Davidson on Value and Objectivity

HALLVARD LILLEHAMMER

ABSTRACT

In a series of articles, Donald Davidson has argued that the objectivity of value is entailed by the objectivity of propositional attitude ascriptions. In this paper I explore the scope and limits of Davidson's argument with particular reference to his claim that there would be convergence among the values of enlightened agents who fully understood one another. I argue that Davidson's argument supports the convergence claim only if that claim is restricted to a limited range of comparatively basic values. I show that this restriction of the convergence claim is in tension with Davidson's contention that value judgements are objective in much the same way as factual judgements are.

1. Values out of mind

According to one version of objectivism about value, ethical and other evaluative claims have a fixed truth-value independently of who makes them or the society in which they happen to live (c.f. Davidson 2004, 42). Subjectivists about value deny this claim. According to subjectivism so understood, ethical and other evaluative claims have no fixed truth-value, either because their truth-value is dependent on who makes them, or because they have no truth-value at all.

In a series of articles, Donald Davidson has argued that the objectivity of value follows from the objective truth of propositional attitude ascriptions, the latter being understood in terms of an account of mental content as determined by constraints embodied in a process of radical interpretation (Davidson 2004, 19-74). According to this

argument, necessary constraints on the attribution of mental content entail the attribution to agents of minimal standards of rationality, generally true beliefs, and a core set of correct (or what Davidson alternatively calls 'real') values (Davidson 2004, 50). A number of other philosophers have been influenced by Davidson's work to make similar claims, including Susan Hurley, Paul Hurley, and R. H. Myers.[1]

Davidson's argument is powerful and illuminating. Yet neither his own work, nor the work of his followers, does much to establish its precise implications. In what follows, I shall examine Davidson's argument for the objectivity of value and evaluate its scope and limits. In doing so, I will not question the controversial interpretationist account of the mental on which Davidson's argument depends.[2] Instead, I will ask what exactly we are entitled to infer from it. In particular, I will examine Davidson's claim that there would be convergence among the values of enlightened agents who fully understood one another. I will argue that for all Davidson's argument shows, there is an indefinite range of substantial evaluative judgements (including paradigmatic moral judgements) that have no determinate truth-value, and are therefore arguably not objective in the sense suggested by Davidson's use of that term.

2. Davidson's argument

Davidson's case for the objectivity of value can be formulated as a transcendental argument. The argument begins from the assumption that propositional attitudes with

---

[1] See Hurley 1989, 189ff; Hurley 2002; and Myers 2004. S. Hurley's account is the most comprehensive of these and goes beyond the range of Davidson's in a number of ways (including her discussion of substantial ethical, political, and legal deliberation). David Lewis considers, but goes on to reject, a Davidsonian account of the relation between mind and value as 'intelligible and interesting, but too good to be true' in Lewis 1989, 135. C.f. Lewis 1996, 307.

[2] One canonical articulation of Davidson's account is found in Davidson 1984, 123-179. The view is articulated further in Davidson 2001 and Davidson 2004, 75-166.

determinate contents can be successfully attributed to some agents. The argument then proceeds by making explicit what the necessary presuppositions of such attributions are. According to Davidson, the successful attribution of propositional attitudes presupposes the existence of at least two agents, each responding in similar ways to similar features of a shared environment, including the responses of other agents. In this way, the fact of interpretability is taken to rule out scepticism both about the existence of other minds and about the external world.[3] Propositional attitude attributions proceed by assigning contents to mental states partly in terms of what is perceived to be their relevant causes. Davidson claims that one agent, the interpreter, can only assign determinate propositional attitudes to another agent, the interpreted, if at least two further conditions are met. These conditions are captured in terms of two principles that in one place Davidson calls 'the principle of coherence' and 'the principle of correspondence' (Davidson 2001, 211). Both principles are subsumed under the heading of what Davidson calls 'the principle of charity' (Davidson 2004, 35-6).

The principle of coherence concerns the internal relations between propositional attitudes. It constrains the interpreter to attribute a significant degree of internal consistency to the interpreted. The relevant consistency comes in two forms. First, there must be basic logical consistency between most propositional attitudes in order for these attitudes to have determinate content. Davidson claims that mental states get their content partly from their relations with each other. Thus, in order for Andy to think that Harry is a bachelor, he must also be thinking that Harry is a man, that Harry is not married, and so on. According to Davidson, the basic norms of consistency are those formulated in

---

[3] See e.g. Davidson 1984, 183-198; 2001. Philosophers committed to some form of metaphysical realism are likely to protest at this point. As my main interest is in what follows from Davidson's account of the mental, I shall grant these controversial implications of his account for the sake of argument.

classical predicate logic with identity (c.f. Davidson 1984, 199-214). Second, 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. Thus, in order for Sheila to propose to Harry, she must have a desire to get married or some alternative desire that instrumentally connects with a belief about the likely consequences of a proposal (and the content of which also depends on the content of her other intentional states). According to Davidson, the basic norms of instrumental consistency are those formulated in subjective decision theory as developed by Ramsey, Jeffrey and others (c.f. Davidson 2004, 151-166).

