Daniel Eggers, Nothing new in Ecumenia? Hare, hybrid expressivism and *de dicto* beliefs.

This is the peer-reviewed version of the article "Nothing new in Ecumenia? Hare, hybrid expressivism and *de dicto* beliefs", published in *Ethical Theory and Moral Practice*, 19, 2016, pp. 831–847. The final publication is available at: https://link.springer.com/article/10.1007/s10677-015-9681-6

One important recent trend in the metaethical debate over expressivism and cognitivism is the emergence of 'hybrid' or 'ecumenical' theories which purport to open up new territory beyond the well-rehearsed oppositions. According to such theories, moral sentences such as 'abortion is wrong' express *both* beliefs, as cognitivism has it, *and* desire-like states, as expressivism has it. The main reason why hybrid approaches have recently become popular is that they promise solutions to the key problems that have long plagued cognitivists and expressivists respectively, such as cognitivism's problem of accounting for the intimate connection between moral judgement and motivation, and expressivism's problem of accounting for the uses of moral predicates in unasserted contexts (better known as the 'Frege-Geach problem').

However, one might think that the hybrid move is not as novel as some of its advocates seem to take it to be and that hybrid expressivists have simply re-invented the idea that moral words have both descriptive and emotive (or evaluative) meaning, which we find in Charles Stevenson and Richard Hare. Stevenson not only argues that ethical words, such as 'good', combine two different kinds of meaning. He also defines his crucial concept of emotive meaning by appealing to a word's power to *express* non-cognitive attitudes of the speaker (see, for example, Stevenson 1944, 33), and he suggests a more general psychologistic or mentalist approach to meaning by offering a parallel analysis of descriptive meaning in terms of the expression of belief (see Stevenson 1944, 34; see also Stevenson

1963a, 9, and 1963b, 79f.). Similarly, Hare not only takes moral words or sentences to have descriptive as well as evaluative meaning. At least the late Hare also seems somewhat sympathetic to the idea of describing the meaning of moral words or sentences in terms of the mental states that they express. Though Hare continues to vehemently reject a causal understanding of 'expression' that assimilates the relevant kind of linguistic expression to the act of 'evincing' certain feelings that well up inside the speaker (see Hare 1952, 10; and Hare 1999, 105), he does not seem opposed to the idea that moral words or sentences express certain states of mind by linguistic *convention*. In fact, Hare even suggests that the attempt to cash out the common idea of linguistic expression in strictly linguistic, non-psychologistic terms faces serious problems (see Hare 1999, 106f.). This may be one of the reasons why, in addition to claiming that moral words or sentences express prescriptions and statements of fact, Hare also, if quite rarely, characterises prescriptions as linguistic expressions of desires (in a broad sense of the word) and statements of fact as linguistic expressions of beliefs (see Hare 1981, 107).

Michael Ridge has recently argued that, despite these affinities, we ought not to see Hare as a hybrid expressivist (see Ridge 2006, 309f.). According to Ridge, the descriptive meaning of moral sentences postulated by Hare is contingent on certain contextual conditions, which means that there can be moral sentences that do not express any descriptive beliefs at all. In Ridge's view, the contingent nature of descriptive meaning prevents Hare from making use of the hybrid machinery for solving the 'Frege-Geach problem'. However, Ridge's reading of Hare has recently been challenged by John Eriksson who has defended the exact opposite conclusions: According to Eriksson, Hare does not take descriptive meaning to be a merely contingent feature of moral sentences and deserves to be seen as a true hybrid expressivist (see Eriksson 2009, 8f. and 16-20). Yet, far from being of only historical interest, Eriksson's defense of Hare's status as an hybrid expressivist is directly relevant to the current

systematic discussion in that it intimates that modern hybrid expressivism should actually reclaim its historical roots: According to Eriksson, it is Hare's particular version of hybrid expressivism, and not Ridge's, that provides us with an appropriate solution to the 'Frege-Geach problem', and for this very reason, the hybrid account of moral utterances developed by Eriksson explicitly follows in Hare's footsteps.[1]

While I am sympathetic to the cause of emphasising the importance of descriptive meaning in Hare against Ridge's somewhat cursory discussion, I think that Eriksson's treatment of the issue does not really settle it. Eriksson glosses over some problems in Hare's discussion of descriptive and evaluative meaning and, as a result, paints a slightly too harmonious picture of Hare's position. Since it seems that these problems are exactly what fuels Ridge's non-ecumenical interpretation of Hare, Eriksson thereby also fails to do full justice to Ridge's worries. Furthermore, it seems questionable whether the position Eriksson attributes to Hare, and which he then embraces for himself, really provides the best version of hybrid expressivism, given that it seems to be open to some crucial objections. In what follows, therefore, I want to take up the issue once more and argue for four conclusions. First, I want to show that we can defend Eriksson's reading of Hare with regard to the official version of Hare's theory, by providing further evidence that Hare takes descriptive meaning to be an indispensable element of moral sentences. I will, secondly, argue that, in line with what we may take to be Ridge's perspective on Hare, this official version faces serious difficulties, resulting from the possibility of unknown speaker standards. These difficulties suggest that any critical modern reconstruction of Hare's position should give up on the inescapability claim. Thirdly, I will demonstrate that a more conservative modern reconstruction of Hare's version of hybrid expressivism, which is suggested by Eriksson in passing and appeals to the idea that moral sentences express what I will refer to as '*de dicto* beliefs', or, for that matter,

[1] For more recent defenses of Eriksson's expressivist account, see also Eriksson 2014, 2015a and 2015b.

any modern version of hybrid expressivism developed along similar lines, fails. One crucial problem with this approach is that it will clearly not provide us with a successful solution to the 'Frege-Geach problem'. Lastly, I will try to argue that the fact that we may legitimately attribute certain '*de dicto* beliefs' to a speaker uttering a moral sentence does not suffice to show that pure expressivism is an unstable theoretical position and that any plausible expressivist analysis of moral language must collapse into a form of hybrid expressivism, something that has been suggested by Michael Smith (see Smith 2001).

I.

