

This is a pre-peer-reviewed version of the article 'Methodological Reflections on Hägerström's Meta-Ethics' which is to appear in Eliaeson, S., Mindus, P., and Turner, S., Axel Hägerström and Modern Social Thought, The Bardwell Press, 2014 (ISBN: 978-1-905622-43-6).

# Methodological Reflections on Hägerström's Meta-ethics

Folke Tersman

## 1. Introduction

The last volume of Anders Wedberg's impressive work *Filosofins historia (A History of Philosophy)* ends with a chapter that discusses the Uppsala school of philosophy.<sup>1</sup> Wedberg was a little worried that this would lead people to interpret him as regarding the Uppsala school as a sort of *climax*, as being the most superior expression of the Western philosophical tradition that so far had emerged. However, if one actually *reads* the chapter, one is effectively cured from the temptation to make that interpretation. Wedberg was *in general* a rather grumpy person. But when it came to Phalén, Oxenstierna, Hägerström and the others, he doesn't appear to have felt any inhibitions at all. Words like 'obscure' and 'bizarre' are tossed around frequently. In one section of the chapter, Wedberg addresses Hägerström's meta-ethical ideas, his so-called 'value nihilism'. Like others, he takes Hägerström to offer a theory about language, or more specifically about certain sentences, namely sentences of the following kind:

- 'One should resist one's jealousy'
- 'Its our duty to help people in need'
- 'Pettersson is a good man'
- 'Its better to be robbed than to rob'

Such sentences are called 'value sentences', and Wedberg attributes to Hägerström the following views, where the first is supposed to provide an argument for the second while the second is in turn supposed to be an argument for the third:

- A. Utterances/assertions of value sentences are expressions of emotions.
- B. They are not expressions of beliefs (states that are capable of being true and false).
- C. They have no truth-values (cannot be true/false).

Regarding these claims, Wedberg raises certain methodological questions. For example, what would it take for them to be incorrect? Suppose it was found—Wedberg imagines—that on a single, unique, occasion, when a value sentence was gently whispered, it was not used to express an emotion. Would that be enough to refute Hägerström's theory? Wedberg figures that that cannot be right. Instead, he suggests that Hägerström offers a kind of stipulation. That is, he takes him to reason somewhat as follows: Since value sentences are on many occasions used to express emotions rather than beliefs, it is suitable to adopt a notion of a *genuine* utterance of such a sentence that generates the conclusion that A, B and C, when restricted to genuine utterances become true by definition. He says little, however, about what purposes such a definition is supposed to serve and thus little about the grounds for assessing its suitability.

What *sort* of a theory is Hägerström's value nihilism? That is the question Wedberg asks, and my aim today is simply to pursue this discussion a little further. Claims similar to A, B and C are often used to express the relevant positions also in contemporary meta-ethics. To

---

<sup>1</sup> There is an English edition of this work which is published by Clarendon Press, Oxford, in 1984. However, unfortunately, the English version of the third volume does not include the chapter about Hägerström. The chapter is mentioned, however, in a review of the Swedish edition by C.D. Broad which was published in *Philosophical Quarterly* (vol 18, no 72, 1968, 269-71).

see what is at stake in these debates it is crucial, in my view, to consider questions of the kind Wedberg raises. What I am interested in, more specifically, is to find a subject matter that the participants of these debates can plausibly be seen as disagreeing about. This is less easy than it might seem, at least if one wants to define that subject matter in such a way that it is possible to resolve the debates.

Before I go on, however, I should make one further comment on Wedberg's treatment of Hägerström. As Bo Pettersson explains in his paper, Hägerström had a more complex view than the one Wedberg attributes to him. For Hägerström seems to have held that for example C only applies to *some* value sentences. In the case of others, he seems to have thought that they do have truth-values, although he also stressed that, given the content of those sentences, they are all false (at least in so far as they actually predicate a value term to the items being evaluated). Such a composite position raises the same methodological worries as the more simplistic one, however.

## 2. Cognitivism vs Expressivism

Hägerström was an early participant in the still intense debate between *cognitivists* and *expressivists*. On a traditional account of those positions, cognitivists stress that value sentences are used to express beliefs, while expressivists deny this and insist, like Hägerström, that they rather express conative attitudes. Expressivists disagree about the *types* of conative attitudes that value sentences are supposed to express. Some agree with Hägerström in thinking that they express emotional attitudes, while others stress that they rather express desires of some sort. In any case, given the traditional account, the pertinent positions are characterized as follows.