Coherence is a matter of degree. Determinate propositional attitudes can be attributed to agents who are less than perfectly coherent, but only within limits. The lower the degree of coherence among an agent's attitudes, the less determinate it is what they think. According to Davidson, an agent's failure to comply with the principle of coherence is a *normative* failure of (theoretical or practical) rationality.[4]

Whilst the principle of coherence ensures that the propositional attitudes of agents form a relatively coherent set, the content of this set is underdetermined by knowledge of an agent's behaviour and its perceived causes. For any set of behavioural evidence there is an indefinite number of incompatible sets of contents that fit the behavioural evidence equally well. Some of these are quite bizarre, unnatural, or simply unbelievable (c.f. Lewis 1984, 107-8). The principle of correspondence rules out the bizarre, unnatural, and simply unbelievable contents by constraining the attribution of propositional attitudes by

---

[4] For a contrasting view, see e.g. Mellor 2005.

their external causes as perceived by the interpreter. 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 interpreted is responding to one feature of the world rather than another (Davidson 2001, 212-3). Davidson calls this process of external correlation 'triangulating'.[5] Triangulation allows the interpreter to think of the interpreted as responding to the same features of the world that the interpreter would respond to in similar circumstances.

Davidson claims that the principle of charity constrains the attribution of two kinds of mental state. First, the interpreted must be regarded as generally forming beliefs with the same contents as those of the interpreter when confronted with similar features of the world. The resulting commonality of belief entails that the interpreter must regard the interpreted as having generally true beliefs. For purposes of interpretation, beliefs are 'intrinsically veridical' (Davidson 2001, 155).

Second, the interpreted must be regarded as generally forming desires with the same contents as those of the interpreter when confronted with similar features of the world. According to Davidson, the resulting commonality of desire entails that the interpreter must regard the interpreted as generally sharing her desires or values.[6] For purposes of interpretation, desires or values are also intrinsically veridical (or shared).

---

[5] By 'feature of the world' I shall interpret Davidson to mean 'event, object, or state of affairs, actual or possible'. It might be considered a potential problem for Davidson's account that agents do not generally interpret their own mental states on the basis of observing their own behaviour. For Davidson's attempt to reconcile first person authority with his externalist account of mental content, see Davidson 2001, 3-14.

[6] In a number of places, Davidson appears to use the terms 'value' and 'desire' interchangeably (c.f. Davidson 2004, 36, 73). In other places, he does not. Thus, according to Davidson, 'evaluations' come in many kinds, including 'wanting, desiring, cherishing, holding to be right or to be obligatory, and the negative and comparative versions of these' (Davidson 2004, 20; c.f. Davidson 1980, 4). This latter use most strongly suggests a reading on which the evaluative attitudes form a wider category of which desires are an instance (c.f. Davidson's talk of 'pro attitudes' e.g. in Davidson 1980, 3-19). In what follows, I

According to Davidson, the applicability of the principle of charity is a necessary condition for the existence of genuine disagreement between agents. It is only against the background of a shared set of beliefs that it makes sense to say that two agents employ a common stock of concepts in terms of which they can disagree about some determinate feature of the world. Furthermore, it is only against the background of a shared set of criteria for the application of the relevant concepts that such disagreements can be resolved. It follows that there is a substantial limit on the degree to which the beliefs of agents can coherently conflict.[7]

Applying the principle of charity to the domain of desire, Davidson draws a parallel conclusion as he does for conflicts of belief. It is only against the background of a shared set of desires or values that it makes any sense to say that two agents employ a common stock of concepts in terms of which they disagree about the desirability of some feature of the world. Furthermore, it is only against the background of a shared set of criteria for the application of the relevant concepts that such disagreements can be resolved. It follows that there is a substantial limit on the degree to which the desires or values of agents can coherently conflict.

Davidson claims that we should expect what he calls the 'enlightened' values of rational agents to converge (Davidson 2004, 49). To the extent that they must, value judgements 'are true or false in much the same way our factual judgements are' (Davidson 2004, 49). What explains that a judgement is true or false is therefore at least partly the same in either case, namely that the same features of the world cause similar responses in different agents at different times. Where convergence fails to occur we are

_____

employ the disjunctive terminology of 'desires or values' in order to remain as neutral as possible on this interpretive issue.