Hare develops his account of moral language in various places. The central elements of his analysis, however, including the distinction between descriptive and evaluative meaning, remain a permanent feature of Hare's account from *The Language of Morals* up to *Sorting out Ethics*. According to Hare, moral words, such as the word 'good' on which the analysis of *The Language of Morals* focuses, are characteristically used for commending something: While there may be quite specific cases in which we use 'good' with no commendatory intention at all (see Hare 1952, 124), it is misleading to follow descriptivists into thinking that the meaning of 'good' can wholly be analysed in descriptive terms. According to Hare, we use 'good' in order to guide people's choices, and the reason why we can do so is that in using 'good' we implicitly declare our acceptance of certain standards and thereby express attitudes of approval or disapproval.

However, as Hare is keen to emphasise, the word 'good' nevertheless shares important features with purely descriptive terms like 'red'. If we apply 'good' to a certain object, as when we say of a person that she is a (morally) good person, we also convey that the object in question has certain descriptive properties, namely those corresponding to the standard by

which we are judging the object and in virtue of which we are commending it. For Hare, therefore, the word 'good' has *both* evaluative *and* descriptive meaning, and the same is true of other moral words such as 'bad', 'right', 'wrong' or 'ought'. The descriptive meaning is crucially associated with the universalizability and supervenience of moral judgements and ensures the rationality of moral thinking and moral argument. The evaluative meaning is crucially associated with the prescriptivity of moral judgements and accounts for the practical nature of moral judgement and the link between moral judgement and motivation.

Hare makes several attempts to more clearly define the term 'descriptive meaning'. According to Hare, the term refers to the "reasons" (1952, 118) for which a speaker commends something, or to the "qualities" (1952, 148) or "properties" (1981, 88) which he commends, or to the "virtues or good-making characteristics" (1952, 145) in virtue of which he makes his commendation. In *Sorting out Ethics*, we also find more sophisticated characterisations according to which the descriptive meaning of a moral statement is the same thing as the "semantics of the statement" or its "truth conditions" (1999, 52) or as the "application conditions" (1999, 138) of the words contained in it.[2]

The main question to be asked against the background of Hare's works is whether Ridge is right in claiming that Hare makes descriptive meaning thus described a contingent feature of moral sentences and allows for moral sentences that do not possess any descriptive meaning at all. As already indicated, Eriksson has recently rejected Ridge's claims as false. He does so by emphasising the link between descriptive meaning on the one hand and supervenience and universalizability on the other, and by drawing attention to Hare's claim

[2] I will not here delve into the question of whether these various characterisations of descriptive meaning really amount to the same thing. There can be no doubt, however, that this is exactly Hare's view: "It should be evident by now that the same animal is here appearing in different metamorphoses. It does not make any difference whether we speak of criteria of application for a moral word (…), or about the word's descriptive meaning, or about the truth conditions of statements containing it, or about a moral standard or universal principle." (Hare 1999, 138)

that when we call something good or right, we do so for certain reasons. According to Eriksson, such reasons are indispensable for any moral judgement, which he takes to be evident from the fact that we can always respond to a claim of the form 'X is good' by asking 'Why do you think so?' (see Eriksson 2009, 16f.).[3]

Although the role of 'reasons' and the link with supervenience or universalizability are indeed crucial for Hare's understanding of descriptive meaning, one might feel that the evidence provided by Eriksson raises problems of its own – if not for Eriksson, then for Hare. Thus it seems that Eriksson's claim that we can always respond to moral utterances with 'Why do you think so?' is a kind of double-edged sword. Surely, that this question is usually open indicates that we generally expect people who utter moral sentences to have reasons for what they say. Yet, it also seems to indicate that these reasons are not part of what their utterances *mean*. After all, when a speaker utters a descriptive sentence, such as 'my new car is red', we do not take the question 'why do you think so?' to be an open question. If somebody responded to the utterance in this way, we would tend to think that he was not wholly familiar with the meaning of the word 'red'. However, if descriptive meaning is what moral sentences share with purely descriptive sentences, then why should the case of moral sentences be so different?

This variant of Moore's 'open question argument' suggests two things. First, it suggests that we should try to supplement the evidence provided by Eriksson in order to further back up the view that Hare takes descriptive meaning to be an indispensable aspect of the meaning of moral sentences. Secondly, it suggests that we might need to distinguish between the official statement of Hare's position and the most coherent and plausible reconstruction of it. After all, what the above version of the 'open question argument' may be taken to demonstrate is not, strictly speaking, that Hare cannot officially have taken

[3] See Eriksson 2009: 16f.

descriptive meaning to be an indispensable element of moral sentences. Rather, it suggests that he *should* not have taken it to be such an element, because the resulting view might be difficult to defend. In a similar vein, it may be that Ridge's characterisation, while failing to correctly summarize Hare's official tenets, nevertheless provides a correct description of what the most coherent and plausible reconstruction of Hare's account would look like.

Now with regard to the official version of Hare's position, there is indeed ample evidence that Ridge gets Hare the wrong way around. What Hare concedes, and what has already been indicated above, is that moral sentences may be devoid of *evaluative* meaning, not that they may be devoid of descriptive meaning. Nowhere does Hare make this latter claim, and he makes quite a couple of claims that point in the opposite direction. In one of the relevant passages, which is cited by Eriksson, Hare not only emphasises that a moral judgement, such as 'St. Francis was a good man', is based on certain descriptive characteristics of the object we evaluate. He also claims that the judgement is *logically dependent* on the judgement that the object has those characteristics (see Hare 1952, 145). Similarly, in *Moral Thinking*, Hare suggests that the universalizability of moral judgements makes moral judgements logically dependent on the descriptive properties which provide the reasons for the judgement (see Hare 1981, 216f.).

Moreover, in *Sorting out Ethics*, Hare explicitly opposes moral judgements with ordinary imperations, such as 'Do that!', and the aspect he focuses on is that ordinary imperations do not have to be issued for reasons: What distinguishes moral judgements from ordinary imperations, and conjoins them with constative speech acts, is precisely that moral judgements "have to be made for reasons" (1999, 12) or, as Hare also puts it, "that when I say 'I ought to do that' I have to say it because of *something about* the act that I say I ought to do" (1999, 11). All this shows quite clearly that Hare takes descriptive meaning to be an indispensable element of moral sentences, and this interpretation finds further support in the

fact that a few pages later, Hare explicitly refers of the "inescapable factuality or descriptivity of moral principles" (1999, 20). Moreover, in a passage from *Freedom and Reason*, Hare highlights the very same contrast between moral sentences and singular imperatives by describing the latter, but not the former, as "expressions which […] do not carry descriptive meaning at all" (1963, 27).