### *Expressivism*

(E1) Value sentences (are used to) express conative attitudes (desires, sentiments, etc).<sup>2</sup>

(E2) Value sentences do not (are not used to) express beliefs (states that are capable of being true and false)

### *Cognitivism*

(C1) Value sentences (are used to) express beliefs.

(C2) Value sentences have truth-values.

These claims are usually supposed to be related in various ways. For example, (C1) is sometimes supposed to follow from (C2) and is sometimes supposed to imply (C2),<sup>3</sup> and both (C1) and (C2) are in turn related to other views associated with cognitivism, such as the thesis that predicates such as 'right', 'wrong', etc, refer to 'real properties'.

These formulations raise many questions. One set of questions concern their *scope*. Are they supposed to hold for all speakers, languages, and periods, or only about some? Another set concerns the term 'express'. What does it mean to say that a sentence expresses, or is used to express, an attitude, such as a belief? In addition, we may ask: To which items, more specifically, does the phrase 'value sentence' refer? I will start with the last question.

On a traditional view, a sentence is simply a (well formed) sequence of words, characters or phonemes that belong to a particular language. And value sentences are distinguished from others on the ground that they contain certain words, namely, in the case

---

<sup>2</sup> Of course, it sounds wrong to say that sentences express something by themselves. Rather, it is *we* who express beliefs, by using sentences. This should be kept in mind in what follows.

<sup>3</sup> For example, Frank Jackson and Philip Pettit thinks that (1) entails that ethical sentences have truth conditions. See their 'A Problem for Expressivism'.

of English, words such as ‘should’, ‘better’ and so on. As for the notion that an attitude is expressed by a sentence, it is sometimes assumed that to say that a sentence expresses, say, the belief that *p*, for a given speaker, is to say that, by sincerely uttering the sentence, the speaker represents herself as having the belief that *p*.<sup>4</sup>

Now, one problem is that, given these views, for example C1 does *not* seem to capture anything that expressivists and cognitivists can sensibly disagree about. For expressivists must surely concede that value sentences *are* sometimes used to express beliefs in the pertinent sense. For example, consider the following sentence:

‘It is morally wrong for Mormons to have premarital sex’

...and suppose that it is uttered as an answer to the following question: ‘What do Mormons think about sex and marriage?’. Since what Mormons think about sex is of course a straightforwardly factual matter, this example conflicts with E2.

One way to handle such trivial counter-examples is to restrict the pertinent claims ((E1), (C1), and so on) to certain *contexts*, namely those that are in some sense ‘typically moral’. If restricted in that way, the fact that ethical sentences express beliefs in *other* contexts is irrelevant. But a further problem is that, however we pick those contexts out, expressivists cannot reasonably deny that value sentences sometimes express beliefs in them as well.

This holds most clearly for ‘thick’ sentences, such as ‘Socrates was courageous’ which expresses straightforward beliefs about Socrates’ way of facing dangers.<sup>5</sup> But many expressivists concede, and should concede, that ascriptions of ‘thin’ ethical terms, such as ‘right’, ‘wrong’, and so on, might *also* acquire the capacity to express beliefs, *even* when they occur in typically moral contexts. For example, Richard Hare suggests that such terms may acquire that capacity ‘by reason of the constancy of the standards by which they are applied’.<sup>6</sup> In a similar vein, Charles Stevenson, another early expressivist, writes that:

to say that a man is ‘good’ may be to suggest that he has such traits as honesty, humility [...], and so on. [...] Within communities with well-developed mores these varied suggestions become fixed, and people tend to define ‘good’ in a way that makes the word strictly designate what it formerly suggested.<sup>7</sup>

Do Hare and Stevenson, by holding those views, commit themselves to thinking that, if such constancy were to occur in the case of an idiolect of a given language community, then their expressivism would be false of that idiolect? For example, consider the Swedish of the 13th century. In medieval Sweden, there were generally recognized moral authorities (the church and the priests), and, or so I have been told, considerable agreement over the standards by which actions were judged to be right and wrong. On the basis of the views Hare and Stevenson hint at, this may be enough for concluding that their value sentences expressed straightforward beliefs. Would that mean that Hare and Stevenson have to concede that it is cognitivism that gives the best account of 13<sup>th</sup> century Swedish and restrict their expressivism to other periods and idiolects, such as those of more modern societies in which the old authorities have fallen into disrepute and the consensus has withered? That would be an awkward result, given that expressivists, just like cognitivists, typically want to claim that their doctrines hold generally, for all languages, periods and speakers.

