal interactions with the world.

The general idea is that any coherent set of propositional attitudes that a thinker can have must be anchored to the world, and that this anchoring is content fixing. This in turn fixes it that any thinker must have mostly true beliefs. Towards the end of Chapter 2 I suggested that it seems that to grant Davidson's general presumption is to accept that there are a class of 'basic beliefs' (perhaps mostly of the form 'a is F') which must have some true instances because their content is determined by what causes them. That is to say, it cannot be the case that on every occasion on which a thinker holds a basic belief of this sort, the belief is false. Davidson of course acknowledges that the content of a great many other beliefs can be given by their relation to further beliefs, or are caused by misleading sensations, and remarks that 'any particular belief or set of beliefs about the world around us may be false' (Davidson 2001c, 2014).

I also suggested that if this is true, then granting Davidson's general presumption does not seem to prevent the possibility that some specific set of beliefs or attitudes, for example about ethics, religion or pseudo-science (e.g. phrenology, astrology), may in general be false or misguided. The general presumption by itself does not fix which propositional attitudes are true or correct, and so seems to leave it open that there might be a philosophically

significant class of propositional attitudes none of which are true or correct, e.g., the ethical judgments. That is to say that whatever the significance of Davidson's general presumption, it does not as it stands do enough to secure a claim to which Davidson appears to be committed – that most ethical judgments must be true, and that there is more agreement on ethical matters than there might seem to be.

Whether Davidson's general presumption rules out the possibility I have just mentioned depends, it seems to me, on whether or not the general presumption can be spelled out in such a way as to include ethical judgments, or concern with ethical concepts, amongst the 'basic beliefs'. That is, the issue is whether or not the interpreter is constrained so as to include amongst the basic beliefs of any thinker not simply beliefs about the colours or shapes of objects for example, but beliefs about certain ethical properties of objects. This is perhaps what Davidson has in mind when he says that 'values are rooted in things'. However, nothing in Davidson's work as I have presented it up to this point seems to have established this yet.

The idea is that, according to Davidson, it seems to be a condition for being a thinker (a 'condition of mindedness as such') that thinkers have ethical beliefs amongst their basic beliefs. My aim in what follows is to put some pressure on this idea. I will argue that nothing in Davidson's blanketing ontological claims can secure this point. I will then go on to argue that a full understanding of ethical properties suggests that they are unlikely candidates for featuring amongst the 'basic beliefs' of all thinkers, or for playing a role in 'constraints on mindedness as such'. I will also suggest that the fact that the nature of ethical properties leaves open a number of questions regarding ethical objectivity which Davidson fails to address. I will start however by putting Davidson's remarks about values in a wider philosophical context. To this end, I will place Davidson's remarks about value judgments and statements in the context of some well-known issues that are raised by non-cognitivist and error theoretic approaches to values.

### **1. Non-Cognitivism and Error Theory:**

Straightforward belief statements are understood to have cognitive content, to be true or false. It is sometimes thought to be less clear that this is the case for statements that express value judgements. There is a large body of work in ethical philosophy which holds that many or all value judgements do not in fact have truth values. I have in mind the non-cognitivists of various sorts who have argued that statements such as 'setting the cat on fire is cruel' should not be understood as having truth conditions at all. They claim that such statements

are not in the business of expressing judgements which are true or false in any substantial sense— they are not typically expressing states of mind which are like beliefs or which are cognitive in the way that beliefs are. Statements of this kind should not, according to the non-cognitivist, be understood as predicating properties at all, but should rather be understood as expressing attitudes. They are in fact expressing non-cognitive attitudes more similar to desire, approval or disapproval<sup>31</sup>. It should not be particularly surprising to anyone who is familiar with Davidson’s wider work in the philosophy of language that he, along with many others, thought that this approach is mistaken and that ethical judgments and ethical statements simply must be regarded as true or false, and for a number of reasons.

The general line that he takes with non-cognitivism is to insist that these value statements must have truth conditions in virtue of their form. In more than one place he urges us to consider what exactly it is that we do when we judge that an action or event is ‘good’, ‘evil’, ‘morally desirable’, ‘obligatory’, ‘cruel’ etc. He stresses that when we do this for even a moment ‘the semantic nature of such judgements becomes clear’ (cf. Davidson 2004, 48). We must be attributing some property or other to an entity or group of entities. This is to say that when we state that ‘setting the cat on fire is cruel’ this *can only mean* that we attribute the property ‘is cruel’ to the event ‘setting the cat on fire’ (that is, we attribute some property to the subject of the sentence). Such statements may be true or false but they must have truth conditions – either the subject of the sentence (the entity or event) has the property or it doesn’t. To think otherwise would, according to Davidson, be to ‘bifurcate language in an unacceptable way’ by leaving the semantics of sentences which express value judgements detached from the semantics of other types of sentences (cf. Davidson 2004, 25).

Davidson held that there are numerous other considerations that support this conclusion. To this end he has at various times highlighted the logical form of compound sentences, generalisations and embedded sentences. Take for example his discussion of the nature of compound sentences. When a speaker says ‘I ought to give more money to charity and I will do so in future’ nobody would doubt that the second conjunct of this statement has a truth value, is true or false. Davidson challenges us then to explain what ‘and’ could mean in this

---

<sup>31</sup> Most philosophers who work under the ‘non-cognitivist’ banner occupy subtle positions within the general outline I provide. As van Roojen remarks, ‘many non-cognitivists hold that while *primary* function of evaluative judgements is not to express beliefs, they may express them in a *secondary* way. Others deny that their contents are true or false in any *robust* sense but not that they can be true or false in a *deflationary* sense according to which there is no substantial property separating true and false sentences’ (van Roojen 2009).

statement. He says that no one has explained the role of a conjunction except by saying 'a conjunction is true if and only if each conjunct is. It follows that the first conjunct is true if the sentence as a whole is' (Davidson 2004, 48).

Davidson has also cited the role of value judgements in practical reasoning as further support for his position. When we reason about what to do or about the value of our actions or the actions of others, we will often combine value judgements with other types of judgement and beliefs. We conclude from our judgement that the eradication of poverty would be good and our belief that giving money to charity will help eradicate poverty, that we should (if we can) give money to charity. If practical reasoning of this sort can be shown in some cases to be valid, Davidson argues that the 'premises and conclusion must have truth values, as validity is defined as a truth preserving mode of reasoning' (cf. Davidson 2004, 55).

In light of these considerations and many others, Davidson has urged that we must face the fact that it is, as he puts it, 'difficult, if not impossible' to avoid the conclusion that value judgments have truth values. This holds without an insistence that value judgements are exactly like other types of judgements or beliefs (though they may be regarded as a class of belief), or a denial of the many differences between evaluative concepts and other types of concepts. The types of argument for this claim just sketched follow naturally from Davidson's longstanding emphasis on the importance of the distinction between questions of the logical form or grammar of sentences and the analysis of individual concepts. The deep differences between descriptive and value statements do not show at the level of the former, hence we should not balk at the idea that value judgements have truth conditions and truth values.

I think that we should be willing to grant that at the level of logical form value judgements or statements poses no special problem for Davidson's account of interpretation<sup>32</sup>. By this I simply mean that we can understand value judgments or statements as being true or false (or perhaps neither), and recognise that the speaker holds them true or false. It seems however that we might hold this view even if we think that there are good reasons for regarding some sets of judgements or statements, for example ethical judgments or statements, as being mostly false or indeterminate because of the nature of the concepts or

---

<sup>32</sup> For my purposes here I need only to have provided a sketch of the sort of reasoning behind Davidson's rejection of certain forms of non-cognitivism. I think Davidson's arguments are good ones, but I do not need to claim that he has presented a conclusive case against all forms of non-cognitivism.

properties involved. And in fact many worries about the truth values of evaluative judgements are found at the level of conceptual analysis.

As Bernard Williams argued in an late essay, that ethical statements are statements is suggested by the fact that 'they are involved in speech acts of the assertive kind, they permit embedding and various other kinds of syntactic manipulation which are associated with the sorts of things that are statements, assertions, bearers of truth value'. Further, these 'surface facts' about semantics, as we might call them, will have to be honoured in any account of ethical statements. However, Williams went on to insist that many of the substantive questions in ethics 'about realism, objectivism, cognitivism, and so on' are questions not just about the existence of the surface facts, nor indeed about the adequacy of the surface facts to support the application of the word 'true' (Williams 1995a, 242). In a similar vein, and many years earlier, Michael Dummett famously suggested that the really important issues in ethics do not concern whether in practice the words 'true' or 'false' are applied to value statements, but whether, if they are so applied, 'the point of doing so would be the same as the point of applying them to statements of other kinds, and, if not, in what ways it would be different' (Dummett 1959).

One way to articulate the idea that ethical judgments could be mostly false or indeterminate would be via a form of error theory. Error theorists are suspicious of certain types or areas of discourse, and most famously of ethical discourse. An error theorist about ethical discourse is a cognitivist about ethical statements. The error theorist thinks that statements such as 'setting the cat on fire is cruel' are (typically) assertions. They would agree with Davidson that the speaker is attributing the property 'is cruel' to the subject 'setting the cat on fire'. (That is, they would honour what Williams calls the 'surface phenomena' of ethical statements). But the error theorist holds that such statements are systematically false or indeterminate *due to the nature or source of the properties* that are being attributed in these statements.