II.

We may conclude, therefore, that, as long as we take it to be a description of Hare's official position, Ridge's rendition of Hare's account of descriptive and evaluative meaning is misleading. As already indicated, however, we may also want to read Ridge's characterisation as some sort of critical reconstruction emerging from the view that Hare's official position faces certain problems. That something like this might be what lies behind Ridge's reading becomes more plausible once we look at the reasons why Ridge thinks that descriptive meaning is contingent in Hare. The argument provided by Ridge is that there are cases in which the standard by which a speaker is judging is neither shared nor even known by his hearers, which means that the hearers will be ignorant of the exact descriptive properties or qualities in virtue of which the speaker does, or does not, commend the object to which he applies a moral predicate. For Ridge, it obviously seems misleading to say of such cases that the sentence uttered by the speaker has descriptive in addition to evaluative meaning.

There are a couple of reasons why such a critical way of reading Hare demands serious consideration. The first is that Hare explicitly allows for the possibility of unknown speaker standards. In fact, Hare even seems to allow for cases in which the speaker himself is not fully aware of the standard by which he is judging (see Hare 1952, 58f.). In other words: the particular element on which Ridge ultimately bases his reading of Hare is undoubtedly

part of Hare's official position, even if what Ridge concludes from it may be not. The second reason is that, even if the possibility of unshared or unknown moral standards were not part of Hare's official position, one may want to argue that it *ought* to be part of it. It seems impossible to deny that in many, if not most, cases, we do not know the exact reasons that prompt a speaker to utter a moral sentence of the type 'X is wrong' or 'X is bad'. It seems equally impossible to deny that individuals sometimes utter moral sentences of the type 'X is wrong' or 'X is bad' without themselves having a clear conception of why X is wrong or bad. That the exact standard by which a speaker is morally judging an action or a person is often unknown to himself is strongly suggested by recent work in social psychology. In their empirical studies on moral reasoning, Jonathan Haidt and his co-researchers presented American undergraduate students with stories of deviant sexual behaviour and pressed the participants on their instant moral judgements, finding that they strongly clung to their initial judgements and often resorted to saying things like 'I don't know, I can't explain it, I just know it's wrong' or 'I know it's wrong, but I just can't come up with a reason why' (see Haidt, Björklund, Murphy 2000; and Haidt 2001). In view of all this, we may take the possibility of unknown moral standards to be a general constraint that any plausible theory of moral language needs to satisfy. Moreover, that the issue is of more than just exegetical relevance is demonstrated by the fact that Eriksson himself explicitly subscribes to the position he attributes to Hare and obviously takes it to be the one that modern hybrid expressivists should opt for.

The third reason is that Ridge's conclusion, according to which the unknown standard of a speaker uttering a moral sentence cannot be taken to infuse his utterance with descriptive meaning, seems quite intuitive. One of the most important strategies in arguing for hybrid theories of moral language is to refer to the existence of thick moral predicates (see also Ridge 2006, 309f.). The reason is that it is not at all easy to see how a non-ecumenical

interpretation of terms such as ‘industrious’, ‘tidy’, ‘honest’ or ‘courageous’ could be justified. While all these words imply a certain (positive) evaluation, it seems equally true that they also refer to quite specific descriptive properties and would not be applied to persons who lacked these properties, even if these persons were to qualify for a positive moral evaluation for other reasons. We can support this idea with the help of the ‘open question argument’: it does not seem to make sense to respond to the sentence ‘Peter is courageous’ by saying something like ‘I understand that you approve of Peter, but do you also think that he is not easily frightened?’ There are good reasons, therefore, to argue that by using thick moral predicates, a speaker conveys something about the standard by which he is judging or, to put it differently, conveys something about the descriptive beliefs he holds with regard to the object of his evaluation. Moreover, that he so conveys something about his beliefs is the result of certain linguistic conventions, regulating the use of the terms like ‘courageous’. Since the concept of linguistic meaning seems to be closely related to the idea of conveying information by way of linguistic convention, it seems natural so say that moral sentences making use of thick moral concepts have descriptive meaning or express descriptive beliefs.

It is exactly the distinction between thick and thin moral concepts, though, which raises a crucial challenge for modern hybrid theories and for Hare’s account of descriptive and evaluative meaning because it is far from being clear that the argument works for thin moral concepts as well. As we have already argued, we seem to have no reason for thinking that a speaker making use of terms such as ‘good’ or ‘right’ must have any particular descriptive belief about the object to which he is applying them. Neither do we seem to have reason to claim that the speaker conveys information about the descriptive properties of the object to his audience, even if we may agree that he conveys approval. It is not at all obvious, therefore, why we should take certain descriptive properties to be part of either speaker or hearer meaning. Rather, it seems that we should conclude that the role of descriptive meaning

is exactly what marks the crucial difference between thick and thin moral concepts: whereas thick moral concepts express descriptive beliefs and have descriptive meaning, thin moral concepts do not.

In fairness, it ought to be noted that Hare somewhat anticipates these worries and makes an attempt to respond to them. What is relevant here is not so much Hare's claim that thin and thick moral predicates are distinguished by the fact that in the case of the latter, the descriptive meaning is *primary*, whereas in the former, the descriptive meaning is *secondary* to the evaluative meaning (see Hare 1952, 118ff. and 149ff.). Since the task is to justify that thin moral predicates have a distinctive descriptive meaning *at all*, this distinction, as plausible as it may otherwise be, will not get Hare off the hook. However, we find a second relevant move in Hare. Stressing his partial analogy between descriptive and evaluative terms, Hare claims that the looseness or vagueness we encounter in the descriptive meaning of moral predicates such as 'good' is something we find in exactly the same way in purely descriptive predicates such as 'red'.[4] In both cases, Hare argues, we usually share a certain understanding of the relevant descriptive characteristics, even though, in both cases, there may be occasions on which we do not know the exact standards of a speaker or are mistaken about them. For Hare, therefore, the indeterminacy of the descriptive meaning of moral predicates does not provide any good reasons not to refer to it as a form of meaning, unless we also want to conclude that ordinary descriptive predicates such as 'red' do not have descriptive meaning, either.

