

## Do Expressivists have an Attitude Problem?\*

This paper is published in *Ethics* published by the University of Chicago Press. This is a postprint; that is, it is the full and correct text. The definitive version is available from the University of Chicago Press. The full citation is: Köhler, Sebastian 2013. Do Expressivists have an Attitude Problem?. In *Ethics* 123: 479-507.

### 0. Introduction

One objection raised for meta-ethical expressivism is the *Moral Attitude Problem*.<sup>1</sup> This is the objection that (i) expressivists must give an account of the nature of the attitudes constituting moral thinking, (ii) they can do so by saying either that they are *sui generis* attitudes or that they are reducible to attitudes fully describable in non-moral terms, but (iii) that the first option is uninformative and incompatible with naturalism and the second incompatible with our intuitions about moral thinking. In this paper I argue for two conclusions regarding this apparent dilemma. First, that expressivists can and should address this problem by drawing on strategies meta-ethical cognitivists use to address a problem, whose analogy to the Moral Attitude Problem has not been appreciated in meta-ethics. Second, that piggybacking on these strategies not only makes the second horn of the apparent dilemma viable for expressivists but—somewhat surprisingly—also clears the way for expressivists to take its first horn, holding that moral thinking consists in *sui generis* attitudes.

I proceed as follows: Section one clarifies what the Moral Attitude Problem is. In section two I argue that cognitivism faces an analogous and equally pressing problem, explain what

strategies cognitivists have developed to address it, and argue that expressivists can use the same strategies to address the Moral Attitude Problem. Moreover, I argue, these strategies not only allow the more typical expressivist approach—on which moral thinking is reducible to attitudes fully describable in non-moral terms—to address the Moral Attitude Problem. They also reveal how we might develop a novel expressivist approach on which moral thinking consists in *sui generis* attitudes. In section three, I then sketch just enough of one such non-standard expressivist position to provide at least some license for optimism that adopting the dilemma's first horn is *in principle* feasible.

### 1. Expressivism and the Moral Attitude Problem

The Moral Attitude Problem is an argument directed against meta-ethical expressivism. As I am understanding it here, *expressivism* consists in two theses: that the meaning of moral sentences is to be explained in terms of the mental states they express, and that the mental states constituting moral thinking and descriptive thinking are interestingly *different*. Descriptive thinking, say expressivists, consists in *beliefs* which are, using a familiar metaphor, mental states with mind-to-world direction of fit. Moral thinking, on the other hand, consists in practical mental states with world-to-mind direction of fit. I will call these states, for convenience's sake, *conative attitudes*.<sup>2</sup>

Expressivism's dialectical rival in the debate in which the Moral Attitude Problem is proposed is cognitivism.<sup>3</sup> According to *cognitivism*, the mental states constituting moral thinking and those constituting descriptive thinking are both beliefs. The Moral Attitude Problem is, consequently, an objection to expressivism's second thesis. It has, in varying forms, been pressed

by several philosophers.<sup>4</sup> Judith Jarvis Thomson, for example, presents the objection in the following remarks:

“[T]he [expressivist] needs to avail himself of a special kind of approval and disapproval: these have to be moral approval and moral disapproval. For presumably he does not wish to say that believing Alice ought to do a thing is having toward her doing it the same attitude of approval that I have toward the sound of her splendid new violin. The problem I point to here is a familiar one; [...] It pays to stress its seriousness, however. For if there is no way of saying what the attitude of moral approval consists in other than by saying that having it toward a thing *just is* believing a favorable moral sentence, then this [expressivist] thesis is uninformative [...].”<sup>5</sup>

Similarly, Michael Smith writes:

“[Expressivists] insist that it is analytic that when people sincerely make normative claims they thereby express desires or aversions. But which desires and aversions [...], and what special feature do they possess that makes them especially suitable for expression in a normative claim? How do they differ from mere desires and aversions that aren’t suitable for such expression? The difficulty involved in supposing that there are any such desires or aversions at all cannot be overestimated. It is, after all, agreed on all sides that the psychological state we express when we make a normative claim has many of the functional features of belief. The difficulty, to anticipate, is that it is hard to see how desires or aversions could have exactly these functional features [...].”<sup>6</sup>

The argument can be set up in the form of a dilemma in three steps.

The first is to note that the conative attitudes supposed to constitute moral thinking do significant explanatory work on the expressivist account: both the meaning of moral sentences and the nature of moral thinking are explained in terms of these attitudes. Without an account of the nature of these states, which shows how they do this explanatory work, expressivism would consequently be objectionably incomplete. So, expressivists must provide such an account. Meeting this challenge is especially important because it is not obvious that moral thinking

consists in conative attitudes: First, as Crispin Wright has pointed out, the phenomenology of moral thinking is not decisively like that of familiar conative attitudes.<sup>7</sup> Second, as Michael Smith has noted, moral thinking seems to have the functional role of *both* beliefs *and* conative attitudes.<sup>8</sup> Smith's point is particularly crucial for expressivists, as they deny that mental states could have both functional roles, and consequently need to explain this appearance differently.

The second step is to note that expressivists have only two routes to give an account of the nature of the moral attitude. First, they can hold that to account for moral thinking, we need to introduce an *additional type* of conative attitude into our theory of mind, over and above those already part of an account of non-moral thinking such as emotions, desires or intentions. Following this approach, expressivists will hold that moral thinking consists (characteristically) in a distinctive, *sui generis* type of conative attitude which cannot be (completely) reduced to attitudes already part of an account of non-moral thinking.

The second route is to hold that to account for moral thinking, we do *not* need to introduce an additional type of conative attitude into our theory of mind over and above those already part of an account of non-moral thinking. Instead, moral thinking can be completely accounted for in terms of attitudes that are already part of an account of non-moral thinking. Of course, since the relevant conative attitudes will *in themselves* not be moral attitudes, expressivists taking this approach must identify some feature that distinguishes moral instances of these attitudes from non-moral instances. This could, e.g., be that they play a distinctive functional role or result from distinctive causal processes.

Finally, we can note three conditions expressivist accounts of the moral attitude must satisfy. First,

**Informativity:** The expressivist account of the moral attitude must be informative.

Imposing Informativity is justified by those considerations that raise the challenge to provide an account of the moral attitude in the first place. Since those conative attitudes that supposedly constitute moral thinking pull significant explanatory weight for expressivists, they must give an *informative* account of these attitudes that pulls this weight. Without such an account, expressivists have asserted, but not shown that the relevant features of moral language and thought can be accounted for in terms of conative attitudes.

Second,

**Naturalism:** The expressivist account of the moral attitude must be compatible with a naturalistic world-view.<sup>9</sup>

Although expressivism is not necessarily tied to a naturalistic world-view, two considerations support imposing Naturalism on expressivist accounts of moral thinking: First, all expressivists in the meta-ethical debate endorse a naturalistic world-view, so Naturalism applies to all actual expressivists challenged with the Moral Attitude Problem. Second, expressivists often argue that a strong argument for expressivism is its ability to fit moral judgements into a naturalistic framework. To reject Naturalism would deprive expressivists of this argument for their view.

Third,

**Plausibility:** The expressivist account of the moral attitude must, at least in clear cases, agree with competent speakers' intuitions about what is, and what is not, an instance of moral thinking.

Plausibility is justified as follows: Expressivists give a systematic and informative analysis of a phenomenon of which we already have some (rough and mostly implicit) understanding. This

understanding is exemplified by competent speakers' intuitions about possible cases. Since these intuitions are our best epistemic guide to the phenomenon in question, agreement with these intuitions is required just to avoid changing the subject: if the expressivist analysis disagreed substantially with competent speakers' intuitions about moral thinking, it would become plausible that the analysis is not of moral thinking, but of something else. Therefore, for the expressivist analysis to be plausible as an analysis of moral thinking, it must agree with competent speakers' intuitions, at least in clear cases.

With these remarks in place, we can now present the dilemma that constitutes the Moral Attitude Problem:

If, on the first horn of the dilemma, expressivists claim that the moral attitude is a *sui generis* attitude not reducible to conative attitudes already part of an account of non-moral thinking, they will fail to satisfy Informativity and Naturalism: What can they say about the moral attitude on this approach? They can say that it is expressed by moral language, constitutes moral thinking, and that its functional role seems like that of both beliefs and conative attitudes. However, neither of these claims is informative, nor does this kind of account satisfactorily explain any of the features that are, from the perspective of a naturalistic world-view, supposed to be problematic about moral thinking. For example, it merely presupposes, without explanation, that there could be attitudes with the functional role of both beliefs and conative attitudes. Nor does it explain *what* moral thinking is. As an explanatory account of how moral language and thought fit into a naturalistic world-view this seems unsatisfactory and to violate both Informativity and Naturalism.