The error theorist about ethical discourse does not think that all sentences which contain ethical concepts are false. Primarily the type of statement they are concerned with are those that imply or presuppose the instantiation of an ethical property. For example, while the statement 'Pat thinks Patricia is unchaste' is straightforwardly true or false depending on what Pat thinks, the truth value of the statement 'Patricia is unchaste' is regarded as being somewhat more problematic. The error theorist might claim that the properties in question

simply do not exist. This seems to be Mackie's position (there is 'really' no such property as 'chaste' or 'unchaste' whether Pat thinks there is or not).

We are left then with the following possibility: by itself, the claim that value judgments must possess some truth value (must be true, false or perhaps neither) in virtue of their logical or grammatical form appears to be compatible with the thought that the truth value of many ethical judgements, say, may be indeterminate or false. Halvard Lillehammer appears to have noticed this quite general point too. Lillehammer remarks that Davidson often explains his conception of the objectivity of evaluative judgements in terms of their possession of some truth-value, noting that so understood evaluative judgements could be regarded as having truth values 'even if all evaluative claims were indeterminate or false' (Lillehammer 2007, 217). If Davidson wants to say something stronger than this, and he certainly does, then it is clear that he needs something more than these general arguments against non-cognitivist approaches to ethical discourse. This is where the argument from interpretation, and the general presumption in favour of the truth and coherence of any thinker's beliefs and attitudes that follows from it, seems to come in to Davidson's account.

Davidson has argued that not only are ethical judgements properly understood as being true or false, but also that many are true. He supports this view with a claim about how value judgements acquire content. He seems to appeal to the general presumption about the truth and coherence of beliefs and attitudes in this respect. But, I have argued, this alone is not enough to secure the claim that most ethical judgments must be true. In order for the general presumption to secure this claim, Davidson must show that ethical judgements, or beliefs about the 'ethical properties' of things, are included amongst the basic beliefs of any thinker. If Davidson can show this, then the error theory fails to get any grip on ethical properties and the judgments that predicate them. These judgements could not be mostly false, because their content would generally be determined by what causes them. But Davidson's blanket ontology does not tell us anything about specifically ethical properties without this point.

To summarise, the argument from interpretation is purported to show that the correct interpretation of a speaker's language by an interpreter must result in a large degree of agreement about ontology, about the sorts of things that exist. But, as I claimed in Chapter 2, we can't get from any general claim about agreement on basic ontology to a specific claim about a core of agreement on what exists when we consider ethical properties without saying more than this. What he must say, I am claiming, is that it is a condition for being a

thinker (that is, a condition on mindedness as such) that thinkers have ethical beliefs amongst their basic beliefs.

## **2. Error Theory, Primary Properties and Secondary Properties**

There are a cluster of questions which arise within accounts of value regarding the nature and source of value properties, such as: Do value properties really exist? Are they 'in the world'? Do they exist independently of contingent human sensibilities? Are certain value properties natural or unnatural? At first glance, these questions might seem relevant to the idea that judgments about ethical properties could (must) figure amongst the basic perceptual beliefs of any thinker. Certainly, it would be odd if beliefs about contingent cultural or social properties were to be included amongst the basic beliefs which lead to 'constraints on mindedness as such'.

Davidson however is quite spectacularly unconcerned with these questions. He thinks that we can address the question of whether or not value judgements can be intelligently regarded as true or false, and even show that many must be true, without settling questions of whether or not value properties are real or exist in the world, or whether or not they are independent of contingent human sensibilities. Davidson's approach is to concentrate on what he calls the 'epistemological problem' of value judgements and let the 'ontological problem' take care of itself. To address the epistemological problem is to look at how the content of value judgments is determined. To address the ontological problem is to look at the nature and source of value properties. Davidson states that 'if we were to solve the epistemological problem we would lose interest in the supposed ontological problem' (Davidson 2004, 44).

I think that Davidson's views on this point must at least in part be down to his commitment to certain claims about the relationship between semantics and ontology which I mentioned in the previous chapter. Davidson's account of how the content of value judgements is determined is essentially connected to his account of interpretation, which in turn supports his argument for general presumption of truth and coherence among the beliefs and attitudes of any thinker.<sup>33</sup> One of the things that this account is supposed to show is that an interpreter and a speaker must agree to a large extent on matter of basic ontology. In Chapter 2 I claimed that we can't get from any general claim about agreement on basic

---

<sup>33</sup> 'How do we tell what the content of a particular moral judgment is? This is a question of interpretation, of the understanding by one person of the utterances of another, since there is no other context in which the content of a judgment can be agreed to or disputed' (Davidson 2004, 48).

ontology to a specific claim about a core of agreement on what exists when we consider ethical properties without saying more than this.

If this is right, then it is my belief that the epistemological point that Davidson has made does not provide us with a good enough reason to lose interest in what might broadly be described as the ontological problems about value properties. In what follows I will focus on Davidson's explicit remarks on these ontological problems – that is, on the issues that arise when we consider the nature and source of value properties. I will attempt to show how these problems might cast doubt on the idea that concern with ethical properties will provide constraints on mindedness as such.

Davidson finds it strange to ask whether or not value properties exist, or whether they are 'in the world'. He remarks that while the entities and events to which we attribute value properties certainly exist or are 'in the world' (in most cases anyway), the 'properties we predicate of such things are neither here nor there, for properties have no location'. He then states that he can happily agree with philosophers such as Richard Hare and Simon Blackburn when they claim that 'it adds nothing to an account of values to insist that they are real, part of the furniture of the world, something waiting to be found or discovered' (Davidson 2004, 45).

In some ways this is a controversial claim. Many philosophers have thought it an extremely important issue whether or not value properties are in some sense 'made' or are 'found' or 'discovered' in the world<sup>34</sup>. Davidson does not deny *this*. He remarks that discussions of value are almost always 'infected' by this issue. So why does Davidson make this controversial claim, and how does he support it? It is perhaps helpful to return to the error theorist.

One way for an error theorist to set up their argument is in terms of the distinction between primary and secondary properties. Primary properties are thought to be properties that are independent of any thinker, such as extension, solidity, number, and motion. These properties would be expected to feature in an account of the world provided by a completed physics. Secondary properties are thought to be properties that produce sensations in observers, such as colour, taste, smell or sound. These properties would not be expected to feature in an account of the world provided by a completed physics. The (very crude) idea seems to be that only primary properties 'exist in the world', and so only judgments which

---

<sup>34</sup> An example of an important contemporary work which addresses this question directly is 'Truth, Invention and the Meaning of Life' by David Wiggins (Wiggins 1998, 139-184).

predicate primary properties of objects or entities can be properly regarded as being true or false.

The error theorist then might set up his argument is like this (E1):

1. It is a presupposition of ethical discourse that ethical properties exist 'in the world' in the same sense that primary properties exist 'in the world' – that is, that they would feature in an account of the world that would be provided by a completed physics.
2. Ethical properties do not exist 'in the world' in the same sense that primary properties exist 'in the world' – that is, they would not feature in an account of the world that would be provided by a completed physics.

Therefore

3. Statements that attribute ethical properties to entities or events are systematically false.

In so far then as we can take Davidson to engage with this type of argument, we can take him to cast doubt on the first premise of the argument. That is to say that we can take Davidson to deny that a specific fact about ethical properties – that they are not primary properties – would be sufficient for the claim that they do not exist, or that we cannot make objective judgements regarding them.

It is to this end it seems that Davidson discusses what are often referred to as 'companions in innocence' in this context, though Davidson does not use this term. That is to say that he discusses value properties in connection with secondary properties. In one place, Davidson notes that plenty of secondary properties (colour properties, such as the property of 'being green', are paradigm examples) are supervenient on the more fundamental properties of things<sup>35</sup>, without this counting against the objectivity of attributions of these properties to entities or events that are certainly in the world, or the truth or falsity of those attributions. Value properties and secondary properties such as colour properties might not be definable in purely physical terms, but it is still true that some things are green (and that most are not)

---

<sup>35</sup> When I say that certain properties are supervenient on more fundamental properties I typically mean something along the following lines: a property or set of properties *A* is supervenient on another property or set of properties *B* just in case no two things can differ with respect to *A*-properties without also differing with respect to their *B*-properties. There are, of course, subtly different ways of understanding the supervenience relation.

and that some entities or events have certain value properties, are cruel or fair, and others do not, and it is still true that some actions are courageous etc. and that others are not.

I have claimed that Davidson needs to show that ethical properties are suitable candidates for featuring in basic propositional attitudes. The suggestion here is that (so far) there is nothing in the distinction between primary and secondary properties that might reasonably cast doubt on this. Certainly it seems to be far too simplistic to say that value properties 'do not exist' or 'are not real' (that is, to generate the sort of worry that the error theorist at E1 has in mind) on the basis they are not primary properties<sup>36</sup>. As John McDowell among others has repeatedly stressed, and Davidson would agree, the contrast between primary and secondary properties is not a contrast between the real and the unreal, or between the veridical and the illusory (cf. McDowell 1998, 136). Indeed, the very suggestion that value properties might need to be like primary properties has looked to many philosophers to be very odd. Susan Hurley describes this thought as 'insane'.