However, Hare clearly overstates the similarity between ordinary descriptive and evaluative predicates at this point: While it may be true that our understanding of what makes a thing red is somewhat vague and that the views of different speakers may vary, these

[4] See Hare 1952: 115: "It is important to notice that the exactness or looseness of their criteria does absolutely nothing to distinguish words like 'good' from words like 'red'. Words in both classes may be descriptively loose or exact, according to how rigidly the criteria have been laid down by custom or convention. It is certainly not true that value-words are distinguished from descriptive words in that the former are looser, descriptively, than the latter. There are loose and rigid examples of both sorts of word."

variations seem quite different from the variations which, on Hare's account, we potentially have to deal with in the case of moral utterances, where the descriptive meaning of different speakers' uses of 'wrong' may range from 'causes more pain than pleasure' to 'is incompatible with the categorical imperative' or from 'causes a feeling of disgust in me' to 'is forbidden by the Bible'.[5]

Moreover, Hare's analogy between descriptive and evaluative predicates, if fully convincing, would ultimately prove too much. Even if we were to grant that there is far-reaching agreement among different people concerning the descriptive properties that make actions morally good, we would surely have to admit that we find at least as much agreement in the *attitudes* different people hold with regard to certain objects. However, we find such an agreement in attitude not only with regard to forms of behaviour which can plausibly be cited as the descriptive element of thick evaluative concepts such as 'courageous', but also with regard to objects for which no such interpretation is plausible, such as strawberries (to use one of Hare's own examples). Thus, if the fact that most people agree about the descriptive qualities that make an action 'right' or a person 'good' were enough to justify the claim that 'right' or 'good' have descriptive meaning, why does the fact that most people enjoy the taste of strawberries not likewise justify the idea that the term 'strawberry' has evaluative meaning? Once we accept Hare's analogy between descriptive and evaluative terms, therefore, it seems that we can no longer restrict the notion of evaluative meaning to those established examples of thin and thick evaluative concepts to which Hare himself wants to restrict it.

Hare's extended analogy between descriptive and evaluative words, therefore, does not really help him to meet the challenge set by the distinction between thin and thick moral

[5] Something like this also seems to be Urmson's idea who argues that, on Stevenson's related account, ethical predicates turn out to be, not only vague (as both Stevenson and Hare claim), but in fact downright ambiguous (see Urmson 1968, 78).

predicates. It seems that there are three other ways for Hare and modern defenders of hybrid expressivism to respond to this challenge. The first is to simply bite the bullet and confine the proposed hybrid analysis to thick moral concepts. Theorists taking this route would choose to be hybrid expressivists about thick moral concepts and pure expressivists about thin moral concepts. As far as I can see, nobody has explicitly argued for such a view, and one of the reasons may be that by endorsing this view, one would give up on the extra resources which make ecumenical theories attractive in the first place, most importantly the resources for solving the Frege-Geach problem (which problem obviously arises for thin moral concepts just as it does for thick moral concepts).

The second strategy is to argue for a wider conception of meaning which does not tie linguistic meaning to what a speaker conveys, intends to convey or is himself aware of in uttering something and which is, therefore, immune to worries based on the 'open question argument'. There is not much evidence that Hare himself wanted to have his theory understood in this way or would have been sympathetic to the enterprise. It is true that Hare embraces a somewhat wider notion of 'meaning' that appeals not only to semantic aspects, but also to logical properties, to aspects of syntax and to aspects such as illocutionary force which we may classify under the heading of pragmatics (see Hare 1999, 43-51). However, the conception of meaning I have sketched above is already a broader understanding of meaning in this very sense. Moreover, there are a couple of passages which suggest that Hare accepts the link between meaning and the idea of conveying something by way of linguistic convention (see, for example, Hare 1952, 118; and 1963, 6f.).

Of course, this need not keep modern hybrid expressivists from employing this second strategy. It should be noted, however, that the strategy comes at a certain cost. It not only commits the hybrid expressivist to giving up on our common sense understanding of linguistic meaning and to appealing, for example, to the kind of Kripke/Putnam style

semantics employed by Cornell realists such as Richard Boyd and David Brink (see Boyd 1988; and Brink 1989). The hybrid expressivist runs the risk of thereby losing some of the dialectical advantages over metaethical naturalism, in particular those provided by the 'open question argument'.

The problem here is not so much that hybrid expressivism becomes vulnerable to the 'open question argument' in its own right. Though hybrid expressivists employing the strategy in question will certainly face the charge that their account of the descriptive meaning of moral sentences is somewhat arbitrary, they are still in a position to fend off the 'open question argument' in its classic form: they can still explain why the question 'I agree that X has the descriptive property D, but is X also morally wrong?' is open while the question 'I agree that X has the descriptive property D, but does X also have the descriptive property D?' is closed because they can argue that, unlike the sentence 'X has the descriptive property D', the sentence 'X is morally wrong' additionally expresses a negative attitude which a person asking the question need not have. The real problem is that this appeal to attitudes is no longer necessary in order to respond to the 'open question argument' in the first place and that, for this very reason, the hybrid expressivist can no longer turn the argument against sophisticated naturalists such as Boyd or Brink: if the descriptive meaning of moral sentences is not transparent to competent speakers but rather something of an *a posteriori* matter, then this alone seems sufficient to explain why the question 'I agree that X has the descriptive property D, but is X also morally wrong?' appears to be open because the semantic equivalence of 'X has the descriptive property D' and 'X is morally wrong' might simply be intransparent to competent speakers. By employing a less conventional conception of semantic meaning, therefore, the hybrid expressivist loses the ability to appeal to the 'open question argument' in order to claim superiority over naturalist analyses of thin moral concepts. Nor can he continue to appeal to an interpretation of moral disagreement as 'disagreement in attitude' by simply

pointing to the everyday experience that people morally disagree over something even where they share all relevant descriptive beliefs: once the hybrid expressivist concedes that the descriptive property to which moral terms refer is not transparent to us, any naturalist can use this concession to claim that in every case of moral disagreement, there *is* in fact a relevant disagreement in belief, notwithstanding our difficulties in identifying the belief in question.

