The Report Committee for Christopher Aaron Simpson
Certifies that this is the approved version of the following report:

Can Non-Cognitivism Account for Ethical Explanation?

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor:
_____
Daniel Bonevac

Co-Supervisor:
_____
Jonathan Dancy

Can Non-Cognitivism Account for Ethical Explanation?

by

Christopher Aaron Simpson, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2014

Dedication

For my parents, James and Deborah, and my wife, Sandra.

Acknowledgements

Abstract

Can Non-Cognitivism Account for Ethical Explanation?

Christopher Aaron Simpson, MA

The University of Texas at Austin, 2014

Supervisors: Daniel Bonevac, Jonathan Dancy

In this report I argue that a popular account of the nature of ethical thought and talk – non-cognitivism – cannot make sense of our attempts to explain why some things are right or wrong, good or bad, just or unjust. After introducing the process by which we attempt to explain these sorts of ethical features (a process I call ethical explanation), I consider how we might test whether non-cognitivism can account for this process. We can test whether non-cognitivism can account for ethical explanation, I argue, by testing whether non-cognitivism can account for the meanings of ethical explanatory sentences, the sentences we use to express explanatory thoughts in ethics. After considering how non-cognitivism might account for ethical explanatory sentences (and so the thoughts these sentences express), I develop a series of problem cases on which, I argue, no plausible non-cognitivist account of these meanings of these sentences is possible. Because non-cognitivism cannot account for the meanings of ethical explanatory sentences, I conclude, non-cognitivism cannot account for ethical explanation.

Table of Contents

Introduction

        If an act is wrong, we expect there to be some explanation of this. No act is merely right or wrong in itself. Rightness and wrongness, like many other ethical features, require grounds. Suppose in the course of his teachings Socrates did something wrong and, in doing so, corrupted the youth of Athens. In that case we will expect there to be some explanation of the wrongness of Socrates' act, an explanation that cites the grounding features of his act's wrongness. In virtue of what, we might say, was Socrates' act wrong? What made his act obtain the ethically serious character of *wrongness*, rather than obtain some other character? Perhaps we will cite the grounds given by his accusers – that he was a maker of new gods and a denier of the old – or perhaps we will search for other grounds. But it seems that between ourselves and Socrates' accusers, themselves no fans of philosophy, there will be at least this much agreement: If Socrates did something wrong, there must be some explanation of this.

        The fact that many ethical features are not brute, but require explanation, is familiar and uncontroversial.[1] We find this presumption in both philosophic and everyday attempts to say why an act is right or wrong or good or bad, and in arguments attempting to show that there could be no such explanations.[2] This presumption underlies our conviction that ethical features can be made some sense of, and our discomfort when we find ourselves unable to make such sense of them.[3] Yet ethical explanation, the process by which we attempt to explain ethical features, remains not well understood. What is it we are doing when we engage in this process? Could this process ever be successful, and if so, how? Any complete answer to these questions is bound to be controversial, yet such answers must ultimately be given if we are to successfully illuminate this important and ubiquitous process. In this paper I attempt to contribute to such an illumination by answering a basic question with respect to this process, namely, "Can non-cognitivism account for ethical explanation?"

---

[1] Recent discussion of explanation in the context of ethics includes Rosen (2010), Zangwill (2011), and Väyrynen (2013). Probably the best known expression of puzzlement with regard to explanation in ethics is due to Mackie (1977) p. 41: "…it is wrong *because* it is a piece of deliberate cruelty. But just what in the world is signified by this 'because'?"

[2] Joyce (2007) p. 2 provides a nice list of arguments attempting to show that there could be no such explanations.

[3] Haidt (2012) provides numerous examples of subjects' discomfort at affirming that a given act was wrong while being unable to explain *why* that act is wrong – the phenomena known as *moral dumbfounding*.

I will ultimately argue that non-cognitivism cannot account for ethical explanation, but before we can proceed to my main argument, I will need to lay some necessary groundwork. I begin laying this groundwork in chapter 1 by locating the challenge for non-cognitivism concerning ethical explanation. Here, I argue that we ought to understand this challenge as a particular instance of the embedding problem; it is the challenge non-cognitivism faces in accounting for the meanings of ethical explanatory sentences, the sentences we use to express explanatory thoughts in ethics. Having located the challenge, I move on in chapter 2 to describe the embedding problem. I then consider one way of responding to this problem – Gibbard's fact-prac worlds account of sentence meaning – with the hope of developing a clearer picture of the resources the non-cognitivist has available to address the particular instance of the embedding problem we're interested in. In chapter 3 I proceed to my main argument. I begin by describing a set of problem cases for non-cognitivist accounts of the meanings of ethical explanatory sentences. These are cases involving disagreements between distinct ethical explanatory sentences on which, I argue, no plausible non-cognitivist account of the meanings of these sentences can be given. I move on to consider Gibbard's account of the meanings of ethical explanatory sentences, demonstrating that his specific proposal does indeed result in the sort of problem cases I describe. I then consider alternative ways a non-cognitivist might respond to these cases, ultimately determining that none appears satisfactory. I therefore conclude that non-cognitivism cannot account for ethical explanation.