[7] One standard statement of Davidson's position on disagreements of belief is Davidson 1984, 183-198.

entitled to assume that there is an explanation for this failure consistent with the claim that some of these responses are correct and others not.

Davidson concedes that the analogy between belief and desire is imperfect in two respects. First, while paradigm beliefs are non-controversial truth-bearers, paradigm desires are not. Davidson's response is to claim that the positive attitude taken by an agent towards the object of a desire can be made linguistically explicit by formulating the desire as an evaluative judgement that carries its truth-evaluability on the surface (Davidson 2004, 37). According to Davidson, the fact that an agent who desires something desires it on the basis of some of its perceived features entails that '… all pro-attitudes may be expressed by value judgements that are… implicit' (Davidson 1980, 86).[8] On this view, desire is what Scanlon and others have called a 'judgement-sensitive' attitude (see e.g. Scanlon 1998, 18ff). It follows that there can be no conceptual analysis of value in terms of desire reductively understood, contrary to what has been proposed by some contemporary naturalists about ethics and value (c.f. Blackburn 1998; Lewis 1989).[9]

---

[8] P. Hurley interprets Davidson as holding that the content of a desire that P is given by the evaluative judgement that it is *desirable* that P (Hurley 2002, 6-15). What Davidson actually says is less unequivocal: 'The simplest view would be… to identify desiring a sentence to be true with judging that it would be desirable if it were true… And it is… hard to see how these two attitudes can be allowed to take entirely different directions. On the other hand… there is no simple… connection between our basic preferences for the truth of various sentences and our judgements about the moral or other values that would be realized if they were true. The full story here is… very complicated, and touches on many much disputed matters' (Davidson 2004, 37). Some of these complications are noted in Myers 2004, 107, Note 2.

[9] Davidson's account would be problematic if it could be shown to entail that all desires have objective truth conditions regardless of their content or their conditions of possession. Thus, it is natural to think that at least some intelligible tastes (such as a preference for ice cream over pickled eels) are rationally arbitrary. S. Hurley may be on to this when she writes: 'Evaluative constraints on eligibility operate at the level of *pro tanto* intelligibility, not correctness… The evaluative assumptions about intelligibility… are not of controversial solutions to ethical problems in which fundamental values conflict, but rather assumptions that it is correct to describe these problems in those value-laden terms in the first place…' (Hurley 1989, 105).

Second, while paradigm beliefs are caused by objectively existing features of the world that give them their content, paradigm desires do not on Davidson's view have distinctively evaluative ontological correlates to provide evaluative contents for them. Davidson's response is to separate the 'epistemological' problem of objectivity from the 'ontological' problem of existence (Davidson 2004, 44; c.f. Davidson 1984, 199-214). On this view, the correct application of concepts as given by their shared criteria of application is sufficient for truth, and thereby also for objectivity (Davidson 2004, 56, Note 12). To pick out the (real) values, all that is needed is to pick out those (non-evaluative) features of the world that are correctly valued. Once we have solved the epistemological problem involved in doing so, we should lose our interest in the ontological issues that have traditionally been associated with the objectivity of value. According to Davidson, these questions are 'infected by the inherently unintelligible question where values are' (Davidson 2004, 43).[10]

Davidson's argument has three noteworthy implications for ethics and the philosophy of value. First, there is agreement in value among enlightened agents as a matter of *necessity*. Second, the question of the objectivity of value is not best interpreted as a question of *ontology*, but rather as a question of *convergence* in evaluative judgement. Third, the truth of Davidson's account would go some way to explain the methodological centrality of pre-theoretical *intuitions* in evaluative thought. For if rational agents value the same things as a matter of necessity, and if what they value is by

---

[10] While Davidson is quite clear that some beliefs do have substantial ontological correlates, he is less than perfectly clear about what distinguishes those beliefs that do from those beliefs (and desires) that do not, and the extent to which this distinction is itself clear, sharp, or objective. Analogous problem cases here are abstract objects like numbers, properties, propositions, logical relations and the like. For Davidson on abstract objects, see e.g. Davidson 2001, 53-4. Susan Hurley is less nervous than Davidson about ontologizing talk of objective values, being happy to describe the 'value-ladenness' of desire as 'a species of the world-ladenness of the mind in general' (Hurley 1989, 55-57).

and large bound to be correct, then it is hardly a surprise that pre-theoretical intuitions should be regarded as evidential in the course of ethical and other evaluative inquiry.