The third and last strategy is to try to defend the idea that moral sentences have descriptive meaning against the background of a more conventional understanding of meaning and the possibility of unknown speaker standards. What a hybrid theorist opting for this strategy would first and foremost have to do is to provide a plausible *specification* of the descriptive meaning of thin moral concepts, or of the descriptive beliefs that are expressed by speakers applying such concepts. It is this way of responding to the worries about Hare's position which Eriksson seems to favour, and if the strategy can successfully be pursued, this would be good news to any modern hybrid expressivist who does not want to face the prospect of having to shoulder some of the crucial burdens of naturalism. In the next section, I will examine the prospects of this strategy and ask whether there is a way to concede that the exact standards of a speaker uttering a moral sentence may be unknown to his hearers or even to himself, but to nevertheless argue that his utterance conveys that the speaker holds certain descriptive beliefs.

III.

In what sense, then, can we plausibly claim that a speaker uttering a moral sentence, such as 'abortion is wrong', expresses a particular descriptive belief – a belief which her hearers may attribute to her in virtue of linguistic conventions, even when they are ignorant of the actual reasons that led her to negatively evaluate abortion? It seems that there is at least one thing we

know about the beliefs of the speaker, namely that she believes that there is something in abortion which makes it wrong. Even if we concede that the speaker may not be fully aware of the descriptive properties which led her to her judgement, it seems that we must at least attribute to her the belief that abortion has *some* such properties. To think otherwise seems incompatible with the ideas of supervenience or universalizability. Accordingly, we may not sensibly respond to a speaker uttering 'abortion is wrong' by saying 'I understand that you commend not to allow abortions, but do you also think that abortion has any descriptive properties in virtue of which you are commending this?'

It seems, then, that we can identify at least one belief that we may attribute to a speaker using thin moral predicates such as 'good', 'right' or 'ought'. This belief is what we may refer to as a '*de dicto* belief': it is not the substantial belief that the evaluated object, e.g. a particular action, has certain specified properties, but only a kind of summary belief to the effect that the descriptive properties which the action possesses (and which are left unspecified) warrant a certain reaction. The possibility of interpreting the belief expressed by moral sentences along these lines is explicitly pointed out by Eriksson who suggests it as a possible way of escaping the problems resulting from unknown speaker standards but who does not further discuss the prospects of the strategy (see Eriksson 2009, 29). However, given the way in which the *de dicto* belief can be said to express the very idea of supervenience or universalizability, it seems that it might be this strategy that actually conforms best with the spirit of Hare's original approach to which Eriksson is so obviously indebted. In fact, in a somewhat older paper, Michael Smith has suggested that we should read Hare's account of moral language in something like these terms and argued that, in virtue of the fact that beliefs of the type in question are plausibly expressed by all moral sentences, pure expressivism is not a stable position at all but necessarily collapses into this kind of Harean hybrid expressivism (see Smith 2001, 95ff.). Finally, some remarks made by Daniel Boisvert in the

course of criticizing speaker-relative accounts seem to suggest a similar interpretation of the descriptive content of moral sentences. Boisvert emphasises that the descriptive properties picked out by the predicates 'right' and 'wrong' are not specific properties but the more general properties of rightness and wrongness (see Boisvert 2008, 178), and one way of reading this claim is to interpret it as saying that 'right' and 'wrong' pick out those descriptive properties that make an action wrong, whatever those properties happen to be.[6]

There is one further reason why the view that moral sentences express the kind of *de dicto* just described demands serious consideration. It is that, on the face of it, the account in question claims to differ in an important respect from two other hybrid theories which appeal to the idea of *de dicto* beliefs (or something very much like it), namely from the hybrid account of Stephen Barker and the original version of Michael Ridge's *ecumenical expressivism*. The crucial idea of both theories is that we may identify a constant descriptive content of moral sentences by unspecifically referring to the property towards which the attitude expressed by such sentences is directed. According to the simpler of Ridge's two proposals, which he describes as the "plain vanilla" (2006b, 57) account, sentences such as 'there is moral reason to X' can be taken to conventionally express (a) an attitude of approval of a certain kind toward actions insofar as they have a certain property and (b) a belief that X has that property. Similarly, according to Barker's analysis, sentences of the type 'X is good' express (a) the belief that X instantiates the kind of general property the speaker approves of, and (b) the speaker's approval of this kind of general property. However, as has convincingly been argued by Daniel Boisvert and Mark Schroeder, the problem with these accounts is that they conceive of the constant descriptive content of moral sentences as speaker-relative and

[6] Note, however, that in a later passage, Boisvert himself introduces a more specific property, namely 'maximizing general welfare', as a possible candidate for the descriptive property denoted by the term 'right' (see Boisvert 2008, 185).

are therefore unable to deal with so-called 'attitude ascriptions', as in 'John believes that abortion is morally wrong' (see Boisvert 2008, 195f.; and Schroeder 2009, 284ff.).

A second but somewhat neglected problem with such speaker-relative versions of hybrid expressivism is that they cannot give a convincing account of moral disagreement.[7] Both Barker and Ridge claim that the object of the desire-like attitude expressed by moral utterances is not the particular action or person that is being evaluated, but a more general property which is instantiated by the action or person in question. Moreover, as Mark Schroeder has plausibly argued, it is only in virtue of conceiving of the desire-like attitude in this way that hybrid approaches are able to solve the 'Frege-Geach problem', which, as we have seen, is one of the main motivations behind the ecumenical move (see Schroeder 2009, 276-78).[8] On Ridge's and Barker's approach, therefore, different speakers not only express different beliefs when they make a moral utterance, but the attitudes they express may relate to quite different general properties. However, if the opposing attitudes expressed, for example, by John's utterance 'abortion is wrong' and Peter's utterance 'abortion is not wrong' do not relate to the same general property, then we do not seem to have a case of proper moral disagreement at all, because John and Peter might just be talking about different general properties that are equally instantiated by abortion and might, after all, possess an absolutely identical set of belief-attitude pairs.