Chapter 1 – Locating the Challenge

1.1 What is Non-Cognitivism?

      The non-cognitivist views I target in this paper are metaethical views about the nature of ethical thought and the meanings of ethical sentences. According to *non-cognitivism*, 1) ethical thoughts can be explained solely on the basis of the functional role these thoughts play in motivating an agent to act, and 2) all sentences (including ethical sentences) can be explained solely in terms of the thoughts those sentences express. Simon Blackburn and Allan Gibbard are the principal modern adherents of this sort of non-cognitivism.[4] According to Blackburn, ethical sentences express values, conceived of as relatively stable dispositions to act in certain ways, while according to Gibbard, ethical sentences express plans, conceived of as decisions about what to do in various situations.[5] In this paper, I focus primarily on Gibbard's views for reasons of simplicity, though my target is more general. My argument applies to anyone who accepts non-cognitivism, understood as I have defined it above.

1.2 How to Test Whether Non-Cognitivism can Account for Ethical Explanation

      Our question is whether non-cognitivism can account for *ethical explanation*, the process by which we attempt to explain ethical features. In this section, I argue that an appropriate way of testing whether non-cognitivism can account for ethical explanation is by testing whether non-cognitivism can account for the meanings of ethical explanatory sentences. I will first explain which sentences are ethical explanatory sentences, and then I will move on to explain the appropriateness of my proposed test. I will end this section by accounting for this test's primary advantage – it will allow us to locate the challenge for non-cognitivist accounts of ethical explanation in a familiar place, as a particular instance of the embedding problem.

      Ethical explanatory sentences are the sentences we use to express explanatory thoughts in ethics. I have in mind sentences like the following:

---

[4] Gibbard (2003), Blackburn (1998).

[5] Gibbard (2003) pp. 49-50, Blackburn (1998) p. 67

(1) Lying is bad *because* lying is willful deception.

(2) Right acts are right *in virtue of* the fact that they maximize happiness.

(3) The fact that Socrates ought to Φ given his circumstances *depends upon* the fact that Φing in his circumstances would be endorsed by an ideal reasoner.

These sentences are *ethical sentences*. Ethical sentences, for the purposes of this paper, have two defining features. First, an ethical sentence includes a paradigmatically normative or evaluative term, a term like "good," right," "ought," or their contraries. And second, an ethical sentence expresses a thought that is capable of playing a functional role in originating an agent's motivation to act.[6] These sentences meet both of these criteria. They each include a normative or evaluative term, "bad," "right," and "ought," respectively. In addition, each of the thoughts these sentences express is capable of playing a functional role in originating an agent's motivation to act. For instance, an agent who thinks (1) will generally be motivated not to lie. An agent who thinks (2) will generally be motivated to act so as to maximize happiness. And an agent who thinks (3) will generally be motivated to Φ, assuming he thinks he is Socrates in the specified circumstances.

In addition to being ethical sentences, sentences (1), (2), and (3) are also explanatory. *Explanatory sentences* answer "Why" questions in a given domain, questions about why various objects of explanation falling within that domain occur or obtain. Sentences of this form are common across all domains where philosophers have investigated explanation, and many domains that have yet to be investigated. Here are three examples of this sort of sentence from the domains of history, science, and metaphysics, respectively.

(4) The Empire of Japan surrendered to the Allies in 1945 *because* of the threat of another atomic bombing.

---

[6] It is important to distinguish the thoughts expressed by ethical sentences, thoughts that originate an agent's motivation to act, from other thoughts that may play some subsidiary or enabling role in motivating the agent to act. Only thoughts that originate said motivation are ethical thoughts.

(5) The fact that the particle is accelerating obtains *in virtue of* the fact that it is being acted on by some net positive force.[7]

(6) The fact that that object is a table *depends upon* the fact that it is composed of a series of atoms, arranged table-wise.