3. Objectivity and convergence

On my reading of Davidson's argument, the scope and limits of objectivity depend on the extent to which the values of rational agents would converge, at least in favourable circumstances. Thus, Davidson writes:

> '… I want to… suggest that we should expect enlightened values – the reasons we would have for valuing and acting if we had all the (non-evaluative) facts straight – to converge; we should expect people who are enlightened *and fully understand one another* to agree on their basic values. An appreciation of what makes for such convergence… also shows that value judgements are true or false in much the way our factual judgements are.' (Davidson 2004, 49)

In order to evaluate this claim, we need to be clear about what is meant by a) basic values, b) enlightened values, and c) converging values.

a) *basic values*. There is more than one way to understand the notion of a basic value. First, it can be understood as a foundational value, as opposed to a derived value. Alternatively, it can be understood as a general, as opposed to a particular value. Third, it can be understood as a large subset of values, as opposed to a small subset of values. Fourth, it can be understood as a core, as opposed to a peripheral value. Whilst perhaps not contradicting anything Davidson actually says, the first and second understanding are

in potential tension with Davidson's holism and his coherentist remarks about knowledge and justification (see e.g. Davidson 2001, 137-57). The third understanding is purely quantitative, and says little about the relative importance of different values. It is the fourth understanding that appears to make best sense of Davidson's talk of holistic networks of values and norms, none of which may be analytically immune from revision, but some of which are more central to interpretation than others.[11] It is in accordance with this fourth understanding that I shall interpret Davidson's talk of basic values in what follows.

b) *enlightened values*. Davidson thinks of enlightened values as values given by desires informed by the facts, where the notion of a fact is understood non-evaluatively. It is not obvious what this category includes, given that on Davidson's view all mental concepts are normatively laden. On one natural understanding of non-evaluative facts, mental facts are therefore excluded. The problem with this interpretation is that it is impossible to even ask whether the enlightened values of agents would converge without being able to talk about propositional attitudes like beliefs, desires, or intentions. On a second understanding, non-evaluative facts would include mental facts, perhaps because mental descriptions are true of mental events that are token-identical with physical events, and as such enter into genuine causal relations in a way that values possibly do not (c.f. Davidson 1980, 207-227; Davidson 2001, 72). While this reading may not be faithful to everything Davidson says about the mental, it is the reading of Davidson's view I shall adopt in what follows.

---

[11] Davidson writes: 'The common framework is the area of overlap, of norms one person correctly interprets another as sharing… [T]he 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).

c) *converging values*. There are at least two ways in which desires can be said to converge. First, Harry and Andy could both want pickled eels *for themselves*. Second, Harry and Andy could both want pickled eels for *Harry* or *Andy* (or *both*).[12] Yet as Hume famously points out, if the 'internal frame or external situation' of agents are different, it is hardly plausible that they will all want all the same things for themselves.[13] Harry's sweet tooth could make him prefer ice cream to pickled eels while Andy would rather have the eels in all their sliminess. Even if Harry and Andy differ in what they want for themselves, it does not follow that they are in any kind of conflict about what either should want, even given their differing internal frame and external situation. In what follows, I shall interpret Davidson as holding that while there must be *some* convergence in desire across agents in both senses, there need not be complete convergence in desire across agents in the first sense. Nor shall I interpret Davidson as denying that the appropriateness of desires can relative, either to 'time, place, person, culture, tribe, or legal system' (Davidson 2004, 40).[14]

Davidson mentions at least two conditions in which disagreement in value may occur. First, the scope of disagreement in value increases in proportion to the distance of the values in question from basic evaluative commitments the attribution of which is necessary to make minimally coherent sense of agents as acting on reasons at all. Davidson writes that questions of fairness, justice, and social welfare are in this sense remote from the basic values that constitute agency, and that they are therefore predictably subject to *de facto* lack of convergence (Davidson 2004, 74). These

---

[12] For a discussion of different modes of self-identifying desire, see e.g. Lewis 1984, 27-50.

[13] As Hume also points out, the same point applies to the same agent at different times (c.f. Hume 1993, 149-152).

[14] For a discussion of how Davidson can accommodate the existence of so-called 'agent relative' values, see Myers 2004, 123-24.

comparatively substantial moral values contrast sharply with the basic laws of logic, as well as what Davidson calls 'the principle of total evidence for inductive reasoning, or the analogous principle of continence' (Davidson 2004, 195). According to Davidson, all rationally minded creatures are necessarily committed to these (relatively) formal values, it being impossible to coherently interpret them as thinking rational agents otherwise.

Second, Davidson suggests that the scope for disagreement in value is sensitive to the differential specificity and theoretical role that different values play in our understanding of the world and each other. In one place, he suggests that disagreement on general evaluative principles is less intelligible than disagreement about the value of particular acts or events (Davidson 2004, 72). Elsewhere, he suggests that theoretical disagreement is in some cases more intelligible than disagreement about what he describes as 'more evident', including (most probably) the objects of basic (or revealed) preference and choice (Davidson 1984, 169).[15] Whether, and if so how, these claims are mutually compatible is a question I shall not pursue here (c.f. Davidson 2004, 87ff).