If, on the second horn of the dilemma, expressivists say that the moral attitude can be

analysed in terms of attitudes already part of our theory of non-moral thinking they will—according to proponents of the Moral Attitude Problem—violate Plausibility. Ironically, there seems to be some support for this thesis in an argumentative device expressivists use to argue for their position, namely the Open Question Argument.<sup>10</sup>

The *Open Question Argument* is an argumentative strategy originally introduced by G.E. Moore that is supposed to establish that a certain kind of concept F can *in principle* not be analysed in terms of another kind of concept G. This is done by pointing out that an analysis of F in terms of G is successful, only if for some sub-set G\* of the set of concepts belonging to G, competent speakers regard the question 'x is G\*, but is x also F?' to be conceptually closed (they have to think that for some sub-set G\* one could not coherently imagine some x which is G\*, but not F). The next step then consists in strengthening the intuition that it is plausible that for any sub-set G\*, competent speakers will regard this question to be conceptually open. Absent further explanation why these intuitions are mistaken or another way of justifying this divergence, analyses of F in terms of G therefore fail.

How might the Open Question Argument be used to press the second horn of the dilemma? Remember that on this horn, expressivists want to analyse moral thinking in terms of some conative attitude C that is already part of a theory of non-moral thinking and has some non-moral naturalistic feature N. Such an analysis succeeds only if competent speakers hold the question “S has attitude C with feature N, but is she engaged in moral thinking?” to be conceptually closed: they must hold that we could not coherently imagine someone who has attitude C with features N, but who is not engaged in moral thinking. However, it is plausible that for any conative attitude C already part of our account of non-moral thinking, when combined with some non-

moral naturalistic feature N, competent speakers will hold this question to be conceptually open. Absent further explanation why this intuition is mistaken or another way to justify this divergence, expressivist analyses of moral thinking in terms of conative attitudes already part of a theory of non-moral thinking will therefore fail. And so, this second route to an account of the moral attitude fails to satisfy Plausibility.

In sum, whichever route they take, expressivist accounts of the nature of moral thinking seem to violate at least one of the conditions such accounts should satisfy. As it stands, I think we shouldn't see this dilemma as a knock-down argument against expressivism, but rather a challenge to escape the dilemma. Expressivists have generally tended to pursue the second horn by giving analyses of moral thinking in terms of conative attitudes already part of an account of non-moral thinking that are supposed to agree with competent speakers' intuitions. However, so far the prospects of such accounts look unpromising: They have difficulties meeting Plausibility, as we see by considering the proposals of two prominent defenders of expressivism.

The first example is Simon Blackburn's account. According to Blackburn moral attitudes are complex *combinations* of conative attitudes already part of a theory of non-moral thinking: attitude C is a moral attitude if the person who has C is (a) disposed to approve of those sharing C and disapprove of those not sharing C and (b) disposed to approve of those who share approval of sharing C and disapproval of those not sharing C and to disapprove of those who do not share approval of sharing C and disapproval of those not sharing C.<sup>11</sup> On this account, the named conative attitudes are individually attitudes describable in non-moral terms. They become moral attitudes, only if combined in the relevant ways.

The problem with Blackburn's account is, however, that the question is open whether

someone with the above complex combination of conative attitudes is engaged in moral thinking. Specifically, his account cannot distinguish moral thinking from other modes of *normative* thinking: for example, someone who holds (a) that those who share her aesthetic judgements have good taste and those who do not have bad taste and that (b) those who deny that such people have bad taste or tolerate such bad taste have bad taste themselves could plausibly be interpreted as holding the web of attitudes Blackburn proposes as characterizing moral attitudes, without it being plausible that she holds *moral* attitudes. If this is true, Blackburn's account fails to satisfy Plausibility.<sup>12</sup>

To take the second example, Allan Gibbard suggests that moral thinking, specifically thinking that  $\Phi$ -ing is wrong, is thinking that it is rational to feel guilty for  $\Phi$ -ing and be angry with those who  $\Phi$ .<sup>13</sup> “Rational” is thereby supposed to capture the concept in play “when we talk about 'what it makes sense' to do or to believe”<sup>14</sup> and “thinking something to be rational” is to be understood in terms of the conative attitude of acceptance of norms that permit it, an attitude supposed to be already part of an account of non-moral thinking.

Gibbard's account, however, also faces the problem that the question seems open whether someone who thinks that it is rational for her to feel guilty for  $\Phi$ -ing thinks that  $\Phi$ -ing is wrong. For example, imagine someone going to a conference. At this conference she will meet old friends whom she knows will want to spend time with her, but she will also meet other people important for her career. Suppose that this person decides against spending time with her old friends in order to have more time to socialize with the other group. It seems possible for her to think that it is rational for her to feel guilty and for her friends to be angry with her for this choice, but nevertheless not to think that she is doing something morally wrong. If this is true,

Gibbard's account fails to satisfy Plausibility.<sup>15</sup>

As we see, both Blackburn's and Gibbard's analyses seem to fail to satisfy Plausibility. Now, although some analysis of this kind might be able to meet that condition, I think that approaching the Moral Attitude Problem the way expressivists have tended to do in the past is wrongheaded: In my view, expressivists have not recognized that they can use resources naturalistically minded cognitivists made available *when cognitivists faced the same dilemma*. Let me explain what I have in mind and then explain how expressivists can benefit from lessons cognitivists learned.

## **2. Cognitivism and the Moral Attitude Problem**

Anyone working in meta-ethics should be familiar with an argument that is structurally similar to the Moral Attitude Problem, but actually applies to their opponents, at least against those cognitivists who subscribe to a naturalistic world-view. The problem of specifying the nature of the mental state constituting moral thinking is clearly not restricted to expressivists: it also applies to cognitivists. Moral thinking has many features which, *prima facie*, do not fit with the features beliefs exemplify. After all, that moral thinking displays the functional role of *both* beliefs *and* conative attitudes, should be problematic for *both* expressivists *and* cognitivists. Of course, cognitivists normally try to explain the nature of moral beliefs not in terms of their *psychological* features, but in terms of their *content*: they try to explain the nature of moral beliefs in terms of the nature of the states of affairs they represent. Nevertheless, to provide a successful explanatory account of the nature of moral thinking, cognitivists must also give an account of moral thinking, or, more specifically, of the content that explains the nature of moral

thinking.

Like expressivists, cognitivists can proceed in two ways: First, cognitivists can hold that accounting for the content of moral thinking requires introducing an additional type of fact into our ontology over and above those that are already the objects of non-moral descriptive thinking. Following this approach, cognitivists will hold that moral facts consist (characteristically) in a distinctive, *sui generis* type of fact, which cannot be (completely) reduced to those types that are the objects of non-moral descriptive thinking. The second route for cognitivists is to hold that accounting for the content of moral thinking does not require introducing an additional type of fact into our ontology over and above those that are already the objects of non-moral descriptive thinking. Instead, the content of moral thinking can be completely accounted for in terms of facts that are the objects of non-moral descriptive thinking.

Furthermore, like expressivist accounts of moral thinking, cognitivist accounts must satisfy three conditions: First,

**Informativity:** The cognitivist account of the content of moral thinking must be informative.

The justification for Informativity also follows from the considerations that make it plausible that cognitivists (like expressivists) must provide an account of moral thinking in the first place. Without such an account, cognitivists have asserted, but not shown that we can account for the features of moral language and thought in terms of beliefs with certain contents.

Second,

**Naturalism:** The cognitivist account of the content of moral thinking must be compatible with a naturalistic world-view.

Of course, cognitivists are even less committed to a naturalistic world-view than expressivists. However, imposing Naturalism on cognitivist accounts of moral thinking can be justified: One of the greatest challenges of contemporary meta-ethics is to explain how morality fits into a naturalistic framework. A view that cannot accommodate moral thinking within a naturalistic framework will consequently be at a dialectical disadvantage to one that can. So, cognitivists should avoid rejecting Naturalism.

Third,

**Plausibility:** The cognitivist account of the content of moral thinking must, at least in clear cases, agree with competent speakers' intuitions about what is and what is not an instance of a moral fact.

The justification for imposing Plausibility on cognitivist accounts is the same as the justification for imposing it on expressivist accounts: Cognitivism gives a systematic and informative analysis of a phenomenon we already have some (rough and mostly implicit) understanding of, an understanding exemplified by competent speakers' intuitions about possible cases. Since these intuitions are our best epistemic guide to this phenomenon, agreement with these intuitions is required just to avoid changing the subject: if the cognitivist analysis were to disagree substantially with competent speakers' intuitions about the content of moral thinking, it would become plausible that the analysis is not one of the content of moral thinking, but of something else. So, the analysis should be in agreement with competent speakers' intuitions, at least in clear cases.