It is I think worth dwelling a little on why people might think this, in order to better understand the motivations behind those who have been tempted by this type of error theory. Primary properties, and the causal theories that can be constructed from them, can achieve a certain standard of truth and objectivity. There is an understanding of the term 'realist' in which someone is said to be a realist with respect to a property or concept in so far as they believe that it would feature in some *ideal causal theory*<sup>37</sup>. Such theories have certain standards of proof. Particle physicists, for example, use the 'five-sigma' threshold of certainty, though researchers in other fields can claim a significant finding with just two or three-sigma threshold<sup>38</sup>. People then feel uncomfortable trying to impose ethical views on others once they begin to be aware that ethical propositions are not realist in this sense.<sup>39</sup>

---

<sup>36</sup> I take it that Bernard Williams is making essentially the same point, with his usual clarity, in the following passage: 'The fabric of the world from which the secondary qualities are absent is the world of primary qualities, and (to take for granted the answers to several large and contentious questions) the claim that secondary qualities, as presented, are not part of that world comes to much the same as the claim that they do not figure in an 'absolute conception' of the world on which scientific investigators, abstracting as much as possible from their various perceptual peculiarities, might converge. There is nothing unnerving or subversive about the idea that ethical qualities are not part of the fabric of the world in this sense' (Williams 1995, 177).

<sup>37</sup> As Hurley notes, in this sense Mackie was an anti-realist in that he held that ethical concepts would not feature in such theories.

<sup>38</sup> The 'five-sigma' threshold corresponds to a p-value, or probability of error, of  $3 \times 10^{-7}$ , or about 1 in 3.5 million.

<sup>39</sup> There are I think interesting connections between this point and things that Williams says about ethical authority.

In fact it seems to me that this understanding of the term 'realism', which might more usefully be called 'scientific realism', is quite common. Furthermore it should not be particularly surprising, given the prevalence and prestige of the scientific standard of truth and objectivity in modern societies, that it has managed to permeate the everyday concepts of truth and objectivity in these societies to some degree. Value judgements or propositions perhaps lack truth or objectivity in this strict 'scientific realist' sense. However, if one wants to preserve the idea that there can be real truths about colour, about art, courage, twinges of pain or national sentiments, then one must allow that whether or not value properties are properly considered as primary properties or secondary properties does not necessarily affect the truth or objectivity of our attributions of these properties to entities or events that are certainly in the world.

It might however be thought that there is some tension between Davidson's ambivalence towards questions about the nature and status of value properties, and his remarks about causation. As I have previously noted, Davidson held that what makes any thinker's judgements of the descriptive properties of entities or events true or false 'is that the same properties tend to cause the same beliefs in different observers', and he is quite clear that he thinks the same holds for value judgments.

In one place he says: 'Red objects tend to cause us to believe the objects are red, square objects tend to cause us to believe the objects are square, and precious objects tend to cause us to prize them. It is because the objects and events have the properties they do that they cause us to have the attitudes we do' (Davidson 2004, 47). So, we might think, Davidson would say that it is the cruelty of objects (or more typically actions) that tends to cause us to believe that those objects are cruel.

Secondary properties then, and it seems ethical properties, can for Davidson figure in causal explanations. It is not yet clear how far this claim threatens to embroil Davidson directly in the sort of debates about the status or nature of ethical properties that he appears to want to sidestep. That is, he would need to explain how ethical properties can figure in causal explanations, but without admitting them into the *ideal causal theories* which consist of primary properties. It is certainly not straightforward how he would accomplish this.

Davidson would perhaps turn to his thesis of anomalous monism by way of explanation here (cf. Davidson 2001b). It is worth noting however that if Davidson's claims about ethical judgements depend on this thesis then many would regard them as being undermined by that fact. It has been argued that for certain properties, for example ethical properties, to

figure in causal explanations, causal relations between entities or events need to hold in virtue of the entity or events possessing these properties. It is not clear however how these relations can hold in virtue of the entity or event possessing the ethical properties, if the fact that entity or event possesses certain physical properties--which are ipso facto not ethical properties--is sufficient to ensure that the relations hold. The last point follows from claim the thesis known as the causal completeness of the physical. This thesis is widely held, and it can seem that Davidson must either deny that thesis, or reject this requirement on what it is for properties to figure in causal explanations.

I am willing to grant however that in some sense thinkers may be understood to stand in causal relations to various ethical properties ('it was the cruelty of the action which caused my belief that...') while denying that Davidson has shown that the beliefs caused by such relations should be included amongst the 'basic beliefs' of any speaker, or that concern with such properties is a 'condition of mindedness as such'.

### **3. Error Theory and Response-Dependence**

There is another way that the error theoretic position might be construed. Rather than worrying about the fact that value properties do not feature in our ideal causal theories of the world, that they are not primary properties, there is a particular feature of secondary properties that many have found problematic when the analogy with value properties is made. This is the apparent 'response-dependence' of secondary properties. Indeed the distinction between primary and secondary properties is sometimes drawn explicitly in terms of the response-dependence of secondary properties.

McDowell, for example, has suggested that we often distinguish between the two types of property in the following sense: A property is a primary property only if what it is for an entity or event to have the property can be understood *without reference to subjective states*. In contrast, a property is a secondary property only if what it is for an entity or event to have the property cannot adequately be understood without reference to subjective states (cf. McDowell 1985). Secondary properties and, if we persist with the analogy, value properties, are understood to be response-dependent then in that the properties in question can only be explained with reference to the reactions of human subjects.

It is not in fact a simple task to characterise the exact nature of the response dependence of secondary properties, or ethical properties. However, the general worry in this area seems to be this: It seems plausible that secondary properties are relative to human responses in

the following way. If human beings were to evolve (or simply change) so that everything in the world that had appeared brown came to appear red, this is, according to one way of thinking, simply *what it would be* for the world to cease to contain brown things and come to contain only red things. Similar claims could be made about the human capacity to distinguish certain tastes or the ability to hear certain sounds. In contrast, it is highly intuitive to many people that if human beings were to evolve so that they stopped classifying a set of actions as 'cruel', this is not what it would be for the world to cease to contain 'cruel' actions<sup>40</sup>. The challenge is to show why the response-dependence has the consequences it does in the colour case but not in the value case.

I suspect that one of the primary motivations of those writers who have insisted that value properties 'exist in the world' has been to attempt to rule out the idea that value properties are response-dependent in any way that allows this worry to get a grip. That is, to insist that ethical properties are in some important sense 'response-independent'. Another way to set up the error theorist's argument then might be like this (E2):

1. It is a presupposition of ethical discourse that ethical properties are response-independent.
2. Ethical properties are not response independent.

Therefore:

3. Ethical statements which attribute ethical properties to entities or events are systematically false.

Now there is a response to this sort of argument which I think Davidson should take but which he doesn't take, or at least doesn't appear to take seriously. This is to accept, with the likes of McDowell and David Wiggins, that there is a variety of response-dependence which ethical properties share with other secondary properties, but insist that this does not make ethical judgements any less capable of being objectively true or false. Both McDowell and Wiggins suggest that ethical properties are in some sense contingent on the responses of

---

<sup>40</sup> Cf. Blackburn, in Honderich ed. 1985, 14: 'If we were to change so that everything in the world which had appeared blue came to appear red to us, this is what it would be for the world to cease to contain red things. The analogue with moral qualities fails dramatically: if everyone comes to think of it as permissible to maltreat animals, this does nothing at all to make it permissible, it just means that everybody has deteriorated'. Blackburn in fact seems to take this to be so obvious that he does not argue for it.

agents, but unlike the error theorist of E2 they do not think that this brings the reality or objectivity of ethical properties into question.

Again the approach against the error theorist here at E2 would be to cast doubt on the first premise of the argument, and again one approach might be to find companions in innocence – to point out that secondary properties like colours, properly understood, are also conceptually tied to how they are perceived by human agents (are ‘essentially subjective’ as Wiggins sometimes puts it), and yet such properties are rightly thought of as capable of being attributed to entities or events objectively. That is, judgements that predicate such properties of objects are capable of being true or false.<sup>41</sup>

McDowell, for example, has again often used the colour analogy to make his point. He has argued that: ‘An object’s being such as to look red is independent of its actually looking red to anyone on any particular occasion; so, notwithstanding the conceptual connection between being red and being experienced as red, the experience of something as being red can count as being presented with a property that is there anyway – independently of the experience itself’ (McDowell 1998, 134).

But it is important to note that McDowell at least it seems is not simply denying the plausibility of the first premise of E2, he is denying the plausibility of the claim about red and brown that allowed the general worry raised above to get going. That is, for McDowell there is an important sense in which the world, in this scenario, would not cease to contain ‘red’ things.

According to McDowell’s account, even if human beings lose the capacity to experience ‘red’ things, this would not mean that red things had ceased to exist. Human beings without that capacity are defective. This is consistent with the theory of response-dependence, for when McDowell says that what it is for something to be red is for it to be such as to look red to human beings, he means for it to be such as to look red to non-defective human beings. Similarly, in the case of ‘cruel’, McDowell would say that for action A to be cruel is for it to be such as to merit certain subjective responses amongst non-defective human beings. This would ensure that, *even if we were somehow to lose the capacity for such responses*, this would not falsify the claim that action A is cruel.

---

<sup>41</sup> It is worth noting at this point that the same would also be true of what might more obviously be regarded as ‘social properties’, things like ‘money’, ‘contract’, ‘marriage’ etc.

I suspect that Davidson would have taken a similar line to McDowell on this issue ('red' objects would exist and 'cruel' objects would exist even if human beings lost the capacity to respond to them) but does not say enough in this context for me to be sure about this. Taking this sort of line however raises a number of thorny issues, and requires a substantial amount of further argumentation. More specifically, there are a host of issues which surround the notion of 'merit' which is being invoked, and around the notions of 'defective' and 'non-defective' human beings which are being invoked. Saying so little here simply does not help Davidson's case. If he does not think that he is at risk of getting embroiled in these issues, he needs to say considerably more about why this is so.