---

<sup>4</sup> That is, the utterance itself (provided that it is sincere) allows us to conclude that the speaker believes that *p*, even if we know nothing else about him than that he speaks the language to which it belongs.

<sup>5</sup> See, e.g., Gibbard, *Wise Choices*, 113.

<sup>6</sup> See *The Language of Morals*, 7, where he suggests that ‘right’ may acquire the capacity to refer to the property of maximizing utility in a society of convinced utilitarians.

<sup>7</sup> *Facts and Values*, 169, my italics. See also 9, 16, and 221.

### 3. Primary and Secondary Functions

In order to avoid such a result, expressivists must define their position so that it can somehow be reconciled with (C1). Hare tries to do this by stating it in terms of the ‘functions’ of value sentences. He says that, although value sentences may *secondarily* express beliefs, their *primary* function is to express conative states. Cognitivists can accordingly be construed as making the converse claim.

#### *Expressivism*

(E3) It is the primary function of value sentences to express conative attitudes.

#### *Cognitivism*

(C3) It is the primary function of value sentences to express beliefs.

But how is the talk of ‘primary’ and ‘secondary’ functions to be cashed out?

Hare does not say much about that question, but here is an idea: We can give it some content by construing expressivism and cognitivism as being theories, *not* about certain particular words or strings of symbols, but rather about the role, or about the features of their use, that, as it were, *make* them into value terms or value sentences, that make them such that utterances and ascriptions of them are correctly interpreted as expressions of answers to *moral* questions rather than something else (such as questions about what people *think* about morality). Obviously, words like ‘should’ and ‘right’ have this role *contingently*. They can lose that role and it could be acquired by other words. For example, some of the words that are used as value terms *today* were not used in that way before. Consider for example, the term ‘uncool’. In some contexts, that term *is* used for expressing moral evaluations, such as when our children (or my children), say that ‘depleting the world’s resources is so uncool’. Or consider how Mona Sahlin, the former leader of the social democratic party in Sweden once summarized her political vision:

‘Paying taxes is cool!’

So, there are certain features of the use of a set of sentences that *make* them into a value sentences. Which are these features? That is a question we have to ponder when we try to identify the value terms in other languages than our own, which expressivists and cognitivists have to be able to do if they want to claim that their positions hold universally. What does it take in order for a term in the target language to be correctly translated with, say ‘morally right’? Presumably, that that it has a similar role in planning and decision-making, and that the speakers have, in general, some motivation to perform actions they apply the term to. We may also require that the speakers, within broad limits, defend ascriptions of the term in similar ways. The latter type of constraint ensures that we find that there is at least *some* overlap between ourselves and the alien speakers about what to think about moral matters. We are destined to find, for example, that we agree that ‘ought’ implies ‘can’. But note that the required overlap is very slim. The constraint still allows us to find that there is also much disagreement. (The pertinent aspects correspond roughly, I think, to what Michael Smith calls ‘the practicality’ and ‘the objectivity’ of moral discourse.<sup>8</sup>)

We can now state the suggestion about how to understand the talk of a ‘primary’ function of value sentences as follows: It is the primary function of value sentences to express

---

<sup>8</sup> See chapter 1 of Smith’s *The Moral Problem*.

whatever it is that they do express *solely* in virtue of the features of our use of value sentences that make them into value sentences. What they express in virtue of *other*, non-essential features (non-essential relative to their status as being value sentences) is irrelevant. Cognitivists believe that the features are best explained by assuming that the sentences in question express beliefs, and accordingly that *that* is their primary function, while expressivists rather believe that the features are best explained by assuming that the sentences express conative attitudes.

