
*Problems of Rationality* is the fourth collection of essays by Donald Davidson that has recently been published (or republished) by Clarendon, and the first to be published posthumously.\(^1\) He passed away on August 30, 2003, at the age of 86. And although he had already made much of the preparations for the volume, it is his long time companion Marcia Cavell who put the pieces together. It is also she who has authored the Introduction.

The essays in this volume focus on action, value and rationality. One of them is the paper ‘A Unified Theory of Thought, Meaning and Action’ (originally published as ‘Toward a Unified Theory of Meaning and Action’, in which Davidson stresses that interpretation of someone’s speech not only requires that we uncover the contents of her beliefs but also the contents of her desires. None of these factors–meaning, belief and desires–may, according to Davidson, be determined without simultaneously determining the others. Consequently, interpretation and explanation of action are parts of the same project.

Other of the essays focus on the problem of irrationality. In accordance with his ‘principle of charity’, Davidson has stressed that we interpret an agent correctly only if we represent him as satisfying certain norms of rationality. For example, the principle entails that a competent interpreter finds that the speaker has beliefs that are, by and large, consistent and true

\(^{1}\) Page references are to this volume, unless otherwise indicated.
(by the interpreter’s lights), and desires and preferences that satisfy certain
ordering conditions (such as transitivity). In a congenial way, Davidson
stresses that the explanation of actions always involves an element of
rationalization; i.e., it involves showing that, ‘from the agent’s point of view
there was, when he acted, something to be said for the action.’ (‘Actions,
Reasons and Causes’, reprinted in Essays on Actions and Events, Oxford:
Clarendon, 1980, 9).

It might be wondered whether this picture leaves any room for
irrationality. That is a crucial question, since we seem to have good reason
to believe that people sometimes act irrationally (I know that I do). How
can we account for irrationality within a Davidsonian framework? In the
essay ‘Paradoxes of Irrationality’, one of the strategies Davidson pursues
invokes the idea of a ‘compartamentalization’ of the mind. According to
this idea, roughly, the mind is divided into more or less independent
subsystems after the fashion of Freudian structures. Each of the
compartments constitutes a supporting and relatively coherent structure of
reasons, beliefs, expectations, and other attitudes, and irrationality (some
cases at least) is explained as results of conflicts between them. But this
looks more like an ad hoc move to save the theory than a viable
psychological model.

I shall not argue that point, however. Instead, I shall dwell on another
theme that Davidson develops in the book (particularly in the paper ‘The
Objectivity of Values’), namely the thesis that, when applied to moral
discourse, his views on meaning and interpretation support the ‘realist’ or
‘objectivist’ claim that moral judgments may be objectively true.\textsuperscript{2} Not only that, Davidson indicates that they allow him to combine objectivism with the internalist idea that thinking, say, that an action is obligatory necessarily involves being motivated to perform it, which is a view often held to conflict with it.\textsuperscript{3} In other words, the question is: Did Davidson have an exciting novel solution to ‘the moral problem’ up his sleeve?\textsuperscript{4} Some commentators have recently made suggestions to that effect.\textsuperscript{5}

However, personally I am skeptical. Davidson takes his objectivism to entail that ‘evaluative judgments have truth values, that is, are objectively true or false’ (56), and adds that in this respect there is no difference

\textsuperscript{2} Davidson himself does not like the term ‘realism’ as a name of his position. For example, see ‘Objectivity and Practical Reason’, in E. Ullman-margalit (ed.), \textit{Reasoning Practically}, Oxford: Oxford University Press, 2000, 22 (only a part of this paper is republished in \textit{Reasoning Practically}). However, this is due to the assumption that realism entails that moral properties are concrete objects with a location in space-time. And under the ordinary understanding of ‘moral realism’ in the current meta-ethical debate, it has no such implications.

\textsuperscript{3} For Davidson’s sympathies to internalism, see 25 and 57.

\textsuperscript{4} The phrase ‘The Moral Problem’ is Michael Smith’s. Roughly, to solve the moral problem is to develop a theory that allows one to reconcile internalism with the idea that moral judgments can be true, without giving up a Humean view on psychology. See Smith’s \textit{The Moral Problem} (Oxford: Blackwell, 1994), ch. 1.

between them and ‘our ordinary judgments about the physical world’ (41).