With these remarks in place, however, we can now generate the following dilemma:

If, on the first horn of the dilemma, cognitivists claim that the content of moral thinking

consists in irreducible, *sui generis* moral facts, they will be unable to satisfy Informativity and Naturalism: What can someone taking this route say about the content of moral thinking? She can say that it is the content of moral language and thought and that because of this content moral thinking has a functional role that seems to be like that of both beliefs and conative attitudes. However, as with the expressivist account, these claims are neither informative nor do they satisfactorily explain any of the features of moral thinking supposed to be problematic from a naturalistic perspective. For example, it merely presupposes, but does not explain that there are facts which provide beliefs about them with the functional role of both beliefs and conative attitudes, nor does it explain *what* moral thinking is. As an explanatory account of how moral language and thought fit into a naturalistic world-view this is unsatisfactory and violates Informativity and Naturalism.

On the second horn of the dilemma, if cognitivists take the second approach, they will be unable to satisfy Plausibility. The Open Question Argument can again be used to support this: Cognitivists who take the second approach want to analyse the content M of instances of moral thinking in terms of the content N of some mode of non-moral descriptive thinking. Such an analysis succeeds, only if competent speakers hold the question “x is N, but is x M?” to be conceptually closed. But, it is plausible that for any content N of some mode of non-moral descriptive thinking, competent speakers will hold this question to be *open*. Absent further explanation of this divergence, cognitivist analyses of the content of moral thinking in terms of the content of some mode of non-moral descriptive thinking will, therefore, fail. The second route will not satisfy Plausibility.

Now, anyone familiar with the history of meta-ethics knows that this dilemma has haunted

naturalistic cognitivism for a large part of the 20<sup>th</sup> century—since Ayer's *Language, Truth and Logic*. Naturalistically minded cognitivists tried to take the dilemma on its second horn by providing reductive accounts of the content of moral thinking in terms of the content of certain modes of non-moral descriptive thinking supposed to somehow escape the Open Question Argument, but such accounts were long thought to fail. In contemporary debate, however, the second horn of the dilemma is thought much less problematic for cognitivists due to certain strategies cognitivists developed during the last quarter of the 20<sup>th</sup> century.<sup>16</sup> Two such approaches are particularly worth mentioning:

First, consider the approach of some of the “Cornell Realists” (I will call this the “Cornell approach”).<sup>17</sup> According to this approach, we must take into account that two concepts  $F_1$  and  $F_2$  can refer to the same property  $F_P$  not only because there is a reductive analysis of one in terms of the other, but also as a matter of synthetic *a posteriori* fact. If two concepts refer to the same property in the second way, however, this can be discovered *only* through empirical inquiry and *not* solely through competence with the concepts in question. Consequently, competent speakers will hold the question “ $x$  is  $F_1$  but is it  $F_2$ ?” to be conceptually open, although, as a matter of metaphysical fact, it is not. The prime examples are, of course, the concepts WATER and  $H_2O$ : although WATER and  $H_2O$  refer to the same thing this cannot be grasped solely by being competent with these concepts. Instead it was an *empirical* discovery that WATER and  $H_2O$  refer to the same thing. At least prior to this discovery, competent speakers would, therefore, have held the question “This is water, but is it  $H_2O$ ?” to be open, even though the two concepts, as a matter of empirical fact, refer to the same thing.

According to the Cornell approach, the same can hold for moral concepts: It is possible

that the properties moral concepts *refer* to are identical with properties certain modes of non-moral descriptive thinking refer to, but that this is a synthetic *a posteriori* fact, *not* because there is a reductive analysis of moral concepts in terms of non-moral descriptive concepts. In this case, competent speakers will hold “x has non-moral property N, but does it have moral property M?” to be open, although as a matter of metaphysical fact it will not be. If this is true, however, the second horn of the dilemma is not damaging for cognitivists. They can hold that the content of moral thinking is identical to facts already familiar as the content of certain modes of non-moral descriptive thinking, but that competent speakers cannot determine this merely through reflection because it is an *a posteriori* matter. Which means, of course, that the cognitivist's proposal regarding the content of moral thinking cannot be tested by using the Open Question Argument. It must be tested by some suitable *a posteriori* means.

The second approach to evade the dilemma for cognitivists was introduced by philosophers of the “Canberra-Plan” (I will call this the “Canberra approach”).<sup>18</sup> According to this approach there can be a reductive analysis of some concept  $F_1$  in terms of another concept  $F_2$  the truth of which is *non-obvious*. On this account, any conceptual analysis of  $F_1$  will first aim at giving a characterisation of that x that realizes  $F_1$  in purely non- $F_1$  terms (where x is e.g. a *property* if  $F_1$  is a predicative concept). We do this by observing that our concepts are plausibly seen as being characterized by our dispositions to apply them and that these dispositions should, consequently, reveal what we would, and would not, take to realize the relevant concepts. So, we start an analysis by testing how competent speakers apply  $F_1$  to real and hypothetical cases, trying to determine that set of conditions characteristic of the application of  $F_1$ . Call these conditions the “platitudes” that characterize the realizer of  $F_1$  and the set of platitudes central to our uses of  $F_1$

our “folk-theory of  $F_1$ ”. However, since how competent speakers use concepts might to some degree be inconsistent or rest on false assumptions, we should not aim through conceptual analysis to capture *every* way competent speakers apply  $F_1$ . Rather, we should aim to find an analysis that does the job of  $F_1$ , but is free from inconsistency, mistaken assumptions, etc. So, not only will we explicate the platitudes that characterize competent speakers' use of  $F_1$ , but also try to fit these platitudes into a maximally coherent, systematic and comprehensive framework. Call the set of platitudes that characterize that  $x$  which realizes  $F_1$  arrived at in this way our “mature folk-theory of  $F_1$ ”.

Having arrived at this mature folk-theory we can transform it into a *non-circular* characterization of the realizer of  $F_1$  as follows: First, we reformulate the platitudes that constitute our mature folk-theory so that all mentionings of the realizer of  $F_1$  come out as *names* (or noun-phrases). For example, if  $F_1$  is a predicative concept, we reformulate the platitudes in property-name style. Note that depending on the complexity of  $F_1$  we might have to introduce more than one name. For example, the platitudes surrounding our concept COLOUR will make ineliminable reference to certain colours, mentionings of which need to be stripped out for the characterisation of the realizer of COLOUR to be non-circular. To do so, however, we might need to introduce names for all such colours. In what follows I will only present how the approach proceeds in cases in which we need introduce one name, but keep in mind that, *mutatis mutandis*, it will also work for more complex cases.

Once we have reformulated the platitudes, we form a conjunction out of them, which can be represented as a relational predicate true of the realizer of  $F_1$ . Call this relational predicate “ $T_F$ ” and let “ $a$ ” stand for the name we introduced to refer to the realizer of  $F_1$ . We can now formulate

what Lewis called a “postulate” of our mature folk-theory of  $F_1$ , namely

$$T_F[a]$$

The postulate characterizes the realizer of  $F_1$  in virtue of its relations to other things, namely in virtue of how it satisfies  $T_F$ . This allows us to say that the realizer of  $F_1$  is *exactly that* phenomenon characterized by  $T_F$ :  $T_F$  gives us the *functional role* of the realizer of  $F_1$ . To get a *non-circular* characterization of the realizer of  $F_1$  we now exchange the name for a free variable “ $x$ ”, which yields

$$T_F[x]$$

We then restate our theory of  $F_1$  with the following Ramsey-sentence:

$$\exists x \{T_F[x] \ \& \ (\forall x^* T_F[x^*] \text{ iff } (x = x^*))\}.$$

With the Ramsey-sentence, we have derived a description of what realizes  $F_1$  solely in non- $F_1$  terms. Indeed we can use the Ramsey-sentence to *define*  $F_1$  in non- $F_1$  terms. Most importantly, however, this description can be used to determine whether  $F_1$  is *realized* by some unique set of entities captured by a theory expressible in terms of some other set of concepts  $F_2$ . If it is, then there will be a true bi-conditional that identifies  $F_1$  with  $F_2$ . This way we will have provided a reductive analysis of  $F_1$  solely in terms of another set of concepts  $F_2$ .