Now, the kind of ‘constancy’ Hare and Stevenson wrote about is *not* among the features that make a set of sentences into value sentences. For, given that such constancy exists, if it were to evaporate, this would not mean that the sentences would no longer be value sentences. It would just mean that people have started to disagree more often regarding moral issues. To take the translation case: If it were found that all the alien speakers are disposed to apply the target term (i.e., the term we think should be translated with ‘right’) to exactly the same actions as others in their community, and moreover to exactly the same actions that *we* think are right, this would be interesting. But it would *not* be necessary in order for the translation to be correct. Therefore, what we are entitled to conclude on the basis of the existence of such constancy is irrelevant to E3 and C3. This idea explains why expressivists can concede that value sentences that belong to a certain idiolect, such as 13<sup>th</sup> century Swedish, express beliefs in virtue of such considerations and remain being expressivists about it.

A problem is, however, that cognitivists and expressivists are likely to *disagree* about which the relevant features of the use of our value sentences are. How can that be determined, without begging any questions? In a way, that is the central methodological question.

#### 4. Moral Convictions

Notice that, if the suggestion I just made is on the right track, then the reference to language in the traditional formulations of cognitivism and expressivism is really *redundant*. The idea is that certain features of our use of value sentences are relevant to the assessment of those positions only to the extent that they make us interpret assertions of those sentences as expressions of moral evaluations. But then it is the evaluations *themselves* that provide the focus of the interest, rather than the words we happen to use for expressing them.

That thought can be put as follows: Meta-ethics tries to uncover the nature of moral questions, such as the questions of whether euthanasia should be legalized. Do such questions allow for answers that can be correct or incorrect? One way to approach that issue is to explore what it is to *provide* an answer to a moral question. If we provide such an answer, and think, for example, that euthanasia *should* be legalized, then we have formed a certain conviction, a *moral* conviction. But what *kind* of a conviction is this? What *is* it to think, say, that euthanasia should be legalized? Cognitivists insist that it is to have a true or false belief about the legislation of euthanasia, while expressivists believe that it is rather to have a conative attitude towards it:

##### *Expressivism*

(E4) Moral convictions consist of conative attitudes (to think that the legislation of euthanasia should be legalized is to have a conative attitude towards it).

##### *Cognitivism*

(C4) Moral convictions consist of beliefs (to think that euthanasia should be legalized is to believe that legalizing euthanasia has a certain property)

Notice that, on this view about how to construe the central claims of expressivism and cognitivism, those positions should *not* be seen as being theories that are primarily concerned with language. After all, a moral conviction is a *psychological* entity. So, on the present construal, cognitivism and expressivism are best seen as a type of *psychological* theories. Meta-ethicists focus on moral language simply because speech behavior is *one* of the ways in which our moral convictions manifest themselves. But it is not the *only* way and it has ultimately no primacy relative to other ways.

Of course, these formulations *also* raise questions. One set of questions concern the term ‘belief’, which is ambiguous. On the one hand, it could refer to a psychological state (such as when we say that beliefs are dispositions to act). On the other hand, it rather refers to something—such as a proposition—that might provide the *content* of such a state (such as when we say that ‘all those beliefs are false’, thereby denying a given set of propositions, regardless of whether they are in fact believed by anyone). Since a moral conviction is a psychological entity, it is the first sense that is relevant in the context provided by (C4). The question is if moral convictions are instances of the type of psychological attitude we have towards a proposition when *believing* that it is true.<sup>9</sup>

What distinguishes that attitude from other attitudes, such as desires? Well, the crucial difference is that while beliefs can be true or false, desires cannot. Many participants of the debate elaborate this idea by saying that the pertinent attitudes have different ‘directions of fit’. Beliefs are states aimed at ‘fitting with the world’, while desires are states aimed rather ‘at making the world fit with them’.<sup>10</sup> Michael Smith has tried to clarify this metaphorical talk in the following way. To *believe* that *p* is to be in a state such that, if one were to perceive that not-*p*, it would tend to go out of existence. To *desire* that *p*, by contrast, is to be in a state that tends to *endure* when we perceive that not-*p*.<sup>11</sup>

There are problems with Smith’s suggestion, but, in the present context, I am interested in a different question. For we may again ask Wedberg’s question: What *sort* of a theory is, say, C4? Is it an empirical hypothesis, on a par with the claim that water consists in H<sub>2</sub>O? Or, should it rather be conceived of as a kind of definition of the phrase ‘moral conviction’, to be assessed relative to the common usage of that term?