Davidson seems to think that the very question of whether or not value properties are contingent on the existence of human sensibilities or sense organs is a bad one. I find his general ambivalence in this area unsatisfying. At one point Davidson chides Wiggins for saying that the property of 'being red' is in an interesting sense a relative property as colour is a category that 'corresponds to an interest that can only take root in creatures with something approaching our own sensory apparatus' by remarking that 'in this sense all perspectival properties are relative' (Davidson 2004, 47). One wants to say that of course all perceptual properties are relative in this sense, but there are distinctions amongst these properties that are important and interesting. In what follows I will discuss one such distinction, and try and say something about why I think it is important in an ethical context.

### **3. Explanatory Differences**

Bernard Williams has argued influentially that the analogy between secondary properties and value properties breaks down in an important way. Williams argues that our *explanations* of secondary properties can and do in fact relate the way the world seems to us (or what is 'there to be experienced') to the way the world is according to physics (according to our best understanding of primary properties). He claims that our conception of secondary properties rests on the notion that in principle the perception of (say) colours can be explained in terms of perceptual psychology, on the one hand, and the world as characterised by primary properties, on the other (cf. Williams 1995, 177).

The idea is that the explanations of our experiences of secondary properties can show how they are related to physical reality, and how they can lead to knowledge of that reality. As he put the point some years earlier, 'the psychological capacities that underlie our perceiving the world in terms of secondary properties have evolved so that the physical world presents itself in reliable ways' (Williams 1985, 150).

Williams suggests that this is simply not the case when we reflect on value properties<sup>42</sup>.

There is nothing in our understanding of a thinker's ethical values ('what they think important, laudable, hateful, to be condemned or despised and so forth in the actions and reactions of human beings') which could correspond to the scientific explanations we have of our perception of secondary qualities, explanations which relate the way the world is to the way the world seems to us. When we consider the relation between value properties and the world, any explanation will not look at all like an explanation of colour properties.

This is because the concepts and standards that our explanations invoke are those involved in finding our way around not simply the physical world (the world of primary properties) but also a social world, *where this means one social world among many*. Williams thinks that an ethical life, and ethical thinking, is in part essentially local; its characteristic concepts are not intelligible independently of particular cultural perspectives. So, when we explain the relation between ethical properties and the world, the world must already be construed in a 'psychologically and socially richer sense' than the world of primary properties (cf. Williams 1995, 177). Consider a particular thick value concept that some people are disposed to use, the concept of blasphemy. Whether or not an action or utterance is counted as 'blasphemous' (whether or not the concept 'blasphemous' is correctly applied) must seem to the radical interpreter to depend as much on the social or cultural environment of the speaker as it does on the state of the physical surroundings. (This problem exists whether or not we allow that a thinker can stand in a causal relation to a property, 'blasphemous'.)

The point is not simply that primary properties underdetermine the ethical properties of different social worlds. It seems simply true, independent of any considerations about social worlds, that primary properties underdetermine evaluative properties. Williams's point is that ethical properties seem to lack even the clear relation to explanatorily primary properties that other secondary properties have. In this sense they seem to have something in common with 'social concepts'.<sup>43</sup> This matters when we try to explain differences between thinkers on ethical matters, when we try to explain 'why things are like this for them and not for us'. This particular point, I think, also casts doubt on the idea that ethical

---

<sup>42</sup> Davidson would claim that some ethical judgements are amongst the basic perceptual judgments of any thinker. But he would still need to show how the perception of ethical properties could be explained by perceptual psychology and the world characterised by primary properties in order to address the point made by Williams.

<sup>43</sup> Of course, the claim that ethical properties resemble 'social properties' in this way is controversial. Nothing I say against Davidson here hangs on the idea that ethical properties resemble social properties in any strong sense. The case against Davidson, or at least the main case, against turns, specifically, on worries about whether ethical properties can be akin to secondary qualities.

properties are likely candidates for featuring amongst the basic propositional attitudes of any thinker, that substantial ethical concerns provide ‘constraints on mindedness as such’. It seems plausible that a property that figures in a basic belief cannot be a property that objects instantiate only if relevant people think that objects instantiate, i.e. can’t be a social property like money or marriage. This is because the concepts that figure in basic perceptual beliefs do not seem to be like concepts such as marriage or money, because what they are concepts of—that is primary and secondary properties – do not seem to be response-dependent in the same way that social properties such as being married, and being a medium of exchange, are.

I have been attempting to argue towards a claim along the following lines. In order for Davidson’s position to be plausible, he would need to deny that ethical properties could have the status of primary properties. Further, Davidson needs to treat ethical properties as having the status of secondary properties and acknowledge that they have a degree of response dependence. The issue is how they can have this status, given the disanalogy between ethical properties and other secondary properties that Williams has raised. Not all secondary properties are the same, and we might think that there are significant difficulties between ethical properties and other types of secondary properties. Of course, that some ethical properties are problematic does not mean that all are.

One significant issue in this area (a more significant issue) is connected to the thought that Davidson would clearly not want to claim that *any* response-dependent property could be regarded as a *real* property (or a natural property) that could be objectively attributed to any entity or event. What it is for an entity or event to have a value property requires, on McDowell’s account, that the property is ‘there to be experienced, as opposed to being a mere figment of the subjective state that purports to be an experience of it’ (McDowell 1998, 134). I think most people would want to say *at least this* about ethical properties. This raises the question of how we can distinguish between properties that are ‘there to be experienced’ and properties that are ‘mere figments of our subjective states’. It seems difficult to address this question without making broadly ontological distinctions. (We want to say here: because that property *is* there to be experienced, while that property *is* a mere figment of a subjective state.)

The typical example of properties that are thought to be ‘mere figments of our subjective states’ would be those that thinkers experience during hallucinations. However, in ethical disputes it has often been thought very important that we are able to affirm or deny that

properties such as 'being a Kraut' or 'being chaste' (or simply 'cruel') are 'there to be experienced'. It is here that we encounter the type of disagreement that was referred to in Chapter 1, namely the sort of disagreement that can occur between those who do use a certain ethical concept and those who don't. This is an ethically important type of disagreement, and it is a disagreement about what exists.

There is a sense in which Davidson's type of argument, an a priori argument from interpretation, appears to be peculiarly impotent in this context. Davidson in fact acknowledges a specific problem here about drawing ontological consequences from semantics. He says: 'In describing the semantics, and hence the ontology, of a language we must perforce appeal to the entities we think exist. But these are just the entities that belong to the ontology of our own language; there is no way we can progress beyond our own resources by doing semantics' (Davidson 2005, 41).

In light of these remarks, we might consider a case described by Williams (cf. Williams 1995, 185). Williams asks us to imagine 'a people who are filled with terror, perhaps of a rather special, numinous kind, by certain features of their environment. They have a word that picks out things to which they react in this way. It is not a blankly causal, still less a merely individual reaction, and children are instructed in what does and does not merit it'. According to Williams, an interpreter could conceivably come to understand these reactions, and the word that picks out certain entities or events in terms of this reaction. The interpreter would not need to share the reaction, except to the extent that he would need to imaginatively enter into the speaker's view of things: he would not need to share 'the beliefs and attitudes which make this reaction intelligible'. Williams then asks: Is the property for which they have this term 'there to be perceived'? It is part of the speaker's world; it is not part of the interpreter's world. Is it part of *the* world? If general considerations about interpretation can tell us nothing about this case, and this case raises issues about ethical objectivity, then general considerations about interpretation cannot settle the issues of ethical objectivity to the extent that Davidson sometimes seems to suggest. That is to say Davidson's blanketing remarks about the necessary agreement on basic ontology that is required for interpretation to be possible seem to be of little help here. Of course *most* of the properties that we predicate of objects and events there to be experienced. But in this type of case we are after specifics.

I think that someone like McDowell would say that some ethical properties are there to be experienced, but can perhaps only be experienced if the thinker has the appropriate

upbringing or moral education etc. But I am sceptical as to whether this sort of neo-Aristotelian account could actually help Davidson, for the following reason. If concern with certain ethical concepts is connected to contingent facts about upbringing, then it is hard to see how a concern with such concepts could represent a constraint on mindedness as such. That is, if concern with certain ethical properties is only understood with reference to contingent facts about upbringing, then this is perhaps further evidence that such properties may correctly be understood as social properties, or properties of specific social worlds. And again, a property that figures in a basic belief, we might think, cannot be a property that objects instantiate only if relevant people think that objects instantiate.

Of course, any thinker must have some sort of upbringing, must have been brought up into some linguistic culture. The worry must be that having certain ethical beliefs requires some sort of upbringing over and above this one. Davidson could say, very plausibly, that any thinker must have some ethical beliefs, but there may be specific ethical concepts (such as piety) which require more particular kinds of upbringing. Whether this is all Davidson needs to say to secure the substantial ethical conclusions that he attempts to draw from his account of interpretation is discussed further in Chapter 4.

#### **4. Summary**

I have argued that in order to grant Davidson's general presumption we must accept that there are a class of 'basic beliefs' which must have some true instances because their content is determined by what causes them. Further, in order to make substantial claims about ethical judgments on the basis of this general presumption, it needs to be shown that ethical properties are such that they can feature amongst basic beliefs. I have suggested however that the nature of ethical properties should make us wary of this idea. More specifically, my case against Davidson turns on worries about whether ethical properties can be akin to secondary qualities. I have argued that Davidson has not addressed this issue in enough detail. Indeed, one criticism that I have made of Davidson is that he has not set out the nature and extent of his ethical claims in general in enough detail for us to be sure of their full implications.