Moreover, he claims that:

If […] we turn to the problem of understanding what it is like to judge that an act or object or institution is morally desirable or ought to exist or is obligatory, we realize that such a judgment necessarily consists in assigning some property or other to an entity or group of entities. The semantic nature of such judgments is clear; the attributor is classifying one or more things as having a certain property. (48)

In support of these claims, Davidson adduces different arguments. Some are familiar from the meta-ethical debate, such as the so-called ‘Frege-Geach Problem’. It is commonly held that evaluative judgments occur in arguments that are logically valid. This supports the claim that they have truth-values, it is held. For if we deny that claim, we have a hard time explaining what it means for such arguments to be valid.6 However, Davidson also appeals to certain considerations that are related to his own views about meaning and interpretation, and it is on these that I shall focus.

Thus, consider his well-known thesis that there is, necessarily, considerable overlap between our systems of belief. According to Davidson, this thesis applies also to evaluations, and the considerations that explain why also provide support for his objectivism:

We should expect enlightened values [...] 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 or agreement also shows that value judgments are true or false in much the way our factual judgments are. (49)

What is it that ‘makes for’ the convergence? Davidson’s argument for the convergence thesis invokes the principle of charity. Davidson thinks that knowledge about the meanings of a speaker’s sentences can be captured by a ‘theory of interpretation’ for that speaker. Such a theory is finitely axiomatized and entails for each sentence s of the target idiolect a ‘T-sentence’, which is a theorem that specifies the truth conditions of s by using a sentence p of our (the meta-) language:

\[ s \text{ is true in the speaker's idiolect if and only if } p. \]

The principle of charity comprises a set of constraints on such a theory. It entails that the theory is correct only if it ascribes truth conditions to the speaker’s sentences such that the sentences she holds to be true, by and large, are consistent and true (by our lights). Or, in other words, it entails that a theory is correct only if it represents the speaker as having beliefs that, by and large, square with our own.

Of course, the principle of charity does not exclude disagreement or attribution of error altogether. For example, Davidson has emphasized that errors may have to be attributed in order to avoid making the theory too complex (see, e.g., Inquiries into Truth and Interpretation, 196). Suppose that we believe that p, and that a speaker denies s. If a theory of
interpretation for the speaker entails that s is true if and only if p, then it represents the speaker as disagreeing with us over the truth of p. This need not mean that the theory is incorrect. But Davidson suggests in numerous passages that in order for it to be correct it must satisfy two, related, constraints. First, it must represent us as agreeing about many other issues, and, in particular, as, by and large, agreeing over the considerations that are relevant in settling our dispute about p (about the standards relevant to resolving the dispute). Second, the error it attributes to the speaker (error, that is, as seen from our perspective) must be explicable with reference to some (independently specifiable) cognitive shortcoming on behalf of the speaker, such as, e.g., ignorance of relevant evidence or suboptimal perceptual conditions (see, e.g., 47, 50-1). If these conditions are not satisfied we should revise the theory and assume that s has a different truth condition.

The problem is that if we apply these views to moral sentences, we get an implausibly restrictive view about when to attribute moral convictions to other thinkers. At least, this is so if we combine this view with the internalist idea that Davidson is also an advocate of.

Davidson’s internalism entails that a theory of interpretation that represents a speaker as thinking that actions are right, wrong, obligatory etc is correct only if the speaker is motivated to perform the actions she, according to the theory, thinks are right and obligatory, and motivated not to perform those she thinks are wrong (subject, perhaps, to the condition that the speaker is not irrational). If we combine this constraint with the implications of the principle of charity just mentioned we are led to conclude that a theory of interpretation that represents the speaker as
judging actions to be morally right or obligatory is correct only if the following two conditions are satisfied:

(1) There is a phrase of the speaker’s idiolect such that she applies it on actions on the basis of considerations that we take to support thinking actions to be obligatory or obligatory. And in so far as she applies it to an action we think is not right her verdict can be attributed to some cognitive shortcoming.

(2) The speaker has at least some motivation to perform the actions on which she is disposed to apply the phrase in question.

If these conditions are not satisfied we may conclude that no theory of interpretation that represents the speaker as judging actions to be right is correct. That is, we may conclude that she has no such convictions.7

The problem is that, in many cases, it is very likely that those conditions are not in fact not satisfied. All it takes is that the speaker’s idiolect either lacks a phrase she is disposed to apply to actions on the basis of considerations we take to support judging them to be right, or that she has no motivation to perform those actions. For example, suppose that we are hedonistic utilitarians.