Given that genuine interpretability is consistent with *de facto* lack of convergence in values, why should we think the enlightened values of agents would converge? There are at least two options here. The first is that Davidson's convergence claim is merely a restatement of the thesis that intelligible disagreement about values presupposes the existence of *some* agreement on basic values. There is some textual evidence for this claim. Thus, Davidson writes:

---

[15] In at least one place, Davidson talks of an agent's preferring one sentence to be true rather than another as the 'raw data' of interpretation (Davidson 2004, 30). It is natural to think of at least some such preferences as being directed at objects more or less directly revealed in the agent's environment (c.f. Davidson 2004, 151-166).

'… [N]o matter what the subtleties involved, interpreting evaluative judgements rests on the same foundation as interpreting the evaluative attitudes: understanding depends on finding common ground. Given enough common ground, we can understand and explain differences, we can criticize, compare, and persuade. The main thing is that finding the common ground is not subsequent to understanding, but a condition of it… If we understand their words, a common ground exists, we already share their way of life' (Davidson 2004, 51).

If this is all there is to it, Davidson's objectivism about value is less powerful and interesting than some of its formulations may suggest. Basic intelligibility leaves open the scope for intractable disagreement about the application of a vast range of substantial evaluative concepts, including rightness, obligation, fairness, justice, social welfare, ice-cream, pickled eels, and many of the other values that concern people the most. The fact that communists, socialists, liberals, libertarians, and the Taliban can all be said to believe in justice is likely to be cold comfort for anyone concerned unless there is some prospect that the conflict between their mutually incompatible interpretations of this allegedly common value could be reasonably resolved, at least in principle.

The second option is that enlightened values would converge beyond the range of basic values. There is also some textual evidence for this claim. Thus, Davidson writes:

'… [I]f I am right, disputes over values (as in the case of other disputes) can be genuine only when there are shared criteria in the light of which there is an answer to the question who is right… When we find a difference inexplicable, that is, not

due to ignorance or confusion, the difference is not genuine… The importance of a background of shared beliefs and values is that such a background allows us to make sense of the idea of a common standard of right and wrong, true and false.' (Davidson 2004, 50-51).

I shall refer to this argument as 'the argument from criteria'. I will interpret the argument as follows. Attributing value judgements entails attributing value concepts. Value concepts are identified by the associated commitments that define their place in a wider network of concepts. These commitments fix the application criteria of value concepts. These application criteria are subscribed to by anyone who competently applies the value concepts. Thus, if two agents appear to disagree in value judgement, then either at least one of them is mistaken about whether the relevant criteria obtain or they are not applying the same concept. If two conflicting agents are applying the same concept, then at least one of them is mistaken about whether the criteria of application for the concept obtain. To be mistaken about whether the criteria of application for a concept obtain is to be ignorant of some (non-evaluative) fact. It follows that the value judgements of (non-evaluatively) enlightened agents will converge.[16]

Thus interpreted, the argument from criteria fails to take account of at least three sources of indeterminacy that could arise for the application of evaluative concepts. Each source of indeterminacy is compatible with competent mastery of the relevant concepts. One source of indeterminacy concerns the application of agreed criteria to cases. The two

---

[16] As formulated here, Davidson's argument from criteria makes no reference to the principle of charity. It is therefore independent of his interpretationist account of the mental (c.f. Davidson 2004, 55). Pettit 2002 propounds a sophisticated variant of the argument from criteria in connection with a discussion of the possibility of rule following in general. Pettit's argument is vulnerable to some of the same problems of indeterminacy that I raise for Davidson's argument below.

others concern the status of criteria themselves. Each source of indeterminacy arises from facts about evaluative thought that are well known and not particularly controversial.

First, the application of shared criteria for evaluative concepts is not as straightforward as some of Davidson's remarks may suggest. This is not only because it can be difficult to establish whether the relevant (non-evaluative) facts obtain. Such difficulties clearly apply to all concepts, evaluative or not. It is also because (as Davidson must agree) the criteria associated with evaluative concepts can themselves be evaluative. Thus, the agreed criteria for whether it is right to give someone a compliment may include whether or not it will do him any good. To establish whether this criterion obtains it may not be sufficient to be fully informed about the non-evaluative facts.[17] Furthermore, and consistently with Davidson's argument, some evaluative conflicts could be subject to trade-offs with more than one coherent resolution. In such cases, the inference from commonality of concept and genuine disagreement to the non-evaluative ignorance of at least one party is unsound.