Susan Hurley, in her excellent book '*Natural Reasons*' (Hurley 1989), argues from Davidsonian considerations about interpretation and charity to robust claims about ethical objectivity. Hurley argues more specifically that any attribution by an interpreter of desires and preferences to a speaker is inextricable from the attribution of some ethical values. That is to say that, according to Hurley, an interpreter would not be in a position to recognise a

speaker's basic desires or preferences, and so the formal values of coherence and consistency that they exhibit, unless they shared with the speaker certain substantive evaluative and, according to Hurley, recognisably ethical interests and concerns.

I will discuss Hurley's central argument in Chapter 4. I suggest that there are two broad approaches that such an argument could take. The first approach would not identify any particular set of ethical concepts or properties concern with which is required for a successful interpretation, but rather will state that there simply needs to be enough material to make the identification—without a specification of what that material will necessarily be. The second approach would attempt to identify some determinate set of ethical properties as central, that a human being must be concerned with in order to be regarded as a human being. Anything that lacks concern with these ethical properties would not be a human being, therefore anything that is properly interpreted as a human is concerned with these ethical properties. I suggest that either approach faces at least one substantial obstacle.

## **Chapter 4: Shared Values**

Davidson has often referred to the idea that any two speakers must share ‘a range of basic values’ in order for them to be mutually intelligible. It is clear that this much is thought to follow from Davidson’s account of radical interpretation. One idea which he emphasises in more than one place is that whether an interpreter is trying to understand the evaluative attitudes or the evaluative judgements of a speaker, ‘everything depends on our ability to find common ground’, and this common ground ‘is not subsequent to understanding but a condition of it’ (cf. Davidson 2004, 37, 51).

I have been looking at how Davidson might attempt to get from claims about radical interpretation to substantial claims about ethics. One criticism that I have made of Davidson is that he hasn’t set out the full nature and extent of his ethical claims in enough detail for us to be sure of their full implications. In this chapter I will consider what I regard as a development of Davidson’s approach that can be found in the work of Susan Hurley. I will then look at the more specific question of whether or not the argument from interpretation that Hurley presents could show that all speakers, as Bernard Williams has put it, must share in some more or less determinate form ‘the same materials of an ethical life’ (Williams 1995, 141). I conclude by suggesting that the sort of problems I raise for Hurley can be shown to be problems for the approach more generally.

### **1. Basic Values:**

Firstly, I am looking to shed some light the following question: what, if any, substantial ethical conclusions might we hope to establish from the common ground of basic values that are required for interpretation according to the argument from interpretation? In order to answer this question I will need to say a bit more about what might be meant by the ‘common ground of basic values’.

Perhaps I should start with some very obvious points. Whatever an answer to this question will look like, it must be compatible with the nature and existence of de facto disagreement on matters of value. Davidson does not think that all speakers in fact agree on all matters of value, or even that they would agree if they could just understand each other<sup>44</sup>.

---

<sup>44</sup> In one place Davidson does say that we should expect ‘enlightened values’ of speakers who fully understand each other to converge (Davidson 2004, 49). By enlightened values he means ‘the reasons we have for valuing when we have all the non-evaluative facts straight’. What exactly is meant to follow from these remarks is a little unclear. It would be very odd however, and out of line with

In one place Davidson says: 'Just as in coming to the best understanding of your beliefs I must find you coherent and correct, so I must match your values and mine: not of course in all matters, but in enough to give point to the differences. This is not to pretend or assume that we all agree...The only way of knowing what someone else's values are is one that builds on a common framework' (Davidson 2004, 36). Successful interpretation then should not be expected to eliminate all disagreement on all matters of value. This would be absurd. The thought is that successful interpretation makes meaningful disagreement possible, and that this depends on some foundation of agreement.

Davidson acknowledges that the different languages that speakers use allow them to express strikingly different ways of thinking. He acknowledges that there can be real differences in evaluative norms amongst speakers who understand each other, again as long as those differences are placed within a common framework. He also acknowledges that interpreters must allow for the fact that some values are further removed from the values the attribution of which is necessary to make minimally coherent sense of others. The thought seems to be that there are some 'basic' values which are constitutive of rational agency, and which presumably can be contrasted with other 'non-basic' values.<sup>45</sup> These basic values provide constraints on the eligibility of interpretations. How then might we draw the distinction between the basic values that all speakers must share and the other non-basic values? It is I think fair to say that the nature and extent of the distinction is not always particularly clear in Davidson's work. Indeed, the sometimes vague formulations of the principle of charity that were highlighted in Chapter 1 seem to be particularly problematic in the context of Davidson's remarks about values.

Hallvard Lillehammer has suggested that basic values are essentially to be understood as 'core values, as opposed to peripheral values, within a holistic network of values and beliefs' (Lillehammer 2007, 211). Lillehammer adds that while none of these values (or indeed beliefs) are to be regarded as analytically immune from revision, some will be more central to interpretation than others'. I think that this is right as far as it goes but that it doesn't go very far.

---

anything that Davidson says elsewhere, to take them as implying that all such values would converge absolutely and in all cases.

<sup>45</sup> I am not suggesting that we should expect a clean distinction in every case. In one place Davidson remarks that 'the more basic a norm is to our making sense of an agent, the less content we can give to the idea that we disagree with respect to that norm' (Davidson 2004, 50). This is suggestive of a sliding scale.

In the same paper Lillehammer does remark, and with some justification, that whenever Davidson distinguishes between basic values (or 'structurally important' values) and non-basic values, he tends to 'illustrate the former by mentioning the paradigmatically formal values of consistency and coherence' that regulate the attribution of desires and preferences to the speaker. Davidson certainly thinks that certain formal evaluative constraints on intelligibility are constitutive of agency, and the discussion of these formal constraints is prominent in his work. For example, in one place he says that 'the constraints that a Bayesian theory of preference places on the pattern of beliefs and evaluations exert a prima facie claim on interpretation; consistency of preferences with one another and with beliefs is a constitutive pressure on interpretation simply because we cannot rationalise (i.e. explain or understand) deviations from it' (Davidson 2004, 71). We can contrast this with Davidson explicitly stating that, for example, conceptions of justice and fairness have 'no favoured role' in our attributions of propositional attitudes to others (Davidson 2004, 74).

It might seem then that there is some textual support for attributing to Davidson the following view: while the more formal values of coherence and logical consistency are the basic values that are constitutive of rational agency, more substantial values (or more specifically ethical values such as a concern for liberty, justice or fairness) are non-basic and play no role in interpretation. If this were so, we might perhaps expect the extent and nature of the shared values required for interpretation to leave substantial ethical matters pretty much untouched. Of course, to say that conceptions of justice and fairness 'play no favoured role' in interpretation is not to say that concern for justice or liberty plays no role at all. In fact however I think that this later claim would be in line with a fairly common way of reading Davidson, which takes basic values to be essentially the formal values of consistency and coherence, and pays less attention to the role that substantial values might play in interpretation.

On this reading it might look like a mistake to think that substantial ethical conclusions could be derived from the consideration of radical interpretation. This however would leave Davidson's explicit claims about ethics looking strangely weak. I take it that Davidson must want to say, and indeed does say, more than this. I take it that he wants to say, at least, that consideration of radical interpretation can lead us to the conclusion that ethical judgments are objectively true or false, and that there is more and deeper agreement on them than there may often seem to be.

Susan Hurley has argued that there are good reasons for not construing the distinction between basic and non-basic values in the way that I have just sketched, that is, as a distinction between formal and substantive values. These reasons stem from the idea, articulated powerfully by Hurley in work that draws heavily on Davidson's own, that an interpreter would not be in a position to recognise a speaker's basic desires or preferences, and so the formal values of coherence and consistency that they exhibit, unless they shared with the speaker certain substantive evaluative and, according to Hurley, recognisably ethical interests and concerns<sup>46</sup>. That is to say that, according to Hurley, the basic values two speakers must share in order to be mutually intelligible must provide substantive and in part ethical constraints as well as formal constraints on interpretation.

## **2. Hurley's Argument from Interpretation:**

In his later work on radical interpretation Davidson increasingly emphasised the relationship between the theory of interpretation and formal decision theory. In one essay he in fact suggests altering the evidential basis of radical interpretation from sentences held true by the speaker to sentences preferred true by the speaker (cf. Davidson 2004, 13-34). The ethical implications of this were never fully explored in Davidson's own work.

However, the nature of the relationship between the theory of interpretation, formal decision theory, and ethics is explored in detail by Hurley in her book *'Natural Reasons'* (Hurley 1989). One thesis that Hurley argues for in this remarkable and sadly neglected book is that any attribution of desires and preferences to a speaker is inextricable from the attribution of some ethical values. This may in fact be a thought that finds fleeting expression in Davidson's work, though it is certainly never argued in as much detail as it is in Hurley's work.

According to Hurley, any interpreter must attribute substantive and recognisably ethical evaluative interests (values?) to a speaker even in the course of crediting them with basic desires and preferences. This is because it is only by reference to some sense of the speaker's substantive and ethical values, to what the speaker thinks important and worthwhile, that an interpreter could give content to those desires and preferences. The following thought is that the speaker's values must be values that the interpreter could

---

<sup>46</sup> I am following Hurley by using 'substantive' simply to mean 'concerning issues of substance and content rather than form, procedure and consistency' (Hurley 1989, 389).

share, or at least make sense to the interpreter as values that a speaker qua speaker could have.