However, we should *not* assume that such reductive analyses can be tested in the way presupposed by the Open Question Argument, because the truth of such reductive analyses might be *non-obvious*. First, whether some characterisation of the realizer of  $F_1$  in non- $F_1$  terms is appropriate can already be non-obvious: the dispositions that characterize our concepts need not be immediately accessible on reflection, just as the rules of grammar underlying our uses of language reveal themselves in our dispositions to use that language, but are not immediately

explicable by competent speakers. Second, such a characterisation can be non-obvious because it need not capture *every* way the concepts are used but only their *core* in a way free from inconsistency, mistaken assumptions, etc. However, such a characterisation might well appear false to competent speakers both on first sight *and* reflection, as it sometimes yields different applications than speakers' intuitions would yield. Third, the possibility of an analysis of  $F_1$  in terms of  $F_2$  might be non-obvious if  $F_2$  is part of an *empirical* theory arrived at using only *a posteriori* means. In this case, competent speakers will be unable to determine on reflection that  $F_1$  is analysable in terms of  $F_2$ , although this is in fact possible.

So, there can be two sets of concepts where one is analysable in terms of the other, but where the equivalence is *non-obvious* to competent speakers, even upon reflection. If this is true, however, the second horn of the dilemma is not damaging for cognitivists. They can hold that there is a true reductive analysis of the content of moral thinking in terms of the content of certain modes of non-moral descriptive thinking but that, for reasons similar to those given above, this is *non-obvious* for competent speakers even upon reflection and so cannot be tested using the Open Question Argument. So, there can be a true reductive analysis of moral concepts in terms of non-moral descriptive concepts, even if no such analysis passes the Open Question Argument.

As we have seen, cognitivists can deal with the second horn of their "Moral Attitude Problem" in at least two ways. For this paper's purposes we need not go deeper into these kinds of cognitivist approaches. What I want to point out instead is that, if cognitivists can use these resources to escape the second horn of their dilemma, expressivists should also be able to use them to escape the second horn of their Moral Attitude Problem. That is, I want to suggest—

somewhat surprisingly—that expressivists can use either of these approaches to argue that moral attitudes are identical to certain kinds of attitudes already part of an account of non-moral thinking. They can claim that this identity is either synthetic *a posteriori* or that it *is* conceptual, but non-obvious. So, expressivists can, for example, use the resources of the Cornell approach and claim that whatever the specific nature of the moral attitude is, is a synthetic *a posteriori* question subject to discovery in the empirical sciences, e.g. psychology and cognitive science. What philosophers do is to work with the understanding we have of moral thinking and psychology and then propose as an *empirical hypothesis*, that whatever mature empirical psychology and cognitive science will discover about the nature of moral thinking, they will also discover that moral thinking is identical with some kind of conative attitude already part of our theory of non-moral thinking. Or, expressivists can follow the Canberra approach and say that we can analyse the moral attitude in terms of certain attitudes already part of an account of non-moral thinking: They can hold that conceptual analysis allows us to derive a description of moral thinking in terms not mentioning moral thinking (which need not be obviously true) and then propose, as a *hypothesis*, that the mental states so characterized are realized by certain conative attitudes already part of the theory of non-moral thinking that will be developed by empirical psychology and cognitive science. In this case, there will be a true reductive analysis of moral thinking in terms of certain conative attitudes already part of an account of non-moral thinking, but it will be non-obvious.

I also want to suggest, however, that—in addition to the above approaches to the second horn—the resources cognitivists made available open up another route for expressivists to escape the Moral Attitude Problem: Expressivists can use those resources to treat moral attitudes as *sui*

*generis* attitudes which are not reducible to attitudes already part of an account of non-moral thinking, but nevertheless fit into a naturalistic framework. Such an approach has at least two advantages. First, it allows expressivists to evade suspicions about moral thinking being reducible to conative attitudes already part of our theory of non-moral thinking. Second, it helps capture the impression that “moral thinking is what it is and not another thing.”<sup>19</sup> To show that such approaches are feasible and provide some license for optimism regarding the chances of success along this route for expressivists, I will explain in the next section how one such account, based on the Canberra approach, might proceed.<sup>20</sup>

### 3. *Sui Generis* Moral Attitudes for Expressivists

Let me start by noting that it is not necessary for an expressivist account of moral thinking to be compatible with naturalism that it can account for moral thinking in terms of *other* mental states. Instead, it is sufficient that an informative and non-circular characterization of moral thinking can be given that allows us to single it out as a phenomenon and to show that the mental states so characterized *fit into our best naturalistic philosophical theories of psychology*. So, to vindicate that an expressivist account on which moral thinking is *sui generis* is feasible, I will explain how such an account can give the relevant kind of characterization and how it can show that moral thinking so characterized fits into our best naturalistic philosophical theories of psychology.

#### 3.1. Characterizing Moral Thinking

I take it that a characterization of moral thinking needs to provide two things: First, a

characterization of the distinctive modes of moral thinking, e.g. thinking that something is morally wrong, thinking that something is just, etc. Second, a characterization of moral thinking in general, namely of what the distinctive modes of moral thinking have in common in virtue of being modes of moral thinking. How do we give the relevant characterization? We can do so, by using exactly that style of conceptual analysis developed by the Canberra approach. Clearly we already *have* an understanding of moral thinking and its distinctive modes, although this understanding is mostly implicit and rough. This is our “folk-theory of moral thinking”. And using our folk-theory of moral thinking, we can derive a characterization of moral thinking and its distinctive modes, the *functional role* of moral thinking. This should be possible, even if moral thinking is *sui generis*. However, if we follow the Canberra approach, the functional role is all we need for the kind of characterization we are looking for. So, as long as the Canberra approach to moral thinking can be developed, we should be able to give the relevant kind of characterisation even if moral thinking is *sui generis*. To vindicate that such an approach to moral thinking can be developed, I will now flesh out the details of how the necessary steps for doing so are to be carried out.

### **3.1.1. Explicating the Platitudes**

The first crucial step in carrying out the Canberra approach to moral thinking is to explicate our implicit understanding of moral thinking and its distinctive modes, making explicit the claims that—according to our folk-theory—characterize them. These are the “platitudes” that characterize moral thinking and its distinctive modes. In explicating these platitudes, we have to bear in mind that we are explicating our *pre-theoretic* understanding of moral thinking. Since we

should abstain from thinking that our folk-understanding of moral thinking has strong implications favouring one philosophical theory over another, we should, consequently, formulate the platitudes as theory-neutral as possible, especially regarding the theories relevant for the inquiry at hand, namely cognitivism and expressivism.

Candidates for platitudes characterizing moral thinking in general would be, e.g.

- (1) For any x, judging that x is morally wrong, judging that x is just, ..., are instances of moral thinking.
- (2) Any instance of moral thinking can stand in the rational and inferential relations expressed by the logical and sentential connectives with all other mental states that can so stand.

Taking as an example judgements of the form “x is morally wrong”, candidates for platitudes that characterize *particular* modes of moral thinking would be e.g.

- (3) For any x, someone can judge x to be morally wrong, only if x is an action, intention, ..., desire.
- (4) For any x, if someone judges x to be morally wrong, then she will *ceteris paribus* be disposed to avoid x, to be angry with those who x, ..., to feel guilty for x.
- (5) For any x, the mental state of judging x to be morally wrong can stand in all the rational and inferential relations expressed by the logical and sentential connectives with all other mental states that can so stand.

To avoid problems that will become manifest shortly, I thereby suggest the following procedure for the explication of the relevant platitudes: We start with explicating platitudes

surrounding atomic instances of moral thinking, what I will call “atomic moral judgements”, and platitudes surrounding moral thinking in general, in so far those concern only atomic moral judgements. I take atomic moral judgements, thereby, to be those instances of moral thinking expressed by moral sentences that can no longer be broken down into logically less complex sentences. Examples of atomic moral judgements might be “x is morally wrong”, “x is just”, etc. Now, if we want to use the Canberra approach to give a characterization of moral thinking, our theoretical purposes require that all platitudes be formulated so that all mentionings of instances of moral thinking come out as *names* (or noun-phrases) for mental states. It does not really matter at which stage of our investigation we do this: we can do it when we first explicate the platitudes, or reformulate them later when all other theoretical work is done and we are about to form the Ramsey-sentence. For convenience's sake, however, let me say at this stage how this should be done for mentionings of moral thinking in general and of atomic moral judgements.