Second (and as Davidson explicitly agrees) the scope of convergence is limited by indeterminacy arising from the fact that concepts are sometimes extended to cases not previously accommodated by their existing criteria of application. Given the diversity in internal frame or external situation across different agents at different times, there could be more than one coherent and reasonable extension of a concept in such 'difficult or unusual' scenarios (Davidson 2004, 51). If so, the question of how a concept should be extended may have no determinate answer. To the extent that the distinction between

---

[17] Perhaps partly for this reason, value concepts are sometimes labelled 'essentially contested'. See e.g. Gallie 1955/6, 180-95; Hurley 1989, 46-50. My argument is compatible with the essential contestability of evaluative concepts in the sense originally introduced by Gallie, although it does not entail it. See e.g. Gallie 1955/6, 169.

invention and discovery blurs at this point, so does the distinction between 'subjective' and 'objective'.

Not all difficult scenarios are unusual. Questions of justice and fairness are also difficult and contested, yet people are constantly forced to make their minds up about them by the need to make choices, for example when discussing the problem of abortion, issues of resource allocation, or income tax. If these problems are placed at some distance from the core of the basic values about which there must be convergence among enlightened rational agents, it cannot be for the reason that they present competent speakers with unusual cases for which criteria have yet to be proposed. They most certainly have. Yet the proposed criteria remain deeply contested, even among rational and intelligent agents.

The potential for indeterminacy arises also for a third reason. The issue here is that the criteria of application for value concepts are themselves often in dispute between competent speakers. On Davidson's view, concepts are partly defined by a network of commitments that allow speakers to distinguish them from each other. This network is neither absolutely fixed nor an all-or-nothing matter. While some commitments are more central to the identity of a concept than others (like the norm of instrumental rationality is central to the concept of rational agency), the notion of centrality is one of degree. Two agents could in principle associate non-identical sets of commitments with a disputed evaluative concept and remain coherently describable as thinking about the same thing. If they do, they will associate different criteria of application with that concept. In some cases (although not all) they will apply the concept in incompatible ways. Where there is diversity in internal frame or external situation across different agents and times, such

conflicts are possible even where the agents in question are neither obviously unreasonable nor misinformed about the (non-evaluative) facts. Where such scenarios obtain, there will be no determinate answer to the question of what to do in the relevant case.[18]

It would be a mistake to think that the preceding point relies on a controversial conception of concepts as delimited by vague and indeterminate boundaries. A parallel conclusion would follow even if the boundaries of evaluative concepts were assumed to be sharp. If two agents associate subtly different application criteria for the term 'justice', for example, it may be granted that they operate with different concepts. If so, they can agree on all non-evaluative facts, apply the term 'justice' differently, and to this extent not be in genuine disagreement about what is just. This does not dissolve the problem. Wherever the two concepts have incompatible action-guiding consequences a question could arise about which of the two concepts, if any, they are to be practically guided by. Even if the criteria of application themselves are not formulated in evaluative terms, the criteria for their selection could be. To settle on criteria of application for evaluative concepts is partly an evaluative matter. Once more, where there is diversity in internal frame and external situation across agents and times, there could be more than one coherent resolution of the relevant conflict by agents who are neither obviously unreasonable nor misinformed about the (non-evaluative) facts. There would then be no determinate answer to the question of what to do in the relevant case.

On Davidson's account, the scope for indeterminacy left open by the argument from criteria is obviously limited by the principle of charity. It would be impossible for

---

[18] There is no quick solution in appealing to a set of shared higher order concepts for the resolution of disputes about the application criteria of lower order concepts. Any higher order concept would itself be subject to the same problems of indeterminacy as the lower order concepts.

agents to disagree in all their evaluative claims at the same time while remaining able to communicate. Furthermore, there are arguably some criteria of application a commitment to which is necessary for being able to agree or disagree with anyone about anything. In particular, the convergence claim is most plausible when applied to the basic norms or values that fall under Davidson's principle of coherence. If enlightened convergence is a mark of objectivity, then any comparatively formal values picked out by the principle of coherence are objective. In this sense, if in no other, the enlightened values of agents would converge.

It does not follow that any individual substantial evaluative claim is immune from indeterminacy so long as there is compensating agreement in substantial value elsewhere. It therefore does not follow that there is a uniquely fixed and determinate set of particular features of the world the positive or negative evaluation of which all agents must share if they understand each other and are otherwise well informed about the (non-evaluative) facts.[19] True, some candidates have been proposed for the title of *a priori* individually non-intelligible objects of desire, including the intentions of Anscombe's infamous Nazi and her saucer-of-mud collector (Anscombe 1957, 73ff). The problem with individual examples like these is that Davidson's account is arguably consistent with the existence of such particular desires, provided they occur against a background of a basically coherent system of attitudes that can be represented as generally responding to the world as we take it to be. In the absence of further specification of external and internal

---

[19] As previously noted, Davidson does not regard the claim that the comparably formal values of consistency and coherence are objective as licensing the claim that they are ontologically substantial features of the world (C.f. Notes 5 and 10 above).