Hurley's argument stems from a claim about decision theory. Davidson has famously suggested that the structural constraints on preferences invoked by the likes of Frank Ramsey and Richard Jeffery in their work on decision theory could be employed in radical interpretation. But as Hurley puts it, and as Davidson himself clearly recognised<sup>47</sup>, decision theory does not interpret itself. Hurley argues that a decision theorist (and so, for her purposes and ours, an interpreter) must at least tacitly recognise 'a variety of naturalistic and social substantive constraints on desires and preferences, in addition to formal principles of rational coherence' (Hurley 1989, 98). Hurley brings out this point by discussing what she calls, following David Lewis, 'the problem of eligibility of interpretations'. I will briefly set out the problem and Hurley's response. The problem is best illustrated by way of examples. Here is a very simple manifestation of the problem that Hurley herself discusses.

It is generally held that the transitivity of preference is a fundamental axiom of decision theory. If we encounter a speaker who prefers pears to oranges and oranges to apples, then we may assume that the speaker does not (or at least should not) prefer apples to pears. The thought is that this axiom helps us attribute content to the thoughts and utterances of others. But suppose in fact the speaker goes on to exhibit various preferences which seem to be intransitive. Has the axiom been discredited? Not necessarily, as Hurley explains. A defender of the axiom could suggest that the original account of the speaker's preferences was simply incorrect. Perhaps the speaker prefers pears to oranges and oranges to red apples, but prefers green apples above all. Or perhaps the speaker only prefers pears with leaves on to oranges. If the pear has no leaves on it he is indifferent between pears and oranges. Or perhaps the speaker has just changed his mind. He preferred pears to oranges at  $t$ , but not at  $t_1$ .

The manoeuvre in each case is to suggest more finely individuated alternatives or preferences in order to satisfy the axiom. At some point however, as Hurley points out, we will balk at this sort of reinterpretation. At some point we will cross the line between interpretations of the preferences of the speaker which we regard as eligible and those

---

<sup>47</sup> Cf. Davidson 2001b, 147: 'A radical theory of decision must include a theory of interpretation and cannot presuppose it'. For Davidson on the specific problems of using decision theory in the process of radical interpretation, cf. Davidson 2004, 29-34. The problem I discuss here applies to decision theory more generally. Hurley refers approvingly Tversky's suggestion that 'decision theorists are eager to tell people how to act, in light of their values. They are typically reluctant to tell them how to feel or what to value' (Hurley 1989, 105).

which we regard as ineligible. Further, the possibility of our being able to refute or confirm the axiom at all depends on this distinction<sup>48</sup>.

This worry is linked to a more general worry about interpretation, the thought that every interpretation is subject to further interpretation. The general idea that is supposed to stop this worry from getting a grip, both in Davidson and in Hurley, is that the interpreter must recognise constitutive constraints on interpretation, provided in part by what Hurley refers to as 'the normal relations of the mind to the world' (Hurley 1989, 84). It is only in virtue of these constraints that mental states have the contents they have, or could have determinate content at all. The specific idea here is that in order to judge whether the speaker's preferences are consistent or inconsistent, the interpreter needs to be able to distinguish between eligible and ineligible interpretations of the speaker's actions, and so distinguish between eligible and ineligible contents of the speaker's preferences that these actions express. Perhaps there are strict and determinate rules for the distinction between eligible and ineligible constraints in every case.

However, according to Hurley *any* distinction between eligible and ineligible constraints on preferences requires substantive and recognisably ethical constraints as well as formal constraints on the contents of those preferences. According to Hurley, there is simply 'no good reason for excluding from the class of distinctions that respect constraints on eligibility those provided by applications of evaluative or reason giving concepts to the alternatives the agent faces' (1989, 99).<sup>49</sup>

I have previously spoken at some length of the way that the principle of correspondence emphasises the significance of the causal connections that link the interpreter, the speaker, and the shared world. According to Hurley, as well as being interested in these causal connections, thinkers are interested in 'making sense of themselves and one another, in trying to find one another intelligible. Applications of reason giving and evaluative concepts make an essential contribution to the kind of intelligibility and understanding that we seek as persons...To eschew the use of these concepts...would be to frustrate this interest, and

---

<sup>48</sup> According to Hurley the point applies to both descriptive and normative theories of decision. She says: 'a descriptive theory should be refutable, or it will lack empirical content, and a normative theory should be possible to violate, otherwise it will provide no constraints on action at all' (Hurley 1989, 59).

<sup>49</sup> Earlier, Hurley speaks of the 'need to distinguish between characterisations of alternatives or states of affairs that are intelligibly related to values and human goods, thus potentially reason giving, and characterisations which are none of these things' (Hurley 1989, 87).

distort the attitudes to be understood, which are constitutively informed by this interest' (Hurley 1989, 100).<sup>50</sup>

I take it that the general direction of the above remarks is as follows. Something more than just formal constraints is required to distinguish between eligible and ineligible interpretations of the contents of the speaker's preferences. What is required, according to Hurley, can be regarded as some sort of understanding of what might be called a thinker's ethical nature. According to Hurley, any such understanding of ethical nature must allow 'determinations on eligibility with respect to...preferences to be sensitive to distinctions that it is normal and natural to recognise or care about, in the context of familiar reason giving practices and forms of life, or which have an intelligible function in relation to human society or human flourishing' (Hurley 1989, 113). These determinations on eligibility will be constrained by specifically and recognisably ethical considerations. Hurley maintains that specific and recognisably ethical considerations, such as considerations of courage or fairness, contribute to the interpreter's ability to distinguish between eligible and ineligible preferences. Such ethical considerations play a constitutive role in interpretation.

Hurley's claim then is that in order to recognise a speaker as having rationally structured desires or preferences—and therefore in order to recognise them as a speaker at all—the interpreter must share with them some set basic ethical nature which make those desires and preferences explicable. This is an argument from interpretation of the form that we would associate with Davidson: it is claimed that it is an a priori matter, not up to us, that mental states are thus constrained, as without the constraints determinate content would not be possible at all.

### **3. Problems facing the Argument from Interpretation:**

Hurley's account<sup>51</sup> has come under attack, notably by Bernard Williams in his essay '*Saint-Just's Illusion*' (Williams 1995). In this paper Williams expresses his scepticism of the thought that we could get from an argument from interpretation (of the sort that Hurley provides) to the conclusion that all speakers must 'share the same materials of an ethical life' (Williams 1995, 141). The extent to which this scepticism is valid will depend on what exactly Williams

---

<sup>50</sup> Hurley says more about the nature of causal explanation and rational explanation at Hurley 1989, 97. What she says here is much in line with Davidson's own remarks.

<sup>51</sup> Williams is reluctant to attribute Hurley's thesis, as he sets it out, to Davidson. I think that Davidson must agree with Hurley that the attribution of preferences to a speaker is inextricable from the attribution of some substantial values, and for reasons along the lines of those provided by Hurley. It is less clear that Davidson would agree with Hurley on what might follow from this.

means by the same materials of an ethical life. Primarily, Williams seems to be concerned with the question of how constraints on interpretation can guide the interpreter's attempts to make sense of the speaker's ethical nature, by which he means 'what they think important, laudable, hateful, to be condemned or despised and so forth in the actions and reactions of human beings' (Williams 1995, 139).

He takes Hurley's position to be roughly as follows. The interpreter may find considerable differences between the speaker's values and his own. Thus a speaker from one culture may esteem certain qualities more than we do and certain others less. Williams suggests, for example, that we should not be surprised if there is a special emphasis put on certain kinds of courage or certain forms of solidarity among speakers in a culture where people have to hunt for their food (Williams 1995, 141).

The interpreter may even be able to understand that the speaker has values which are not candidates for serious consideration given his own (the interpreter's) contingent historical circumstances. However, underneath any local variations (such as different rankings of preferences perhaps) there is a common human ethical sensibility of a fairly structured kind (at a deeper level there are no really alien values). Local variations are thus to be understood as surface adaptations of a shared ethical nature. Crucially, according to Hurley, this much is supposed to follow 'from its being an argument from interpretation – this must be so, because it represents a condition of understanding these people's lives as human at all' (Williams 1995, 141).

Williams thinks that Hurley, or anyone else working from an *a priori* argument from interpretation, will be poorly placed to make such a claim. He is happy to acknowledge that in interpreting other people 'we have to take it that they and we have a good amount in common'. He accepts that we can take the Davidsonian point that for an interpreter to interpret the speech or actions of any thinker, he must suppose that their beliefs and desires can be incorporated in a pattern that is in essential respects like the pattern of his own beliefs and desires. Further, Williams acknowledges that it may well be that some of what they have in common might be, if in a schematic form, some *ethical* values. He gives the example of a concern with justice in some highly indeterminate form (Williams 1995, 138).

However Williams argues that while an interpreter must assume a certain amount in common between himself and the speaker when he is trying to make sense of the speaker, it does not follow that part of what we must assume in common is in any substantial or determinate sense an ethical life, or, to repeat the phrase Williams uses, the shared

materials of an ethical life. He allows of course that any interpreter must use his own ethical concepts and reason giving concepts, such as they are, in the interpretation of others<sup>52</sup>. But Williams is sceptical about the thought that an argument from interpretation could show that all speakers must share some specific and substantial set of ethical concerns, concepts or attitudes (this is essentially what I take Williams to mean by the 'materials of an ethical life'). His point is that this is what is required for Hurley's thesis to be able to provide substantial conclusions in ethics, specifically regarding issues of objectivity.