We can refer to moral thinking as a general phenomenon by using the name “moral thinking”. But how do we refer to atomic moral judgements using only names? We do not, of course, want to introduce a new name for any x and for any kind of atomic moral judgement one could make about x, so that e.g. we have to introduce a distinct name for judging murder to be wrong and judging stealing to be wrong. My suggestion to avoid this is the following: We treat atomic moral judgements as *relational states* between subjects (the subjects making those judgements) and objects (broadly speaking, meaning whatever the respective modes of moral thinking can conceivably be about; the relevant restriction will be determined by platitudes surrounding those modes).<sup>21</sup> And we refer to atomic moral judgements by introducing names for these relational states. We must, thereby, treat each distinctive kind of atomic moral judgement as

a distinctive relational state, introducing a distinct name for each. On this notational treatment the platitudes about atomic moral judgements (taking again the example of judgements about something being morally wrong) would look as follows:

- (5) For any x, someone can be in *the state of making a wrongness judgement* with regard to x, only if x is an action, intention, feeling or desire.
- (6) For any x, if someone is in *the state of making a wrongness judgment* with regard to x, then she will, *ceteris paribus*, be disposed to avoid x, to be angry with those who x, ..., to feel guilty for x.
- (7) For any x, *the state of making a wrongness judgment* with regard to x can stand in all the rational and inferential relations expressed by the logical and sentential connectives with all other mental states that can so stand.

It should be noted that this notational treatment is theory-neutral: expressivists can say that the states so characterized are realized by conative attitudes towards the relevant objects, while cognitivists can say that they are realized by beliefs in incomplete propositions, where incomplete propositions are functions from the relevant objects to full propositions. Another natural notational treatment, namely treating instances of moral thinking as relational states between subjects and propositions, would not be theory-neutral, so it should be avoided.

With the question resolved how to refer to atomic moral judgements using only names, we now have to address further difficulties that arise when we explicate the platitudes surrounding those judgements. According to some of those platitudes atomic moral judgements can stand in the rational and inferential relations expressed by logical and sentential connectives such as negation, disjunction, modal operators, etc. with other mental states. The following, e.g. will be

platitudes about the state of judging something to be wrong:

- (9) For any  $x$ , judging  $x$  to be wrong is inconsistent with judging  $x$  to be not wrong.
- (10) For any  $x$ , and any descriptive concept  $F$ , judging  $x$  to be wrong or  $F$  is inconsistent with judging  $x$  to be not wrong and judging  $x$  to be not  $F$ .
- (11) For any  $x$ , judging  $x$  to be wrong and unjust is inconsistent with judging  $x$  to be not wrong or judging  $x$  to be not unjust.

There will obviously be many platitudes of this kind for any kind of atomic moral judgement.<sup>22</sup> These platitudes do two important things in the context of developing a Canberra approach to moral thinking. First, they introduce mentionings of logically complex mental states into our platitudes, e.g. *judging  $x$  to be not wrong*, *judging  $x$  to be wrong or  $F$* , *judging  $x$  to be wrong and unjust*, etc. Roughly speaking, logically complex mental states are states expressed by complex sentences formed by means of logical and other sentential connectives. Any such mental state stands in certain inferential and rational relations to other mental states, namely those mental states expressed by the parts of the sentence that expresses the logically complex mental state. Call mental states expressed by the parts of sentences that express logically complex mental states the “parts” of the logically complex states.<sup>23</sup> The second important thing such platitudes do is to reveal that for any *kind* of logically complex thought (thoughts expressed by negation, thoughts expressed by disjunction, etc.) there can be logically complex mental states of that kind of which atomic moral judgements are parts.

These two features of such platitudes are important in our context, because they generate two complications for the kind of characterization of moral thinking the Canberra approach aims

at:<sup>24</sup> The first complication arises because we can use the Canberra approach to give a non-circular characterization of moral thinking only if we can substitute *any* reference to instances of moral thinking in our final characterisation for a variable. This is achieved by using names to refer to instances of moral thinking. Mentionings of logically complex mental states with atomic moral judgements as parts, however, make reference to instances of moral thinking that must be stripped out of the final characterization for the characterization to be non-circular, namely to the atomic moral judgements that form their parts. For example, mentioning the logically complex state “Thinking x to be not wrong” makes reference to the atomic moral judgement “Thinking x to be wrong” in a way that cannot appear in the final characterization of moral thinking, if it is to be non-circular. And the crucial question is how to strip away such references. It would be *very* bad if we could only do this by removing any reference to logically complex mental states *themselves*: That would require introducing a name for every possible logically complex thought that has an atomic moral judgement as its part, making the number of names required for our characterisation of moral thinking grow out of hand. So, we need a different way to address this issue. The second complication is that some logically complex mental states will *themselves* be instances of moral thinking. For example, the thought “If someone lies, then they did something morally wrong” seems to be a moral thought. So, some logically complex mental states will be among the things that a theory of moral thinking needs to characterize.

How do we deal with these two complications? Before I explain how this is done, let me first make some crucial preliminary remarks about the *role* of logically complex thoughts in our theory of moral thinking. It is important to note that while some logically complex mental states will be characterized by a theory of moral thinking, certain logically complex mental states (e.g.

logically complex *non-moral* thoughts), as well as the general nature of logically complex thoughts will play a role within our theory of moral thinking *only* as things *characterizing* moral thinking. This has an important consequence, namely that our characterisation of moral thinking will successfully *pick something out*, only if we have a robust theory of logically complex mental states. After all, functional characterisations identify phenomena in virtue of their relations to other things. So, without knowing what these other things are, we cannot identify the phenomena we are after.

Now, it is unlikely, of course, that cognitivists and expressivists will subscribe to the same theory of logically complex mental states. Note, however, that in giving the *characterisation* of moral thinking we do not need to presuppose such a robust theory of logically complex mental states. After all, even without a *robust* theory of logically complex thoughts, we should be sufficiently aware of the connections between moral thinking and logically complex thoughts and of the characteristics of logically complex thoughts relevant in the context of giving an account of moral thinking. We only need to keep in mind two things when using logically complex mental states in giving a characterization of moral thinking: First that we need to do so in a way compatible with all plausible theories of the nature of logically complex thoughts. Second that our characterization of moral thinking has to be supplemented with a robust theory of logically complex mental states before we can determine what states realize our theory of moral thinking. With these remarks out of the way, let me now explain how to handle the two complications with logically complex thoughts, starting with the first.

The first thing needed to handle the first complication is a theory-neutral characterisation of logically complex mental states. Let us call logically complex mental states “commitments”.

“Commitment” is thereby to be read as weakly as possible, as any mental state  $C$  being in which rationally commits someone to be (or not be) in a set of other states  $S_1, S_2, \dots, S_n$  in the sense that being in  $C$  and failing (or continuing) to be in  $S_1, S_2, \dots, S_n$  when one could (and perhaps believes one could) would manifest irrationality. There will be a distinct kind of commitment for each distinct logical or sentential operator. This characterisation should be compatible with any plausible approach to the nature of logically complex mental states.

We can now use this notion of a commitment to treat logically complex mental states as relational states relating subjects and mental states (namely whatever states are the parts of the relevant logically complex thought). Note that this treatment is neutral between different approaches to the nature of logically complex mental states: cognitivists can say that commitments are realized by beliefs in incomplete propositions (where here these are functions from propositions of the parts to the relevant logically complex propositions), but commitments understood along these lines should also be realizable by whatever states expressivists identify as the thoughts expressed by logically complex sentences.

Now, the important thing about this treatment of logically complex states for developing a characterisation of moral thinking using the Canberra approach is that it allows us to treat instances of moral thinking as related to commitments when we formulate the platitudes about moral thinking that involve logically complex mental states. This will enable us to strip out all mentionings of atomic moral judgements in those platitudes and substitute them for variables, without having to strip out mentionings of logically complex mental states themselves. For example, the platitudes about how atomic moral judgements can stand in rational and inferential relations I gave above can now be formulated as follows:

- (12) For any x, being in a state of making a wrongness judgement with regard to x is inconsistent with being in a negation commitment with regard to the state of making a wrongness judgement with regard to x.
- (13) For any x, and any state of descriptive thinking F, being in a disjunction commitment with regard to the state of making a wrongness judgement with regard to x and the state F is inconsistent with being in a negation commitment with regard to the state of making a wrongness judgement with regard to x and a negation commitment with regard to the state F.
- (14) For any x, being in a conjunction commitment with regard to the state of making a wrongness judgement with regard to x and the state of making an unjustness judgement with regard to x is inconsistent with being in a negation commitment with regard to the state of making a wrongness judgement with regard to x or being in a negation commitment with regard to the state of making an unjustness judgement with regard to x.

And as we can see, these formulations allow us to strip out all mentionings of atomic moral judgements and substitute them for variables, without having to strip out mentionings of logically complex mental states themselves. This deals with the first complication.

How do we deal with the second complication, namely, that some commitments are instances of moral thinking? First we need a notational treatment that allows to separate instances of a kind of logically complex thought that are instances of moral thinking from those which are not.<sup>25</sup> Let us call logically complex *moral* thoughts “moral commitments” and logically complex *non-moral* thoughts “non-moral commitments”. So, we will have, e.g., *moral*

and *non-moral* negation commitments, *moral* and *non-moral* conjunction commitments, etc. Moral commitments—like commitments in general—can thereby be treated as relational states between subjects and other mental states. Note that this notational treatment is *not* supposed to suggest that the logical and sentential connectives are ambiguous: it only serves, for our theoretical purposes, to identify and separate a subclass of those states we have called “commitments”. Indeed, there will be platitudes about moral commitments that they share those features with their non-moral counterparts necessary to make them instances of the same kind of logically complex thought.