(attitudinal) background, it is not obvious that either Anscombe's Nazi or her saucer-of-mud collector would necessarily fail this test.[20]

Davidson's account is consistent with an indefinitely wide range of questions about substantial value having no determinate answer. There is ample textual support for this hypothesis in Davidson's discussion of the issue. First, on Davidson's account, the normative status of individual evaluative claims is partly a function of their place in a network of other norms and commitments. For example, Davidson claims that the property of rationality is a purely relational feature of beliefs and desires. He writes:

> Strictly speaking… irrationality consists not in any particular belief but in inconsistency within a set of beliefs… I think we must say much the same about intentions, intentional actions, and other propositional attitudes… They are never irrational in themselves, but only as part of a larger pattern.' (Davidson 2004, 192)

A parallel claim arguably applies to any feature of the world that is judged to be valuable. If no particular features of the world are valuable in themselves, but only as valued by a larger pattern of attitudes, then no particular features of the world are determinately valuable unless they are valued by some suitably unique or otherwise privileged pattern of attitudes.[21] Even when supported by the principle of charity, the argument from criteria

---

[20] For discussion of how apparently worthless and absurd desires could be embedded in a network of generally rational attitudes, see e.g. Blackburn 1998, 59-68; Raz 1999, 22-45; Myers 2004, 118.

[21] The idea that the 'real' values are those picked out by what might be described as the 'most coherent' pattern of desire is one that, while it may be a natural extension of Davidson's argument, is not entailed by anything Davidson says. One possible move in this direction would be a so-called 'coherentist' account of practical reasoning with a convergence claim made about enlightened values held compatibly with sound practical reasoning (c.f. Myers 2004, 134). While there is little textual evidence that Davidson is concerned to defend any such account, invoking one would considerably raise the stakes in the debate over the scope and limits of the convergence claim. While the notion of a 'most coherent' network of attitudes is one that

fails to clearly establish that any uniquely fixed and determinate set of particular features

of the world is so valued. To this extent, the argument fails to clearly establish that any

unique and determinate set of particular features of the world is objectively valuable. [22]

Second, and as already noted, the principle of charity must be consistent with

such genuine disagreement about ethics and value as there actually is. There is evidence

to suggest that some such disagreements, including disagreements on matters of taste or

on abstract moral issues involving justice and fairness, do obtain between rational agents

who are in broad agreement about the non-evaluative facts. It follows that there is no

straightforward route from Davidson's interpretationist account of the mental to the claim

that the enlightened values of rational agents would converge. This modest conclusion

receives additional support from Davidson's own remarks about evaluative disagreement.

Davidson explicitly notes that there can be 'real differences in norms among those who

understand each other', as long as the differences are placed 'within a common

framework' (Davidson 2004, 50). He also concedes that we can 'make a fairly clear

distinction between interpersonal comparisons and the normative judgements based on

them', and that conceptions of fairness, justice, and social welfare have no 'favoured

role' in our attributions of propositional attitudes (Davidson 2004, 74). When making his

convergence claim, Davidson often talks of 'basic' (Davidson 2004, 49, 51), 'leading'

(Davidson 2004, 36), or 'strategically important' values (Davidson 2004, 73).[23] When

---

has been a subject of much philosophical controversy, nothing I say in this paper amounts to a proof that
such a strategy is destined to failure.

[22] Susan Hurley shows some awareness of this issue when she writes that the question of there being
alternatives that ought to be done 'all things considered' comes down to the question of there being a
'theory that does the best job of displaying coherence' (Hurley 1989, 193). The identification of such a
theory is '… left as a substantive issue, to be determined by theoretical deliberation' (Hurley, 1989, 193).
Hurley does not claim to have found a unique theory with this property.