As Williams puts it, while 'to see others as sharing a nature with ourselves is the merest decency', it does not tell us anything substantial about the extent of our shared ethical lives. Quoting Clifford Geertz, he suggests that the problem that the interpreter faces is to deploy his own concepts, some of which are nearer to the speaker's and some further away 'so as to produce an interpretation of the way a people lives which is neither imprisoned within their mental horizons, an ethnography of witchcraft as written by a witch, nor systematically deaf to the distinctive tonalities of their existence, an ethnography of witchcraft written by a geometer'. Williams goes on: 'One must bring with one beliefs, models, patterns of explanation (and concepts): it is they that will determine, in the end, how much identity there will be, and at what levels' (Williams 1995, p143). That two speakers can talk to each other, and talk to each other about ethical judgements, does not show that they share, in an important sense, the same materials of an ethical life. The point is not that different ethical languages are mutually unintelligible. The interpreter can to some extent come to understand what the speaker, in the sense that they can assign meanings to their sentences and come to predict their behaviour. The point is that this level of Intelligibility does not guarantee homogeneity of ethical concepts, which Williams thinks is a requirement for ethical objectivity.

Williams is clearly sceptical of the idea that any a priori argument from interpretation, such as Davidson's or Hurley's, could lead to substantial conclusions about a shared ethical nature. The key word here seems to be 'substantial'. This scepticism, regarding the limits of what philosophy can tell us about ethics, runs deep in Williams's work.

---

<sup>52</sup> When David Lewis says that the speaker should be represented as believing what he ought to believe and desiring what he ought to desire, he goes on: 'And what is that? In our opinion, he ought to believe what we believe and desire, or perhaps what we would have believed and desired in his place. (But that is only your opinion! Yes. Better we should go by an opinion we don't hold?)' (Hurley 1989, 24). This is surely right but also surely obvious.

In various places he has pointed out that questions regarding the nature and extent of our shared ethical makeup have remained largely unanswered and recur in various other disciplines, such as cultural anthropology, comparative linguistics, and cognitive psychology. He has also remarked that the fact that these questions remain, and in such recalcitrant and dispersed forms, discourages him from even allowing for the possibility that they might suddenly be answered by an a priori philosophical argument. His specific response to Hurley's argument is telling in this respect. He says: 'It is not so much that I do not believe it to be true. It is rather that I cannot believe that it has to be true, that reflection on the demands of interpretation should be able to lead to so substantive a conclusion' (Williams 1995, 141).

There is however something more specific about Williams's scepticism than these remarks suggest, something that raises specific problems for the type of argument that Hurley is trying to develop. As I have just mentioned, Williams thinks that Hurley, or anyone else for that matter who is working from an a priori argument from interpretation, will be poorly placed to make substantial claims about the ethical lives of thinkers as such. This seems to be at least in part because he thinks that our best understanding of our ethical lives must essentially involve the consideration of other factors, such as cultural or historical circumstances.

Williams is sceptical of the claim, which he seems to attribute to Hurley or at least take as implied by her argument, that anything an interpreter can recognise as an ethical concern 'must be capable of being mapped onto a structure in which it will intelligibly be related in ethical terms' to ethical concerns that the interpreter himself accepts, as an 'application, extension, limitation, or so forth' (Williams 1995, 210).<sup>53</sup>

As was mentioned in Chapter 3, Williams is often concerned to emphasise the extent to which an ethical life, and ethical thinking, is in part essentially local; the extent to which many of the characteristic concepts of an ethical life may not be fully intelligible independently of particular social or cultural perspectives. Now it seems undeniable, and is indeed much commented on, that the speaker and the interpreter can inhabit, in one sense, very different social and cultural worlds, and that there may be many other different social and cultural worlds. These different social and cultural worlds can include and prioritise very different broadly ethical concepts and attitudes, different ethical interests and priorities. The

---

<sup>53</sup> This is one way of understanding Hurley's claim that by 'constraining scepticism to operate within the conceptual scheme we have, we achieve a kind of objectivism' (Hurley 1989, 57).

claim I am putting forward, and that I think finds expression in Williams's work, is that there is a substantial sense in which speakers who inhabit different social and cultural worlds may be said to have very different ethical lives.

That this is so is important because the different concepts and interests which can shape an ethical life will validate different reason giving explanations and rationalisations for action. In any given situation, the practical judgments which follow from the application of the specific set of ethical concepts used in one culture may not be the same as the practical judgements that follow from the application of the specific ethical concepts used in another culture. The different and differently recognised reason giving and action guiding consequences of certain concepts make this so. And, wherever different speakers use concepts that have incompatible reason giving or action guiding consequences, the question must arise about which of the two concepts, if any, they are to be practically guided by.

I have suggested that this type of question, the type of question that can occur between those who do use a certain ethical concept and those who don't, is particularly important in an ethical context. What we say about this type of disagreement is important because, as Williams has pointed out, one of the things that people want when they talk about objectivity in ethics is for there to be 'one canonical, homogeneous ethical language. They want it to be conceptually homogeneous across cultures, and across disagreements within our culture' (Williams, 1995a, 240). I have also suggested that an a priori argument that takes the consideration of radical interpretation as its starting point will be peculiarly impotent when it comes to addressing this type of question. That is in part because in order to address such questions it appears that we would need to know significantly more about *which* ethical concerns must be shared, rather than simply that *some* ethical concerns must be shared.

Further, as was suggested in Chapter 3, a concern with some specific ethical concepts at least appear to require a specific kind of upbringing. If some ethical concepts require different kinds of upbringing, and these concepts are legitimately regarded as forming part of 'an ethical life', then we have an explanation of why we might lack the relevant conceptual homogeneity that Williams thinks is required for objectivity in ethics.

Of course, Hurley does not deny that there can be substantial differences between the ethical concerns of different speakers. On this point it is very important not to misinterpret the issue between Hurley and Williams. Though her thesis is that interpretation results in a certain form of ethical objectivity, Hurley emphasises repeatedly throughout her book that

her position is compatible with, and indeed facilitates, a wide pluralism in ethics. While the exact nature of this pluralism is difficult to define, she certainly does not think that her position could show us a way to decide, in those cases where a decision is called for, what the right decision is or what we ought to do – she is not attempting to provide an ethical theory that could be axiomatised or presented in a deductive form.<sup>54</sup> Hurley sums up her position here in the following way: ‘I have emphasised that the evaluative constraints on eligibility do not impose controversial solutions to ethical problems when fundamental values conflict, but rather provide a way of determining what the alternatives and hence issues are to begin with’ (Hurley 1989, 125).<sup>55</sup>

The issue between Hurley and Williams hinges on whether or not she has achieved this, and what might follow from it if she has. According to Williams, when we ask what it is that underlies the variety of human ethical practice, the truth is that ‘we simply don’t have a very good idea what an answer could look like’, though he is willing to claim that we shouldn’t expect too much from any answer we might have. He remarks in one place, with typical dryness, that our common ethical nature seems unhelpfully to underdetermine ethical matters just to the extent that leaves open all the disagreement that it was hoped it would resolve. It is trivial that all human beings share a nature to some degree. The question we are considering is whether or not they share an ethical nature, and what can be shown to follow from the fact that they do.

In order for an account of our shared ethical nature to tell us something more substantial, that is something that will enable us to address issues about ethical properties and ethical objectivity, it needs to be a lot clearer what our shared ethical nature might look like. One obvious way around this problem would be to identify some determinate set of ethical concepts or properties as central, that a human being must be concerned with in order to be regarded as a human being. Anything that lacks concern with these ethical properties would then not be a human being, therefore anything that is properly interpreted as a human is concerned with these ethical properties.

---

<sup>54</sup> This should not be surprising. As David Wiggins has remarked, ‘even philosophers of objectivist formation have constantly stressed...the absence of unique solutions and unique determinations of the practical that naïve cognitivism (or realism) would have predicted’ (Wiggins 1998, 127).

<sup>55</sup> In a similar vein, John McDowell has said: ‘Where hard cases occur, the agreement that constitutes the background against which we can see what happens as e.g. disputes about genuine questions, cannot be agreement in the application of the concepts themselves. What matters is agreement about what counts as reasonable argument: consider how lawyers recognise competence in their fellows, in spite of disagreement over hard cases.’

Once we do this however, we are faced with an issue which was raised in Chapter 1. The issue was presented as follows. The radical interpreter is a cognitive imperialist of sorts, who imports his own categories and concepts, beliefs and principles, and applies them to the utterances and behaviour of the speaker. This however seems to neglect the possibility that he may encounter a speaker who has categories and concepts, beliefs and principles that are somehow radically different from his own.

The trouble with this is that once we identify some determinate set of ethical concepts that is sufficiently substantial to allow us to make progress on ethical questions, we seem to leave ourselves open to the accusation that contingent psychological or anthropological assumptions (that is, constraints that pertain only to minds of a particular class) are being smuggled in amongst what are being passed off as necessary constraints on mindedness as such. Anything that Hurley suggests here would have to be sufficiently general so as to avoid these accusations, whilst also being sufficiently substantial so as to be informative. To say simply that successful interpretation makes meaningful disagreement on ethical matters possible, and that this depends on some foundation of agreement, is to say almost nothing.

It seems in fact that any attempt to derive substantial ethical conclusions from an argument from interpretation can take two forms, each of which faces a substantial difficulty. The first approach would not identify any particular set of ethical concepts or properties concern with which is required for a successful interpretation, but rather will state that there simply needs to be enough material to make the identification—without a specification of what that material will necessarily be. I have suggested that this is somewhat unsatisfying, and seems to leave open a whole host of questions regarding the objectivity of ethical judgments.