In contrast, the account suggested by Eriksson, Smith and Boisvert promises an explanation of the belief expressed by moral sentences in *non*-speaker-relative terms, which means that it would allow the hybrid expressivist to overcome the problems with 'attitude ascriptions' and moral disagreement without having to postulate some kind of intransparent descriptive belief on behalf of speakers uttering moral sentences. Yet, the crucial problem

[7] A similar worry has, however, recently been raised by Eriksson (see Eriksson 2015b, 45ff.).

[8] See, however, Strandberg's recent critical discussion (Strandberg 2015) according to which even versions of hybrid expressivism that satisfy Schroeder's criteria fail to provide a solution to the 'Frege-Geach problem'.

with this way of defending hybrid expressivism about thin moral concepts or, for that matter, with a reconstruction of Hare in terms of hybrid expressivism, is that the account just cannot sufficiently make good on this promise. The initial problem, which is not sufficiently addressed by either Smith or Boisvert, is how to specify the content of the belief without invoking any normative terms. The most natural way to describe the content of the *de dicto* belief expressed by 'abortion is wrong' may be to describe it as the belief that abortion has certain wrong-making characteristics – which, as a matter of fact, would be very much in line with one of Hare's characterisations of descriptive meaning. However, the obvious problem with this analysis is that 'wrong' – the moral predicate for which we wanted to provide a hybrid expressivist analysis in the first place – appears as part of the content of the belief, which means that the belief is not, after all, a purely descriptive belief, even though the 'characteristics' to which it refers may be constituted by purely descriptive properties. However, we seem to run into a similar kind of problem if we try to specify the content of the *de dicto* belief in question without using the term 'wrong', such as when we describe it as the belief that the descriptive properties of abortion provide *reasons* for having a negative attitude towards it or *warrant* disapproval. While these specifications may strip the belief of any obvious moral content, they fall short of turning it into a proper descriptive belief.

One idea of how to get rid of the normative content would be to think of the relationship between the belief and the attitude expressed by moral sentences not in terms of 'warrant' or in other normative terms (such as in terms of 'justification'), but in *explanatory* or *causal* terms. According to this idea, the *de dicto* belief expressed by 'abortion is wrong' is not the belief that 'certain descriptive properties of abortion warrant a negative attitude towards it', but the belief that 'certain descriptive properties of abortion *explain* or *cause* a negative attitude towards it'. It seems, however, that we may legitimately attribute such a belief to a speaker judging that abortion is wrong only in virtue of the fact that *he himself*

expresses such a negative attitude towards abortion. Yet, if the true descriptive content of 'abortion is wrong' is given by the belief that 'certain descriptive properties of abortion explain or caused *my* (the speaker's) negative attitude towards it', then we are again left with a speaker-relative content after all. We may surely attribute the belief in question to everybody uttering the sentence 'abortion is wrong', even to a speaker who does not know exactly why he disapproves of abortion. Thus it seems strange to imagine a speaker who disapproves of abortion, but who does not believe that there is an explanation for why he does so which appeals, in some way or the other, to the properties of abortion. Similarly, it seems strange to imagine that a speaker disapproving of abortion would not believe that his disapproval is caused by something in the object he disapproves of. However, the crucial question is whether the above proposal has any advantages over Ridge's and Barker's accounts when it comes to attitude ascription and moral disagreement.

The answer is that the proposal may have some such advantages when it comes to explaining moral disagreement, but that it has the same problem with attitude ascriptions. Moreover, the very same feature that makes the analysis immune to the problems with moral disagreement renders it incompatible with the solution to the 'Frege-Geach' problem that makes Ridge's and Barker's accounts attractive in the first place. To begin with attitude ascription: If we imagine a speaker, call him Robert, to be uttering the sentence 'John believes that abortion is wrong', we quickly see that there is no way to accommodate the descriptive content which the proposal under consideration forces us conceive of as part of John's belief. Thus, the following four options all seem inappropriate:

> *1. Certain properties of abortion explain or caused my (John's) negative desire-like attitude towards it.
>
> *2. John believes that certain properties of abortion explain or caused my (John's) negative desire-like attitude towards it.

*3. Certain properties of abortion explain or caused my (Robert's) negative desire-like attitude towards it.

*4. John believes that certain properties of abortion explain or caused my (Robert's) negative desire-like attitude towards it.

The problem is that both *1 and *2 are beliefs that Robert simply cannot express when uttering 'John believes that abortion is wrong' because they contain indexical references to John, while *3 is a belief that Robert need not have and *4 does not correctly describe what John believes.

The reason why the above proposal has less of a problem with moral disagreement is that it conceives of the attitude expressed by 'abortion is wrong' as relating to abortion rather than to the descriptive properties which cause or explain the attitude. Once we grant that this is a psychologically plausible view, we escape the problem that John and Peter, when uttering 'abortion is wrong' and 'abortion is not wrong', cannot be taken to disagree with each other because they might have both the same beliefs and the same attitudes: While John need not deny that there are both properties in abortion that explain why he, John, disapproves of abortion, and properties that explain why Peter approves of it, John and Peter necessarily disagree with each other insofar as John disapproves of abortion while Peter approves of it. The problem with the proposal, however, is that this general interpretation of the attitude expressed by moral sentences is incompatible with a hybrid solution to the 'Frege-Geach problem'. The solutions offered by hybrid theorists such as Ridge and Barker all work in roughly the same way. The following *modus ponens* argument may serve as an illustration:

P1 Abortion is wrong

P2 If abortion is wrong, then in-vitro fertilisation is wrong.

K In-vitro fertilisation is wrong.

According to Ridge and Barker as well as other hybrid theorists such as Boisvert and Copp, we should interpret the argument in about the following way:

P1 Abortion has property K; boo to actions with property K

P2 If abortion has property K, then in-vitro fertilisation has property K; boo to actions with property K.

K In-vitro fertilisation has property K; boo to actions with property K.

However, if we interpret the descriptive belief expressed by 'abortion is wrong' in terms of the *de dicto* belief described above, the *modus ponens* argument does not even get off the ground:

P1 Certain properties of abortion explain my (the speaker's) negative attitude towards it; boo to abortion.

P2* If certain properties of abortion explain my negative attitude towards it, then certain properties of in-vitro fertilisation explain my negative attitude towards it; boo to abortion/in-vitro fertilisation.