[23] Yet Davidson also appears to say that the convergence claim applies to 'moral' values (Davidson 2004,
47). And the list of values given in the context where he presents the argument for criteria includes 'good,

making room for *de facto* disagreement by distinguishing between central and peripheral values, he illustrates the former by mentioning the paradigmatically formal values of consistency and coherence. Taken together, this evidence goes some way to suggest that a great number of evaluative concepts are too distant from the basic values necessarily invoked as shared in interpretation to imply the existence of a determinate solution to the evaluative disagreements they may involve. This is a significant concession to the subjectivist about value. If objectivity is understood partly in terms of convergence, and if only comparably basic values are ones on which enlightened agents would converge, then Davidson's argument fails to support the objectivity of an indefinitely wide range of non-basic substantial values. Of course, the fact that some substantial evaluative claims are indeterminate in truth-value does not entail that all substantial evaluative claims are. Nevertheless, a restriction of the convergence claim to a set of comparatively basic values required to make sense of rational agents as minimally intelligible amounts to a significant qualification of Davidson's claim that value judgements are objective in much the same way as paradigm factual judgements are. This is not because no factual judgements have ever been thought to have indeterminate truth-values. Some clearly have. It is rather because the range of determinate and ontologically substantial fact is naturally thought to extend beyond the basic range entailed by basic mutual intelligibility. If one consequence of Davidson's argument is that this natural thought is profoundly mistaken, it is a moot point whether the parity between fact and value thus established is better captured by denying the subjectivity of value, the objectivity of fact, or the very

---

bad, moral, blameworthy, right, courageous, courteous, trustworthy, loyal, kind, obedient, cheerful, etc.' (Davidson 2004, 54). It is a notable feature of Davidson's list that it makes no discrimination between so-called 'thick' and 'thin' ethical concepts.

mode of classification on which the traditional debate between subjectivists and objectivists about ethics and value has been based.

In more than one place, Davidson explains his conception of objectivity in terms of the possession of some truth-value, these being 'true or false, or possibly neither' (Davidson 2004, 42, 56). So understood, evaluative judgements would be objective even if all evaluative claims were indeterminate or false (c.f. Mackie 1977).[24] Yet this literal reading cannot accurately represent Davidson's considered view about the objectivity of value. It is beyond reasonable doubt that Davidson thinks a significant range of evaluative claims come out as determinately true. What is not beyond reasonable doubt is the extent to which the conjunction of the principle of charity and the argument from criteria supports this claim.[25]

REFERENCES

ANSCOMBE, E. 1957, *Intention*, Oxford: Basil Blackwell.

BLACKBURN, S. 1998, *Ruling Passions*, Oxford: Oxford University Press.

DAVIDSON, D. 1980, *Essays on Actions and Events*, Oxford: Oxford University Press.

DAVIDSON, D. 1984, *Inquiries into Truth and Interpretation*, Oxford: Oxford University

Press.

DAVIDSON, D. 2001, *Subjective, Intersubjective, Objective*, Oxford: Oxford University Press.

DAVIDSON, D. 2004, *Problems of Rationality*, Oxford: Oxford University Press.

---

[24] In one place, Davidson writes: 'Hume, Bentham, Hare, Blackburn, and a host of others have emphasized the… subjective, emotive, or projective character of value judgements. They are surely right' (Davidson 2004, 57). In light of the foregoing remarks, this intriguing claim is perhaps not surprising.

[25] The writing of this paper has been greatly helped by comments from anonymous referees. Parts of the material included in the paper have also been presented at the universities of Edinburgh, Stirling, and York. I am grateful to the audiences there, as well as to Simon Blackburn and the students in the Metaethics Seminar at Cambridge for useful discussions of this topic.

FRANKFURT, H. 1971, ''Freedom of the Will and the Concept of a Person', *Journal of*

    Philosophy 68, 5-20.

GALLIE, W. B. 1955/6, 'Essentially Contestable Concepts', *Proceedings of the Aristotelian*

    Society 56, 167-198.

HUME, D. 1993, *Selected Essays*, Oxford: Oxford University Press.

HURLEY, P. 2002, 'A Davidsonian Reconciliation of Internalism, Objectivity, and the Belief-

    Desire Theory', *The Journal of Ethics* 6, 1-20.

HURLEY, S. 1989, *Natural Reasons*, Oxford: Oxford University Press.

LEWIS, D. 1984, *On the Plurality of Worlds*, Oxford: Blackwell.

LEWIS, D. 1989, 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*,

    Supplementary Volume 63, 113-137.

LEWIS, D. 1996, 'Desire as Belief II', *Mind* 105, 303-314.

MACKIE, J. L. 1977, *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.

MELLOR, D. H. 2005, 'What Does Subjective Decision Theory Tell Us?', in H. Lillehammer

    and D. H. Mellor (eds.), *Ramsey's Legacy*, Oxford: Oxford University Press, 137-148.

MYERS, R. H. 2004, 'Finding Value in Davidson', *Canadian Journal of Philosophy* 34, 107

    136.

PETTIT, P. 2002, *Rules, Reasons, and Norms*, Oxford: Oxford University Press

RAZ, J. 1999, *Engaging Reason*, Oxford: Oxford University Press.

SCANLON, T. M. 1998, *What We Owe Each Other*, Cambridge: Harvard University Press.