In order to avoid this problem there is another approach. This approach would be to identify some determinate set of ethical properties as central, that a human being must be concerned with in order to be regarded as a human being. Anything that lacks concern with these ethical properties would not be a human being, therefore anything that is properly interpreted as a human is concerned with these ethical properties. This approach runs into the sort of difficulties that were raised in Chapter 1 with regards to the smuggling in of contingent ethical concerns, casting doubt on the objectivity of the conclusions reached.

## ***Conclusions:***

The broad aim of this work was to outline, assess and elaborate on Davidson's work on radical interpretation and its connections to his understanding of values, and particularly on his understanding of the nature of ethical judgments and concepts. The central idea that I consider is that the principle of charity must play much the same role in the attribution of values to a speaker as it does in the attribution of beliefs. This is taken to show that certain general claims about the content of propositional attitudes, which Davidson thinks follow from the consideration of radical interpretation, can be applied to values as well as beliefs.

I have suggested that there is *prima facie* plausibility to the idea that the role of the principle of charity in the attribution of values to a speaker will be similar to the role of the principle of charity in the attribution of beliefs to the speaker. However I argued there are important limitations to Davidson's approach as it is set out, particularly with regards to certain claims that he wants to make about ethical judgements. It is by no means clear that we can get from Davidson's account of radical interpretation to the substantial ethical claims that Davidson wants to make – specifically that ethical judgments must be objectively true or false, and that many indeed must be true.

One criticism that I have made of Davidson is that he has not set out the nature and extent of his ethical claims in general in enough detail for us to be sure of their full implications. A substantial elaboration of Davidson's general approach can be found in the work of Susan Hurley. Hurley argues more specifically that any attribution by an interpreter of desires and preferences to a speaker is inextricable from the attribution of some ethical values. I have argued that the consideration of the type of problems that Hurley's account faces suggests that any attempt to derive substantial ethical conclusions from an argument from interpretation can take one of two forms, each of which faces a substantial difficulty.

The first approach would not identify any particular set of ethical concepts or properties concern with which is required for a successful interpretation, but rather will state that there simply needs to be enough material to make the identification—without a specification of what that material will necessarily be. I have suggested that this is somewhat unsatisfying, and seems to leave open a whole host of questions regarding the objectivity of ethical judgments.

In order to avoid this problem there is another approach. This approach would be to identify some determinate set of ethical properties as central, that a human being must be

concerned with in order to be regarded as a human being. Anything that lacks concern with these ethical properties would not be a human being, therefore anything that is properly interpreted as a human is concerned with these ethical properties. This approach however runs into difficulties with regards to the objection that contingent ethical concerns are being included amongst constraints on minds as such, casting doubt on the objectivity of the conclusions reached.

## ***Bibliography:***

### **Works by Donald Davidson:**

- DAVIDSON, D. 2001a. *Essays on Actions and Events*. 2<sup>nd</sup> ed. Oxford: Clarendon Press.
- DAVIDSON, D. 2001b. *Inquiries into Truth and Interpretation*. 2<sup>nd</sup> ed. Oxford: Clarendon Press.
- DAVIDSON, D. 2001c. *Subjective, Objective*. Oxford: Clarendon Press.
- DAVIDSON, D. 2004. *Problems of Rationality*. Oxford: Clarendon Press.
- DAVIDSON, D. 2005a. *Truth, Language and History: Philosophical Essays*. Oxford: Clarendon Press.
- DAVIDSON, D. 2005b. *Truth and Predication*. Cambridge: Belknap Press.
- LUDWIG, K. and LEPORE, E. ed. 2006. *The Essential Davidson*. New York: Oxford University Press.

### **Further Literature:**

- BLACKBURN, S. 1981. Intersubjective. Reply: Rule-Following and Moral Realism. In: HOLTZMAN & LEICH ed. *Wittgenstein: To Follow a Rule*. London: Routledge.
- BLACKBURN, S. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press.
- BLACKBURN, S. 1985. *Errors and the Phenomenology of Value*. In HONDERICH, T. ed. *Objectivity and Morality*, London: Routledge & Kegan Paul.
- BLACKBURN, S. 1999. Is Objective Moral Justification Possible on a Quasi-realist Foundation? *Inquiry*. Vol. 42. pp. 213–28.
- GLOCK, H. J. 2003. *Quine and Davidson on Language: Thought and Reality*. Cambridge: Cambridge University Press.
- HARE, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- HEIL, J. and MELE, A. ed. 1993. *Mental Causation*. Oxford: Clarendon Press.
- HURLEY, P. 2002. A Davidsonian Reconciliation of Internalism, Objectivity, and the Belief-Desire Theory. *The Journal of Ethics*. 6. pp.1-20.
- HURLEY, S. 1985. Objectify and Disagreement. In HONDERICH, T. ed. *Morality and Objectivity*. London: Routledge and Kegan Paul.
- HURLEY, S. 1989. *Natural Reasons: Personality and Polity*. New York: Oxford University Press.
- KIM, J. 1993. *Supervenience and Mind*. Cambridge: Cambridge University Press.
- LEPORE, E. and K. LUDWIG. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Oxford University Press.

- LEPORE, E. and MCLAUGHLIN, B. ed. 1985. *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Basil Blackwell.
- LEPORE, E. ed. 1986. *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.
- LEWIS, D. 1974. Radical Interpretation. *Synthese*, 27: 331-44.
- LEWIS, D. 1989. Dispositional Theories of Value. Proceedings of the Aristotelian Society. *Supplementary Volumes*. Vol. 63.
- LILLEHAMMER, H. 2003. Debunking Morality: evolutionary naturalism and moral error theory', *Biology and Philosophy* Vol. 18. Pp. 567–581.
- LILLEHAMMER, H. 2007. Davidson on Value and Objectivity. *Dialectica*. Vol 61. Issue 2, pp. 203–217.
- LUDWIG, K. 2003b. *Rationality, Language and the Principle of Charity*. In MELE & RAWLING ed. *The Oxford Handbook of Rationality*, Oxford: Oxford University Press.
- LUDWIG, K. 2007. Book Review. *Mind*. Vol. 116. p405.
- LUDWIG, K. ed. 2003a. *Donald Davidson*. New York: Cambridge University Press.
- LUKES, S. and HOLIS, M. ed. 1982. *Rationality and Relativism*. Blackwell.
- MACKIE, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- MALPAS, J. 2003. Donald Davidson. *The Stanford Encyclopaedia of Philosophy*. Winter ed. ZALTA, E.N. ed. URL = <http://plato.stanford.edu/archives/win2003/entries/davidson/>.
- MCDONALD, C. and MCDONALD, G. ed. 2006. *McDowell and His Critics*. Blackwell Publishing.
- MCDOWELL, J. 1981. Non-Cognitivism and Rule-Following. In HOLTZMAN & LEICH ed. *Wittgenstein: To Follow a Rule*. London: Routledge. 1981.
- MCDOWELL, J. 1985. Values and Secondary Qualities. In HONDERICH, T. ed. *Objectivity and Morality*. London: Routledge & Kegan Paul.
- MCDOWELL, I. 1998. *Mind, Value and Reality*. Harvard University Press.
- MCDOWELL, J. 2009. *The Engaged Intellect*. Harvard University Press.
- MCGINN, M. 1997. *Wittgenstein and the Philosophical Investigations*. London: Routledge.
- MILLAR, A. 2004. *Understanding People: Normativity and Rationalizing Explanation*. Oxford: Oxford University Press.
- MYERS, R. H. 2004. Finding Value in Davidson. *Canadian Journal of Philosophy*. Vol 34. pp.107-136.
- NIETZSCHE, F. 2003 (1973). *Beyond Good and Evil*. Penguin Books.

- PETTIT, P. 1981. Reply: Evaluative 'Realism' and Interpretation. In HOLTZMAN & LEICH ed. *Wittgenstein: To Follow a Rule*. London: Routledge. 1981.
- PETTIT, P. 2001. Embracing Objectivity in Ethics. In LEITER ed. *Objectivity in Law and Morals*. Cambridge: Cambridge University Press.
- PUTMAN, H. 1981. *Reason, Truth and History*.
- QUINE, W. V. O. 1960. *Word and Object*. Cambridge. M.I.T. Press.
- RAMBERG, B. 1989. *Donald Davidson's Philosophy of Language: An Introduction*. Oxford: Basil Blackwell.
- RORTY, R. 1979. *Philosophy and the Mirror of Nature*. New Jersey: Princeton University Press.
- SMITH, M. *On the Nature and Significance of the Distinction between Thick and Thin Concepts*.
- STROUD, B. 2000. *Understanding Human Knowledge: Philosophical Essays*. Oxford: Oxford University Press.
- WIGGINS, D. 1998. *Needs, Values, Truth*. (Third Edition). Oxford University Press.
- WIGGINS, D. 2006. *Ethics: Twelve Lectures on the Philosophy of Morality*. Penguin Books.
- WILLIAMS, B. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- WILLIAMS, B. 1985. *Ethics and the Limits of Philosophy*. London: Fontana Press.
- WILLIAMS, B. 1995. *Making Sense of Humanity*. Cambridge: Cambridge University Press.
- WILLIAMS, B. 1995a. *Truth in Ethics*. Blackwell Publishers Ltd.
- WILLIAMS, B. 2005. *In the Beginning was the Deed: Realism and Moralism in Political Argument*. HAWTHORN, G. ed. New Jersey: Princeton University Press.