K Certain properties of in-vitro fertilisation explain my negative attitude towards it; boo to in-vitro fertilisation.

The proposed interpretation of the *de dicto* belief creates severe problems for rendering the conditional of the argument. For one thing, it is not clear whether we should think of the attitude expressed by the conditional as an attitude towards abortion (as in the first premise), or as an attitude towards in-vitro fertilisation (as in the conclusion), or as an attitude towards both abortion and in-vitro fertilisation. The problem with the two latter options is that the whole idea of the above *modus ponens* argument seems to be to allow for the possibility that

someone who does not yet think that in-vitro fertilisation is wrong may come to think of it in this way *in virtue of the argument*. However, this means that it should be possible to assent to the conditional without having the attitude described by 'boo to in-vitro fertilisation'. Yet, the first of the three options does not provide much help, either. If the attitude expressed by both premises were 'boo to abortion', then it would not be true to say that somebody who assents to both premises must have the attitude expressed by the conclusion, namely a negative attitude towards in-vitro fertilisation.

A second problem following from the nature of the *de dicto* belief is that we also encounter problems in trying to formulate the *belief* expressed by the conditional of the argument. As it stands, my formulation is somewhat ambiguous because the second "it" might refer either to 'abortion' or to 'in-vitro fertilisation'. However, it seems that there is no legitimate way of resolving this ambiguity. If we take the first option and formulate the belief expressed by the conditional as the belief that 'if certain properties of abortion explain my negative attitude towards abortion, then certain properties of in-vitro fertilisation explain my negative attitude towards abortion', we are left with a belief whose content seems self-contradictory. However, the same holds if we interpret the belief as the belief that 'if certain properties of abortion explain my negative attitude towards in-vitro fertilisation, then certain properties of in-vitro fertilisation explain my negative attitude towards in-vitro fertilisation'. An additional problem is that under this interpretation, the content of the belief presupposes that the speaker already has a negative attitude towards in-vitro fertilisation which presupposition, as I have already emphasised, is not warranted. It is for this latter reason that it is of no help to interpret the "it" differently in the antecedent and the consequent of the conditional, either. The belief that 'if certain properties of abortion explain my negative attitude towards in-vitro fertilisation, then certain properties of in-vitro fertilisation explain my negative attitude towards in-vitro fertilisation' is a belief which we may not legitimately

ascribe to a person uttering 'if abortion is wrong, then in-vitro fertilisation is wrong'. The reason is that the person making the utterance need not have any negative attitude towards in-vitro fertilisation in the first place in which case there would be no foundation for any conditional belief about what properties might explain the existence of this attitude.

For this reason, it would not even help us if we followed Eriksson into rejecting the common hybrid solution to the Frege-Geach problem. According to Eriksson, the expressivist can concede that in certain contexts, such as in the conditional of a moral *modus ponens* argument, moral predicates such as 'wrong' do not express any attitudes at all. Eriksson's crucial idea is that the expressivist is not thereby forced to also concede that this changes the meaning of 'wrong', which means that he might escape the problem of equivocation (see Eriksson 2009, 10f.). If this were true, the expressivist could happily accept that, on the *de dicto* interpretation, the two premises and the conclusion of a moral *modus ponens* argument express different attitudes or attitudes towards different objects. However, in order for this alternative solution to the 'Frege-Geach problem' to work, it must still be the case that the two premises and the conclusion express descriptive beliefs which together form a valid argument. Yet, the *de dicto* interpretation under consideration is unable to make good on this requirement.

IV.

The appeal to *de dicto* beliefs expressing the supervenience of the moral on the non-moral, then, does not provide a fruitful strategy for identifying an uncontroversial and non-speaker-relative descriptive content of moral sentences, especially if such a strategy is meant to preserve hybrid expressivism's additional resources for solving the 'Frege-Geach problem'. The overall conclusion of the above analysis, therefore, is that there is no easy way of

reconstructing Hare's position in order to make it meet the demands of modern hybrid expressivism. Unless the modern hybrid expressivist manages to identify some other uncontroversial, constant and non-speaker-relative descriptive content of moral sentences, he needs to embrace one of the other two options sketched above: Either, he restricts his hybrid analysis to thick moral concepts; or, he interprets the meaning of thin moral concepts in terms of a less intuitive and more sophisticated conception of meaning – thereby incurring burdens not dissimilar to the burdens of metaethical naturalism.

A final option which has been propagated by quite a few authors recently, including Michael Ridge, is to defend a version of *relational expressivism* (see Toppinen 2013; Schroeder 2013; and Ridge 2014). The crucial idea of relational expressivism is to view the mental states conventionally expressed by moral sentences neither as beliefs nor as desires nor as a combination of both beliefs *and* desires, but as *relational* states: as relations, that is, between a descriptive belief and a desire-like attitude. According to this idea, the utterance of a moral sentence presupposes, by way of conventional meaning, that the speaker has both a relevant belief and a relevant desire which are suitably related to one another. However, the meaning is not constituted by the *content* of these beliefs and desires, but by the more general relation between the two.

The obvious advantage of this approach is that there is no need to specify any non-speaker-relative content of the descriptive belief. Relational expressivists can allow that the standards employed by different speakers using moral terms may vary because they do not claim that the content of the speakers' beliefs is conveyed by the speakers' utterances. What is being conveyed is only that the speakers are in a relational state of mind with regard to the object they evaluate, and that this relational state has some kind of descriptive belief and some kind of pro- or con-attitude as its *relata*. Since this relational state can be instantiated by quite different beliefs and attitudes, the relational expressivist can do without an identification of

the descriptive content of moral utterances and hence solve the problem discussed above by avoiding it in the first place.

It should be noted, however, that this solution is not so much achieved *within* the hybrid framework but rather at the expense of it: By conceiving of the mental state expressed by moral sentences as one relational state, the relational expressivist gives up on the central idea of hybrid theories and introduces a new paradigm. Furthermore, although relational expressivism may seem like a natural advancement of hybrid expressivism, it clearly faces problems of its own and, in fact, ones that hybrid expressivism does not face. The main question to be asked here is how exactly we should conceive of relational mental states in psychological terms and whether any such states will figure in our best overall theory of human psychology. Moreover, once we allow that, for any two mental states that stand in a relevant relation with one another, we may simply postulate an additional relational mental state over and above these two mental states, it seems that we can quite easily multiply the relational states that we may attribute to a speaker uttering a moral sentence. Not only does our view of the human mind then become quite a crowded one. The question also becomes: why should we view moral sentences as expressions of a relation that holds between an ordinary descriptive belief and a pro- or con-attitude and not, for example, as expressions of a more complex relation that holds between two relational states, or between relational and non-relational states?