In accordance with the first proposal, it might be tempting to conceive of moral convictions as forming a natural kind, whose nature might be uncovered by examining particular samples. Presumably—and this connects with the question about how to determine which features of our use of value sentences that *make* them into value sentences—this requires an independent criterion of what *counts* as such a sample. One option is to consult our *intuitions* about various cases, real or imaginary. But a problem is that our intuitions come apart.

For example, notice that expressivism entails *internalism*, the idea that it is *necessarily* the case that if a person thinks that she is morally required to do something then she has at least *some* motivation to do it. The reason is that, if to think that one is required to do something, is to have a conative pro-attitude towards that action, then having that thought *consists* in being thus motivated. Now, some people intuit that it is possible for a person to have genuine moral convictions even if she has no tendency *at all* to act in accordance with them, so called ‘amoralists’. Since that possibility is ruled out by expressivism, it conflicts with the intuition that fuels amoralism. So, there is at least one set of samples that it seems unable to account for. Moreover, in so far as the intuitions can be seen as manifestations of

---

<sup>9</sup> Do moral judgments have contents and, if so, in what do they consist? If they consist in beliefs, the answer is straightforward: Like (other) beliefs they do have contents, and their contents consist in propositions. However, in an expressivist setting the question is more complex.

<sup>10</sup> This idea is usually attributed to Elisabeth Anscombe. See her *Intention*.

<sup>11</sup> See Smith, *The Moral Problem*, 111-6.

what phrases like ‘moral conviction’ mean in common language it also undermines E4 when seen as a meaning-claim.

However, expressivists can, and often do, respond to such objections by simply insisting that they do not *share* the pertinent intuition. For even if the alleged amoralists share *some* features with people who have genuine moral convictions—we may for example imagine that they *say* that they think they are morally required to do various things—a crucial element is lacking, namely that they are not appropriately motivated. This game leads to a lot of fruitless bickering and allegations of begging the question from both sides. Maybe the correct diagnosis is that they simply talk past each other, by focusing on *different* concepts. Expressivists talk of *their* use of the notion of a ‘moral conviction’ while cognitivists talk of theirs. If so, a common subject matter is yet to be found.

However, there is a third option, as Wedberg indicated. On third proposal regarding that status of E4 and C4, the debate about the possibility of amoralists should be construed, *not* as a discussion about whether *the*, or *some*, existing concept of a moral conviction *does* allow for that possibility but whether it *should* do so. If participants can agree about the purposes such a concept should serve, there is still a room for a substantial disagreement between them.

Of course, to resolve a debate through stipulation may seem unsatisfactory. But what I am considering now is not stipulation just out of the blue. Consider what we may call the ‘finding’ that whales are not fish. It might be tempting to see this as the result of mere stipulation, since the fact that whales, unlike fish, give birth to living children, and so on, was known *before* the ‘finding’ was made. However, the revision was prompted by the fact that a concept of ‘fish’ that excludes whales is more useful in our overall theory about the species and their origins. The reason is that the animals that belong to the extension of the revised concept have certain *other* properties in common, properties that set them apart from other creatures in the sea and are assigned a special significance by the theory in question. This is, ultimately, what justifies treating the fact that whales give birth to living children as a crucial difference.

What this example illustrates is that the distinction between stipulations and ‘real’ empirical discoveries is less clear than one might think. It also illustrates that claims of this type are to be tested *holistically*. I said earlier that construing E4 as an empirical hypothesis may seem to require that there is an independent criterion that allows us to pick out the states that are moral convictions *prior* to determining if they are beliefs or desires. But that is really a rather naive idea. For the concept of a moral conviction is a *theoretical* concept, as we cannot directly *observe* that a subject has a moral conviction. In comparison, consider the thesis that electrons are negatively charged. Surely, there is little hope of determining whether an electron has been involved in a certain interaction *ahead* of finding out whether it is negatively charged. Rather, the claim that electrons are negatively charged is justified, if at all, in view of how well the theory to which it belongs as a *whole* fits the evidence.