Davidson stresses that if a theory of interpretation represents someone as disagreeing with us about the rightness of an action, then it is correct

7 I assume here that the judgment that an action is morally right can correctly be attributed to a speaker only if there is sentence of his idiolect that can correctly be held to be true if and only if some action is morally right. But this assumption is congenial with Davidson’s well-known claim that interpretation and belief attribution are sides of the same coin.
only if we can correctly be seen as sharing standards against which such
ascriptions are to be evaluated (take similar considerations as relevant in
determining whether an action is right). Thus, all it takes in order for no
type of interpretation to be correct that represents a speaker as judging
actions to be right is that she has no motivation to perform the actions that
satisfy the conditions we think necessary and sufficient in order for an
action to be co. And, surely, this happens all the time. For example, suppose
that we are utilitarians, and think that actions are obligatory if and only if
they maximize utility. Then we are entitled to conclude that people who
lack the motivation to perform actions that (they believe) maximize utility
lack convictions about what is obligatory.

In my view, this implication is clearly implausible. But the reason is
not merely, or even primary, that it is so counterintuitive. Rather, the reason
is that it is based on an understanding of the concept of a moral conviction
that entirely undermines its theoretical utility. The concept of a moral
conviction has a role to play in a general account of human psychology,
primarily since attributions of such convictions help to explain why people
manage to coordinate their actions in mutually beneficial ways in certain
collective action situations. If we impose too strict requirements regarding
what is to count as such a conviction, it cannot be used to capture all the
states that have this specific function. It is like having a concept of a
religion that entails that someone lacks a religion if he doesn’t believe in
Christ. For an anthropologist who wants the concept to designate a trait or
phenomenon that plays a similar explanatory role in different cultures such a narrow conception is useless.\textsuperscript{8}

In contrast to Davidson’s view on when it is legitimate to attribute moral convictions to other thinkers, I favor ‘the latitude idea’. Suppose that the speaker denies a sentence $s$ and that we are convinced that an action $A$ is right. According to the latitude idea, if a theory translates $s$ with ‘$A$ is right’ (or entails that the content of $s$ is captured by that sentence), then the theory may be correct even if the speaker’s verdict cannot be blamed on a cognitive shortcoming, and even if her denial of $s$ is (within certain broad limits) based on other (non-moral) considerations than those we take to be relevant when determining an action’s moral status. The limits are necessary to ensure that the commitments that are to be counted as convictions about what is right and wrong, etc, can have the explanatory role indicated above. For example, we must probably assume that an action is evaluated as morally right by a thinker only if she, in reaching such a verdict, is disposed to assign at least some weight to its consequences for the well-being of the affected individuals.\textsuperscript{9} However, these limits still leave room for the tolerance prescribed by the latitude idea.

What are the implications of the latitude idea in the present context? Let me mention two possibilities. On the one hand, if we want to hold on to the idea that the contents of moral sentences and moral convictions are to be explained in terms of truth conditions, and that a moral sentence of our own

\textsuperscript{8} I elaborate the reasoning indicated in this and the subsequent paragraphs in detail in my \textit{Moral Disagreement}, New York: Cambridge University Press, forthcoming.

\textsuperscript{9} For a similar view, see Michael Smith’s \textit{The Moral Problem}, Oxford: Blackwell, 1994, 40f.
language gives the content of its counterpart in another language only if they have the same truth conditions, then we must reject Davidson’s general account of how truth conditions are determined. Instead, we must adopt an account (such as, e.g., some version of the causal theory of reference) that allows for sameness of truth conditions even in the presence of wide disagreement. On the other hand, if we want to retain Davidson’s general account, then we must explain sameness of content in terms of something else than truth conditions. That is a formidable challenge for an anti-realist and a non-cognitivist, as we all know, and is really what the discussion about the Frege-Geach problem is all about. However, if I am right, Davidson’s attempt to vindicate his objectivism can at best be seen as a restatement of this old argument. In particular, the considerations he adduces in support of his thesis about agreement and convergence among our ultimate values do not add anything.

Having said that, I still highly recommend the volume. Davidson is a brilliant stylist, and is always stimulating and exciting to read, whether you agree with him or not (but note that this is the verdict of an aficionado). In fact, the publication of Reasoning Practically is particularly welcome since it contains much material that has previously been difficult to get a hold of (such as his Lindley Lecture, ‘Expressing Evaluations’). It might be argued that some of the articles just as well could have been published in one of the other volumes. But that is due to the fact that Davidson to such an extent was a systematic philosopher, whose thoughts cannot easily be sorted neatly in the traditional categories. So my advice is ultimately: Buy all the volumes!