With this notational treatment in place, we can now, in our characterization, single out those commitments that need to be characterized by our theory of moral thinking, namely the moral commitments, and start collecting the platitudes surrounding those commitments. Among these will be platitudes regarding which features distinguish moral commitments from their non-moral counterparts, platitudes about the relations between moral commitments and other kinds of mental states, general platitudes about the nature of logically complex thoughts and the characteristics moral commitments have in virtue of being such thoughts, and so on. Because moral commitments, like all commitments, are treated as relational states, we should thereby have no difficulties to refer to moral commitments by introducing names for these relational states, which can later be substituted for variables. This deals with the second complication. Given that we now know how the platitudes about moral thinking are to be explicated, let me explain how a Canberra approach to moral thinking proceeds from here.

### **3.1.2. Deriving our Mature Folk-Theory of Moral Thinking**

Once we have collected all the platitudes surrounding atomic moral judgements, moral commitments and what characterizes any of these modes as instances of moral thinking in general,—our “folk-theory of moral thinking”—the next step of a Canberra approach to providing a characterization of moral thinking is to subject these platitudes to our best epistemological methods, fitting them into a coherent and systematic framework and working out any principles underlying and general connections between them. This will enable us to give a more specific and systematic characterization of moral thinking and its distinctive modes, help to eliminate errors underlying our assumptions about these phenomena and enable us to derive claims about them we might not have been aware of when we started the analysis. Subjecting our implicit understanding to this treatment makes sense: Since we regard our implicit understanding as an epistemic device to get at the phenomenon in question, applying our best epistemological methods to our implicit understanding of the relevant phenomenon will provide a much *deeper* grasp and understanding of that phenomenon and help to prevent errors we would otherwise have made.

The coherent and systematic framework of platitudes derived this way is our “mature folk-theory of moral thinking”. With this framework in hand we can now give the characterisation of moral thinking we are looking for.

### **3.1.3. Giving a Non-Circular and Informative Characterisation**

The first thing to do is to form a conjunction out of the platitudes that form our mature folk-theory. This conjunction can be represented as a relational predicate that is true of moral thinking and its various modes. Let us call this relational predicate “ $T_M$ ” and let  $m_1, m_2, \dots, m_n$

stand for the names we introduced for moral thinking and its various modes. We can now formulate the postulate of our mature folk-theory of moral thinking, namely

$$T_M[m_1, m_2, \dots, m_n]$$

The postulate characterizes moral thinking and its modes in virtue of their relations to each other and other things, namely in virtue of how they satisfy  $T_M$ . This allows us to say that moral thinking is *exactly that* phenomenon characterized by  $T_M$ :  $T_M$  gives us the functional role of moral thinking. To get a *non-circular* characterization of moral thinking we now exchange all the names for mental states with free variables, which yields

$$T_M[x, y, \dots, n]$$

We then restate our theory of moral thinking with the following Ramsey-sentence:

$$\exists x \exists y \dots \exists n \{T_M[x, y, \dots, n] \& (\forall x^*, \forall y^*, \dots, \forall n^* T_M[x^*, y^*, \dots, n^*] \text{ iff } (x = x^*, y = y^*, \dots, n = n^*))\}.$$

With the transformation of our theory of moral thinking into this Ramsey-sentence, there will now be no mentioning of moral thinking or its distinctive modes in our theory of moral thinking. Indeed we can *define* moral thinking (or any of its modes) with the Ramsey-sentence. The characterization given by this Ramsey-sentence or any definition derived from it will thereby be non-circular: we define moral thinking and its distinctive modes in terms that do not mention moral thinking or any instance of it.

Will this characterization be informative? I think it is very plausible that it will be: Obviously generating the relevant characterization of moral thinking requires substantial philosophical work. Coincidentally, I think this is a project meta-ethicists are already partially engaged in, namely identifying the features of moral thinking and its distinctive modes.

However, the kind of philosophical work required is very insightful: First, explicating our implicit understanding of some phenomenon often itself yields surprising results about (and grants deep insights into) the phenomenon in question. The history of philosophy is replete with examples: Hume's discovery of the problem of induction plausibly resulted from his explication of our implicit understanding of causality, and Plato's discovery of the Euthyphro dilemma seems to have derived from an explication of our implicit understanding of moral considerations. Second, if we subject our implicit understanding to an exhaustive process of inquiry using our best epistemic methods, this will likely yield a far better understanding of the phenomena in question than would have been obtained through mere reliance on implicit understanding. Again, the history of philosophy supports this: consider, for example, the growth in our understanding of knowledge that the post-Gettier debate produced, or how much our understanding of the nature of mental states has improved since the days of behaviourism. Such considerations strongly support that the theory we will have arrived at will be informative in any reasonable sense of the word. And so, it seems possible to give a non-circular and informative characterization of moral thinking that allows us to single it out as a phenomenon, even if one thinks that moral thinking is constituted by a distinctive, *sui generis* type of mental state.

### **3.2. Fitting Moral Thinking into our Best Theories of Psychology**

Importantly, however, all of this theorizing leaves one philosophical task completely open for expressivists: They must show that moral thinking so characterized is constituted by *conative attitudes*. How can they show this? I will not endeavour to carry out this major research program here, but as an application for a license for optimism, let me note the basic structure I think it

should take.

In my view, expressivists can establish that moral thinking as characterized by the Ramsey-sentence is constituted by *conative attitudes* by doing two things: First, by developing a plausible, empirically informed, and naturalistic philosophical theory of psychology that gives (i) a characterization of the general nature of conative attitudes and what features they can have in principle, (ii) a characterization of the general nature of beliefs and what features beliefs can have in principle, and (iii) a theory of the psychological laws governing the interaction between these two kinds of mental states, the possible interactions within each kind of mental state and their connection to action. Second, by arguing on the basis of this theory that, given the characterization of moral thinking derived by philosophical theorizing, the best account of moral thinking is to regard it as constituted by a distinctive kind of conative attitude. This will be the place, for example, at which expressivists need to show that they can give a plausible *general* theory of logically complex thoughts that fits with their thesis about moral thinking. To do so, they need plausible theories about the nature of commitments in general and the nature of non-moral commitments in particular and they need to explain how these theories fit with their thesis about the nature of moral thinking. Expressivists have several options here.<sup>26</sup> But this paper is neither the place to go into detail with regard to nor to argue for any of these options.

Note, however, that on this approach it will *not* be necessary for expressivists that they can point out any specific kinds of familiar conative attitudes as candidates for moral thinking in order to establish that the best account of moral thinking as characterized by the Ramsey-sentence is to regard it as being constituted by a distinctive kind of conative attitude. Instead, it will be sufficient to show that conative attitudes can *in principle* have the characteristics

possessed by moral thinking and its modes and that—given our characterisation—it is more plausible to think of moral thinking and its instances in terms of conative attitudes than beliefs. Indeed, certain ways in which expressivists have proceeded are already compatible with this approach: their arguments often do not require that instances of moral thinking consist in e.g. desires or plans, but only that they are *desire-like* or *plan-like*, meaning that their behaviour can be explained by pointing to ways that conative attitudes could *in principle* behave. For example, Blackburn's and Gibbard's solutions to the Frege-Geach problem (however plausible they are) do not require that moral thinking *consists* in those mental states they claim can stand in the relevant inferential relations, but could be merely used to show that conative attitudes, and so moral thinking if it is a conative attitude, can *in principle* stand in those relations.

Once this further task in the philosophical theory of psychology is completed, expressivists who hold that moral thinking consists in *sui generis* attitudes will have done everything necessary to answer the question after the nature of those conative attitudes that constitute moral thinking. They will have done so by giving a characterization of moral thinking and then arguing that the mental states so characterized can be only a kind of conative attitude. A remaining question is whether and by what natural facts these mental states are realized. But this task is for psychology, cognitive science and neuroscience, not philosophy. Nevertheless, if we establish, based on a plausible naturalistic philosophical theory of psychology, that there can be attitudes that satisfy the characterization of moral thinking, we can be optimistic that moral thinking will be realized by natural facts.