Could something similar be said about C4 and E4? Well, a context in which it is natural to assess these claims is provided by the framework or theory philosophers refer to as ‘folk psychology’. Folk psychology posits a *number* of different types of psychological states people can be in, such as, besides beliefs, desires and moral convictions, intentions, wishes, hopes and so on. According to the central idea, it is by attributing such states that we best can explain, and predict, intentional behavior. Now, making these distinctions within the theory, and seeing each of the categories as denoting a distinct type, is justified to the extent that the states belong to the type are assigned a special explanatory task by the theory, in that attributions of those states are held to explain something that is not best explained by the attribution of (other) states.

As for moral convictions, one such task is to explain why people sometimes act against their (narrow) self-interest in a way that allows them to avoid ending up with the short end of

the stick in certain coordination cases. In such cases, if each does what will be better for him- or herself, they will jointly cause an outcome leaving all worse off compared to if they all had acted differently. The phenomenon to be explained is that people in many contexts are able to *avoid* ending up with such outcomes, which is, for example, why we have societies. And what explains this, it is held, is that people *do not* always maximize their self-interest. And sometimes they do not do this because they think that that is what *morality* requires. Indeed, according to some, such as me, it is the crucial role of moral convictions in such contexts that explains why the capacity to form such convictions has evolved.

This provides a context in which E4 and C4 can be assessed. For, as cognitivism and expressivism disagree about what having such thoughts entails, they generate different *versions* of explanations of the pertinent kind. We may accordingly ask which of these explanations are the *best*. If the cognitivist versions are better, then cognitivism obtains some support from its role in the explanation of the relevant behavior, and if the expressivist versions are better, the same holds for expressivism.

And the expressivist versions have at least *some* things going for them. For example, in order for the thought that a certain option is morally required to be able to have the pertinent impact, the agent must have some motivation to actually *do* what she thinks she is required to do. And, given that that thought is constituted by a desire of the kind expressivists have in mind, the motivation is *internal* to the thought. Cognitivists must tell a more complex story. For as moral convictions, according to cognitivism, are rather constituted by beliefs, then, given that beliefs cannot generate motivation by *themselves*, they must, in addition to attributing such convictions also assume that the agents have other states that, *together* with their moral views, generate the relevant motivation.

In addition, on a cognitivist view, we are to assume that all who are saved from their coordination problems by thinking that they are morally required to choose the options that save them ascribe the *same* particular property to those options. However, this group includes people from rather diverse cultural settings, who are likely to justify their moral assessments on the basis radically different considerations. Given normal views about how to determine belief-content, such differences make the assumption that they ascribe the same property difficult to sustain. On an expressivist view, by contrast, to think that one is required to do something is *not* to ascribe some property. It just consists in *favoring* the option in a certain way. And the assumption that two persons share that attitude towards an option is not undermined by differences of the relevant kind.

But my aim today is not to defend expressivism. The point is just that *if* the expressivist versions of the explanations are better, then we have some justification for adopting an expressivist view *even* if that can be described as a revision of an existing concept of a moral conviction given that it conflicts with the intuition that fuels amoralism. Indeed, the very idea behind the proposal is that we should try to find a way to go *beyond* intuitions of that sort when assessing the relevant theories and to expose them to a wider range of evidence. In my view, something along those lines is needed if we are to proceed beyond the kind of stalemate that contemporary meta-ethicists have maneuvered themselves into today.<sup>12</sup>

---

<sup>12</sup> For further discussion of the themes in this paper, see Tersman, F., *Moral Disagreement*, New York: Cambridge University Press, 2006.

*References*

- Anscombe, E., *Intention*. Oxford: Blackwell, 1957.
- Broad, C.D., *Review of Wedberg, A., Filosofins historia. Philosophical Quarterly* 18 (1968), 269-71.
- Gibbard, A., *Wise Choices, Apt Feelings*. Oxford: Clarendon Press, 1990.
- Hare, R.M., *The Language of Morals*. Oxford: Clarendon Press, 1952.
- Jackson, F & Pettit, P. 'A Problem for Expressivism'. *Analysis* 58 (1998), 239-51.
- Smith, M., *The Moral Problem*. Oxford: Blackwell, 1994.
- Stevenson, C.L., *Facts and Values*. New Haven, CT: Yale University Press, 1944.
- Tersman, F., *Moral Disagreement*. New York: Cambridge University Press, 2006.
- Wedberg, A., *Filosofins historia. Från Bolzano till Wittgenstein*. Stockholm: Bonnier, 1966.