Now it might be argued that the relational expressivist can escape at least some of these problems by restricting his relationalist analysis to moral *utterances*: While moral utterances express relational states rather than ordinary beliefs and desires, the relational expressivist might argue, moral thinking is constituted, not by relational mental states, but by the relevant beliefs and desires themselves. The problem with this idea, however, is that expressivism is now usually understood as a theory about *both* moral language *and* moral

thought (see Schroeder 2008, 4 and 151f.). In fact, as Schroeder himself emphasises, to provide a plausible account of moral thought that coheres with the proposed account of moral language and allows for a crucial continuity between private moral judgements and public moral utterances is something that we should expect from any plausible theory of moral language. Therefore, once the relational expressivist drives a wedge between these two aspects of our moral practice and forwards a psychologistic semantics that interprets moral utterances as expressions of mental states *different* from the ones that actually constitute moral thinking, expressivism seems to lose much of its initial attraction.

The task for relational expressivists, then, is to show that their idea of a relational mental state is more than just some *ad hoc* epicycle meant to solve the problems of hybrid expressivism and that it can form an integral part of a more comprehensive theory of moral judgement that is both psychologically and semantically adequate. It seems fair to say that relational expressivists are still working on this task. Without further entering into the discussion of the existing accounts, we can at least conclude that the view that is sometimes suggested by proponents of relational expressivism, namely that the move to relationalism comes at no cost at all because relational expressivism offers all the virtues of hybrid expressivism without any of its vices, is misleading. Just as the appeal to a more sophisticated and more controversial conception of semantic meaning the appeal to relational mental states clearly imposes burdens on the expressivist that might ultimately outweigh its benefits.

This leaves us with the question of whether Smith is right in claiming that, on any plausible interpretation of moral sentences, pure expressivism collapses into hybrid expressivism. As we have seen, if we do not want the *de dicto* belief expressing the idea of supervenience to be a normative belief that demands a hybrid analysis in its own right, we must ultimately conceive of it as a belief that fails to play any systematic role when it comes to explaining how moral sentences enter into logical relations with other moral or non-moral

sentences. It is then on par with quite a few other beliefs that we could legitimately attribute to a speaker uttering a moral sentence, such as the belief that 'it is me who is talking just now' or the belief that 'somebody is listening just now'. However, if this is true, then it seems somewhat misleading to draw the conclusion that pure expressivism collapses into hybrid expressivism in virtue of the *de dicto* belief in question, given that we would hardly want to make this claim on the basis of the fact that utterances of the sentence 'abortion is wrong' can be said to conventionally express the belief that 'it is me (the speaker) who is talking just now'. As long as a metaethical position does not acknowledge the expression of any further, non-trivial beliefs with some kind of systematic importance to moral argument and moral deliberation, therefore, it seems 'pure' enough to be described as a variant of pure expressivism.

References:

Barker SJ (2000) Is value content a component of conventional implicature? Anal 60:268–279

Boisvert DR (2008) Expressive-Assertivism. Pac Philos Q 89:169–203

Boyd R (1988) How to be a moral realist, In: Sayre-McCord G (ed) Moral realism. Cornell University Press, Ithaca, pp 181–228

Brink DO (1989) Moral realism and the foundations of ethics. Cambridge University Press, Cambridge

Eriksson J (2009) Homage to Hare: Ecumenism and the Frege-Geach Problem. Ethics 120:8–35

Eriksson J (2014) Elaborating expressivism: Moral judgments, desires and motivation. Ethical Theory Moral Pract 17:253–267

Eriksson J (2015a) Hybrid expressivism, In: Ridge, M, Fletcher, G (eds) Having it both ways: Hybrid theories and modern metaethics. Oxford University Press, Oxford, pp 149–170

Eriksson J (2015b) Explaining disagreement: A problem for (some) versions of hybrid expressivism. Pac Philos Q 96:39–56

Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychol Rev 108:814–834

Haidt J, Björklund F, Murphy S (2010) Moral dumbfounding: When intuition finds no reason. Unpublished manuscript.

Hare RM (1952) The language of morals. Oxford University Press, Oxford

Hare RM (1963) Freedom and reason. Clarendon Press, Oxford

Hare RM (1981) Moral thinking. Its levels, method and point. Oxford University Press, Oxford

Hare RM (1999) Sorting out ethics. Oxford University Press, Oxford

Ridge M (2006) Ecumenical expressivism: Finessing Frege. Ethics 116:302–336

Ridge M (2014) Impassioned belief. Oxford University Press, Oxford

Schroeder M (2008) Being for. Evaluating the semantic program of expressivism. Oxford University Press, New York

Schroeder M (2009) Hybrid expressivism: Virtues and vices. Ethics 119:257–309

Schroeder M (2013) Tempered expressivism, In: Shafer-Landau R (ed) Oxford studies in metaethics, volume 8. Oxford University Press, Oxford, pp 283–314

Smith M (2001) Some not-much discussed problems for non-cognitivism in ethics. Ratio 4:93–115

Stevenson CL (1944) Ethics and language. Yale University Press, New Haven

Stevenson CL (1963a) The nature of ethical disagreement, In: Stevenson CL Facts and values. Studies in ethical analysis. Yale University Press, New Haven, pp 1–9

Stevenson CL (1963b) Relativism and nonrelativism in the theory of value, In: Stevenson CL Facts and values. Studies in ethical analysis. Yale University Press, New Haven, pp 71–93

Strandberg C (2015) Options for hybrid expressivism. Ethical Theory Moral Pract 18:91–111

Toppinen T (2013) Believing in expressivism, In: Shafer-Landau R (ed) Oxford studies in metaethics, volume 8. Oxford University Press, Oxford, pp 252–282

Urmson, JO (1968) The emotive theory of ethics. Hutchinson, London