#### **4. Conclusion**

In this paper I have argued for two conclusions: First, that there is an important symmetry between the Moral Attitude Problem and a problem meta-ethical cognitivism faces, and that this symmetry puts expressivists and cognitivists on an equal footing regarding important theoretical resources they can use to address these problems. Second, that those resources allow expressivists not only to take the Moral Attitude Problem by its second horn—holding that moral thinking is reducible to attitudes fully describable in non-moral terms—, but also to take it by its first horn—holding that moral thinking consists in *sui generis* attitudes. To support this contention, I sketched just enough of one such non-standard expressivist position to vindicate that adopting such an approach is feasible and provide some “license for optimism” regarding the chances of success along this route for expressivists.

\* I want to express special thanks to Matthew Chrisman, Guido Ehrhardt and Mike Ridge. For helpful discussion and comments I would also like to thank Hannah Altehenger, Vuko Andrić, Simon Blackburn, Cameron Boulton, Lars Dänzer, James Dreier, Ben Ferguson, Romy Jaster, Robin McKenna, Timothy Kunke, Emil Frederick Ljungberg Møller, Geoffrey Sayre-McCord, Peter Schulte, Max Seeger, Lee John Whittington, Alan Wilson, audiences at the SOPhia in Salzburg and the GAP 8 in Konstanz, two anonymous referees for *Ethics* and the editors of *Ethics*.

<sup>1</sup> Alexander Miller, *An Introduction to Contemporary Metaethics* (Oxford: Polity, 2003).

<sup>2</sup> Let me make a clarificatory note about my use of the phrase “moral thinking” (the same holds *mutatis mutandis* for my use of the phrase “descriptive thinking”): the theories this paper is concerned with—cognitivism and expressivism—are rival accounts of one and the same phenomenon. However, these theories give very different accounts of this phenomenon and so when we describe it, we are well-advised to do so in terms that are as theory-neutral as possible. For these purposes I use the phrase “moral thinking”, as I think it is a relatively theory-neutral way of picking out that phenomenon which cognitivism and expressivism are concerned with (I use the phrase “moral judgement” in a similar way. See Sebastian Köhler, “Expressivism, Subjectivism and Moral Disagreement,” *Thought: A Journal of Philosophy* 1 (2012): 71-78, 75). It is important though that this terminological choice is not merely stipulative: I think that in ordinary discourse the phrase “moral thinking” is used in exactly this way, namely to pick out the phenomenon that cognitivism and expressivism are interested in. Unfortunately, however, “thinking” as in “moral thinking” has—in ordinary discourse—both a process and a product sense (“thought” as in “moral thought”—which I take to be the best alternative—has the same problem). This might lead some readers to the conclusion that I am suggesting that cognitivism and expressivism provide theories of processes in terms of states, which is, of course, not what I have in mind. So bear in mind that when I use the phrase “thinking” in “moral thinking” I am using it in the product, not the process sense and that I take it that “moral thinking” in the product sense picks out the phenomenon that cognitivists and expressivists are interested in. I thank an editor of *Ethics* for drawing my attention to this issue.

<sup>3</sup> I am ignoring, for simplicity's sake, hybrid views according to which moral thinking consists in *both* beliefs *and* conative attitudes. Of course, such views also face the Moral Attitude Problem, as they have to give an account of the attitude partly constituting moral thinking. However, hybrid theorists should be able to endorse everything I say about how to address that problem.

<sup>4</sup> See e.g. Justin D'Arms and Daniel Jacobson, “Expressivism, Morality, and the Emotions,” *Ethics* 104 (1994): 739-763, Manfred Harth, *Werte und Wahrheit* (Paderborn: Mentis, 2008), 51-53, Antti Kauppinen, “A Sentimentalist Solution to the Moral Attitude Problem,” *Oxford Studies in Metaethics* 5 (2010): 225-257, John McDowell, “Truth and Projection in Ethics,” in *Mind, Value and Reality*, ed. John McDowell (Cambridge M.A. and London: Harvard University Press, 1998), 158-59, David Merli “Expressivism and the Limits of Moral Disagreement,” *The Journal*

of *Ethics* (2008): 22-55, Miller, *An Introduction to Contemporary Metaethics*, 43-51, 88-94, Michael Smith, "Some Not-much-discussed Problems for Non-Cognitivism in Ethics," *Ratio* 14 (2001): 93-115, 107-114, and Judith Jarvis Thomson, "Objectivity," in *Moral Relativism and Moral Objectivity*, ed. Gilbert Harman and Judith Jarvis Thomson (Oxford: Blackwell, 1996), 104-111.

<sup>5</sup> Thomson, "Objectivity," 110.

<sup>6</sup> Smith, "Problems for Non-Cognitivism in Ethics," 107-108.

<sup>7</sup> Crispin Wright, "Moral Values, Projection and Secondary Qualities," *Proceedings of the Aristotelian Society* 62 (1988): 1-26, 11-13.

<sup>8</sup> See e.g. Michael Smith, *The Moral Problem* (Oxford: Wiley-Blackwell, 1994), 4-11.

<sup>9</sup> Of course, it is a quite vexed question what a "naturalistic" world-view consists in, which makes it a vexed question what it takes for a theory to be incompatible with such a world-view (see e.g. David Papineau, "Naturalism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford CA: Stanford University, 2009), <http://plato.stanford.edu/archives/spr2009/entries/naturalism/>, or Michael Ridge, "Moral Non-Naturalism," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford CA: Stanford University, 2010), <http://plato.stanford.edu/archives/spr2010/entries/moral-non-naturalism/>). To avoid these thorny issues, the discussion of which is beyond the purposes of this paper, I only want to presuppose what I take to be a minimal criterion for what it takes for an account of moral thinking to be compatible with naturalism, namely that the account fits with our best scientific theories, in particular our best scientific theories of psychology. Furthermore, I take it that moral thinking exemplifies features which are *prima facie* in conflict with even this minimal criterion and that meta-ethical theories—if they accept naturalism—have the burden of proof to explain these appearances in a way that makes them compatible with the minimal criterion. Consequently, I assume that an account that takes the features of moral thinking at face value without giving a deeper explanation of those features would be in conflict with naturalism. I thank one of the editors of *Ethics* for drawing me out here.

<sup>10</sup> Alexander Miller introduced the Open Question Argument in the context of the Moral Attitude Problem (see Miller, *An Introduction to Contemporary Metaethics*, 47-51). He proposes two versions of the Open Question Argument one could use against expressivism: The first just relies on the intuition that competent speakers will hold the question to be open whether some conative attitude already part of our account of non-moral thinking is an instance of moral thinking. The other identifies some feature  $F_M$  characteristic of moral thinking and then argues that competent speakers will hold the question to be open whether some conative attitude already part of a theory of non-moral thinking has  $F_M$ . Since the difference between these versions is irrelevant for what follows I will present the argument in its simpler form.

<sup>11</sup> See Simon Blackburn, *Ruling Passions. A Theory of Practical Reasoning* (Oxford: Clarendon Press, 1998), 8-14, 51.

<sup>12</sup> For similar arguments, see Smith “Problems for Non-Cognitivism in Ethics,” 107-114 and Miller, *An Introduction to Contemporary Metaethics*, 88-94. An anonymous referee suggested that a better interpretation of Blackburn is to read him as proposing that moral attitudes are complex *dispositions* of the following kind: thinking  $\Phi$ -ing to be wrong, for example, is being disposed to avoid  $\Phi$ -ing, to insist that others avoid  $\Phi$ -ing, to feel guilty for  $\Phi$ -ing, to feel angry with those who  $\Phi$ , to care whether others share these feelings, etc. The referee suggests that this account might evade the Open Question Argument. I’m unsure whether this is a better reading of Blackburn than mine: Blackburn’s remarks in *Ruling Passions* seem compatible with both. But, independently of the correct Blackburn exegesis, it is of course worthwhile to ask whether the referee’s suggestion deals with the Moral Attitude Problem. However, I doubt that Blackburn, on this reading, does fare better regarding the Open Question Argument. Although it might *appear* that the question is closed whether someone who is in that complex dispositional state is in a state of thinking that something is wrong (and *vice versa*), we can think of cases in which they come apart. Such a case has, for example, been proposed by Michael Smith when he used the Moral Attitude Problem to argue against Blackburn (see Smith, “Problems for Non-Cognitivism in Ethics,” 107-114): It seems plausible that someone’s moral judgements *cease* being *moral* judgements when they start feeling alienated from them. Accounts of the nature of moral thinking should respect this. It is not hard, however, to imagine someone with the above set of dispositions feeling alienated from them. Maybe they feel that these dispositions are the result of an upbringing they now disapprove of, for example. To avoid this kind of move Blackburn needs to say something additional, e.g. any of the things I propose later. As a straight reductive analysis of moral thinking even this idea fails.

<sup>13</sup> See Allan Gibbard, *Wise Choices. Apt Feelings. A Theory of Normative Judgement* (Cambridge M.A.: Harvard University Press, 1990), 126-150.

<sup>14</sup> *Ibid.*, 6.

<sup>15</sup> For more detailed arguments, see D’Arms and Jacobson, “Expressivism, Morality, and the Emotions,” and Kauppinen, “A Sentimentalist Solution to the Moral Attitude Problem,” 226-230.

<sup>16</sup> Of course, there is still some debate about the extent to which these approaches are successful (for example, as Terence Horgan and Mark Timmons have argued, these approaches, in so far as they are concerned with the second horn of the dilemma, might face problems with a variant of the Twin Earth thought experiment (see e.g. Terence Horgan and Mark Timmons, “Troubles for New Wave Moral Semantics: The ‘Open Question Argument’ Revived,” *Philosophical Papers* 21 (1992): 153-175, or Terence Horgan and Mark Timmons, “Analytical Moral Functionalism meets Moral Twin Earth,” in *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, ed. Ian Ravenscroft (Oxford: Oxford University Press, 2009)). However, for the dialectical purposes of this paper we can bracket these issues: independently of whether or not these approaches are fully successful, it should be conceded that they provide powerful, fruitful and promising resources to make progress on the dilemma that

cognitivists face. And all that this paper is arguing is that expressivists can use those very same powerful, fruitful and promising resources to tackle both horns of the Moral Attitude Problem. Of course, once this conclusion is established further debate can be pursued about the extent to which these expressivist approaches will be successful and one might even raise the question whether expressivists could not be *more* successful in using these resources to avoid the Moral Attitude Problem than cognitivists are in using them to avoid their version of the dilemma. These questions, however, are the topic of a different paper. Thanks to one of the editors of *Ethics* for helpfully pressing me on these issues.

<sup>17</sup> See e.g. Richard Boyd, "How to be a Moral Realist," in *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord (Ithaca: Cornell University Press, 1988), or David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), 163-167.

<sup>18</sup> See e.g., Smith, *The Moral Problem*, 36-39, Frank Jackson, *From Metaphysics to Ethics* (Oxford: Clarendon Press, 1998), 28-37, or Frank Jackson and Philip Pettit, "Moral Functionalism and Moral Motivation," *The Philosophical Quarterly* 45 (1995): 20-40, 22-29. The approach develops ideas of David Lewis. See e.g. David Lewis, "How to Define Theoretical Terms," *Journal of Philosophy* 67 (1970): 427-446, David Lewis, "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50 (1972): 249-258, or David Lewis, "Dispositional Theories of Value II," *Proceedings of the Aristotelian Society* 63 (1989): 113-137, 129-132.

<sup>19</sup> This should enable expressivists to escape the worry that expressivist accounts of the nature of moral thinking face Moral Twin Earth-type objections (see Merli, "Expressivism and the Limits of Moral Disagreement").

<sup>20</sup> In addition to using the Canberra approach, expressivists might, to provide an account of *sui generis* moral attitudes, also draw on Nicholas Sturgeon's work, a Cornell Realist who argues that moral facts are *irreducible natural* facts (see e.g. Nicholas L. Sturgeon, "Moral Explanations," in *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord (Ithaca: Cornell University Press, 1985). According to Sturgeon entities are *natural* entities, iff they play an ineliminable role in our best explanatory account of the world. Using this principle, expressivists could argue that moral thinking consists in *sui generis* attitudes as follows: First, moral psychology and the entities it postulates, namely instances of moral thinking, play an explanatory role in our best explanatory account of human behaviour and mental life. Second, these entities cannot be fully accounted for in terms of mental states introduced by our account of non-moral thinking. So, moral psychology and the entities it postulates play an ineliminable role in our best explanatory account of the world. However, so the final step of the argument, moral psychology provides good reasons to see moral thinking as consisting in conative attitudes. Therefore, moral thinking consists in naturalistically respectable *sui generis* attitudes. To satisfy Informativity we might then argue that moral psychology will make interesting discoveries about these attitudes. I thank an anonymous referee for demanding that I elaborate on this strategy.

<sup>21</sup> This notational treatment is inspired by Lewis' suggestion that functional role accounts of beliefs treat beliefs as relations between subjects and propositions to avoid the problem of having to introduce a new noun-phrase for any proposition which can be believed (see Lewis, "Psychophysical and Theoretical Identifications," n. 13). Note that on notational treatments of this kind the platitudes about the states so characterized will contain universally quantified variables—variables standing for propositions on views like Lewis' or objects on my notational treatment. Although inspired by Lewis' proposal, let me emphasize that my proposal is *not* supposed to come with philosophical commitments about the nature of the objects to which moral thinking relates subjects. For example, a common philosophical assumption is that we can only be related to or quantify over things that *exist* (or *vice versa* that we are committed to the existence of everything we quantify over or say we stand in relation to). My account is not supposed to share this assumption, because "objects" are supposed to be "whatever moral thinking could conceivably be about", where this is determined by our *folk-theory* of what these modes could be about and it is implausible that our folk-theory is committed to robust philosophical assumptions. So, "objects" could be things the existence of which is only possible or even metaphysically impossible, if the folk are committed to that (which might even be philosophically feasible. See e.g. Nathan Salmon, "The Logic of What Might Have Been," *Philosophical Review* 98 (1989): 3-34, and Scott Soames, "Actually," *Proceedings of the Aristotelian Society Supplementary Volume* 81 (2007): 251-277).

<sup>22</sup> We might reduce the number of platitudes about logically complex mental states required for a characterization of moral thinking, by giving recursive compositional rules for mental states instead of listing one-by-one in what rational and inferential relations atomic moral judgements can stand to what mental states (see e.g. Wayne A. Davis, *Nondescriptive Meaning and Reference. An Ideational Semantics* (Oxford: Clarendon Press, 2005), 222-226). Going into detail how this might work, however, is beyond this paper's scope.

<sup>23</sup> Note that by calling mental states "logically *complex*" and others their "*parts*" I do *not* want to commit myself to the picture that mental states expressed by logically complex sentences are constituted by, constructed out of, etc. combinations of mental states which are *literally* the parts of those states. Instead, calling mental states "logically complex" or "parts" is merely supposed to pick out that they stand in certain rational and inferential relations to other mental states. Specifically, rational and inferential relations such that they are legitimately expressed by logically complex sentences in the former and rational and inferential relations such that they are legitimately expressed by the parts of logically complex sentences in the latter case. I thank an anonymous referee for demanding that I be clear on this issue.

<sup>24</sup> I thank an anonymous referee for pressing me on this. I would also like to thank Mike Ridge for very helpful conversations in the process of developing a response to these complications.

<sup>25</sup> In this context the question might be raised how we actually *distinguish* moral from non-moral logically complex

mental states. After all, while it is often clear how to categorize logically complex thoughts, sometimes it is not. For example, some logically complex mental states that have moral thoughts as their parts seem themselves not decisively moral, e.g. “ $2+2=4$  or murder is wrong”. The answer to this question is that this will be determined by our folk-theory of moral thinking. We obviously have an intuitive grasp on what distinguishes moral from non-moral thinking in general and moral from non-moral logically complex thoughts in particular, since it *is* often clear whether to categorize thoughts (logically complex or not) as moral or non-moral. So, our folk-theory of moral thinking includes platitudes regarding what characterizes moral thoughts in general and distinguishes them from non-moral thoughts, as well as regarding what characterizes moral logically complex mental states and distinguishes them from non-moral logically complex thoughts. In the case of moral logically complex mental states a candidate for such a platitude is, e.g., that someone in them is, *ceteris paribus*, disposed to praise those sharing them, be angry with those not sharing them, feel guilty for not adjusting their mental states in accordance with them, etc. It is *these* platitudes that determine how to distinguish moral from non-moral logically complex thoughts and have to be considered to settle unclear cases. Of course, interesting questions can be raised about this proposal, especially about the extent to which it works. Answering such questions, however, is beyond this paper's scope and not required for its purposes, given that difficulties with systematically distinguishing moral from non-moral thinking are problematic for expressivists and cognitivists alike. I thank an anonymous referee for drawing me out on these issues.

<sup>26</sup> See e.g. Simon Blackburn, “Attitudes and Contents,” in *Essays on Quasi-Realism*, ed. Simon Blackburn (Oxford: Oxford University Press (1993), Allan Gibbard, *Wise Choices. Apt Feelings*, 83-102, Michael Ridge, “Ecumenical Expressivism: Finessing Frege,” *Ethics* 116 (2006): 302-336, or Mark Schroeder, *Being For: Evaluating the Semantic Program of Expressivism* (Oxford: Clarendon Press, 2008).