Explanatory sentences have a common form. That form is "X *because* Y" where "X" (but not necessarily "Y") is or expresses an object of explanation within the domain in question and "*because*" stands for a paradigmatic explanatory idiom like "because," "in virtue of" or "depends upon."[8] Objects of explanation vary from domain to domain. And which objects of explanation are appropriate for explanation in a given domain will depend on the specific details of one's view. For the purposes of this paper, I wish to remain metaethically neutral with regard to the objects of ethical explanation. In this vein I will claim only that the appropriate objects of ethical explanation either are or are expressed by ethical sentences. Sentences (1), (2), and (3) meet the criteria for being explanatory sentences. Each of these sentences has the common form of explanatory sentences, "X *because* Y." In addition, each of these sentences answers a "Why" question about an object of explanation in the ethical domain. Those questions are "Why is lying bad?" "Why are right acts right?" and "Why ought Socrates Φ given his circumstances?" respectively. And those objects of explanation are or are expressed by "Lying is bad," "Right acts are right," and "Socrates ought Φ given his circumstances," respectively. Because these objects of explanation are ethical, in addition to being explanatory sentences, sentences (1), (2), and (3) are also *ethical explanatory sentences*.

Having explained which sentences are ethical explanatory sentences, I will now explain the appropriateness of the test I proposed at the start of this section. That test, you'll recall, is intended to answer the question of whether non-cognitivism can account for ethical explanation. We can test whether non-cognitivism can account for ethical explanation, I claimed, by testing whether non-cognitivism can account for the meanings of ethical explanatory sentences, the sentences we use to express explanatory thoughts in ethics. This test is appropriate for answering our basic question because ethical

---

[7] This example is taken directly from Fine (2012) p. 1

[8] Explanatory idioms can reverse this pattern, becoming "Y *explains* X". For example, "The fact that lying is willful deception *explains* the fact that lying is wrong."

explanation, at least on non-cognitivism, *requires* the explanatory thoughts expressed by ethical explanatory sentences, and the meanings of these sentences are explained (according to non-cognitivism) in terms of the thoughts these sentences express. This means that a test concerning the meanings of these sentences can itself serve as a test for whether or not non-cognitivism can account for ethical explanation. If non-cognitivism cannot account for the meanings of ethical explanatory sentences then, ipso facto, they will not be able to account for the explanatory thoughts those sentences express. But since ethical explanation requires these thoughts, if non-cognitivism is not able to account for these thoughts, they will not be able to account for ethical explanation.

Let me say a bit more about ethical explanation and why it requires explanatory thoughts (at least on non-cognitivism). Ethical explanation, I keep saying, is the process by which we attempt to explain ethical features. It is difficult to say anything very detailed about this process without making controversial metaethical assumptions. Nevertheless this is a familiar and ubiquitous process, engaged in by both philosophers and laypeople alike. One of the ways philosophers engage in ethical explanation is by constructing ethical theories. These theories attempt to answer, in a general way, various "Why" questions in ethics, questions about why different sorts of objects of explanation in ethics obtain. For example, in addition to asking why a particular act is right, I can also ask the more general question, "Why are right acts right?" Consequentialism gives an answer to this question. According to this theory, right acts are those that produce the best consequences and those acts are right *because* they produce the best consequences. The Humean Theory of Reasons can be thought of as addressing the general question, "Why is R a reason for A to φ?" The answer, according to one version of the theory, is that R is a reason for A to φ because R explains how A's φing fulfills one of A's desires. While philosophers often focus their attention on answering the most general sorts of "Why" questions in ethics, laypeople (and philosophers in everyday life) often pursue answers to more particular sorts of "Why" questions in ethics. For example, in the 2012 U.S. election, many voters answered the questions "Why should I vote for Barack Obama?" and "Why should I vote for Mitt Romney?" in an attempt to decide whom to vote for. Just this morning, I answered the particular question "Why ought I work on this paper?" in the course of deciding what I would do today. And you have probably recently

answered the question "Why ought I read this paper?" in the course of deciding what you are now doing right now.

We can see from these examples that, ordinarily, when we attempt to answer these sorts of "Why" questions in ethics, questions about why a given ethical feature obtains, we engage in ethical explanation; ordinarily, when we attempt to say why a given act is wrong (for instance), we're attempting to explain why that act is wrong.[9] Further, I claim, we can see from these examples that *correctly answering* these sorts of "Why" questions in ethics is required in order for ethical explanation to succeed; we cannot succeed in explaining why a given act is wrong (for instance) without correctly answering the question of why that act is wrong. Although I doubt many would deny that explaining why a given act is wrong requires answering the question of why that act is wrong, some might doubt the general claim I made in light of specific examples such as this: that successful ethical explanation requires answering a certain sort of "Why" question. After all, one might argue, there are other sorts of questions one might attempt to answer, the correct answering of which we might intuitively think of as instances of successful ethical explanation. To give just two examples, the questions "In virtue of what qualities was Adolf Hitler a bad man?" and "How did Adolf Hitler become a bad man?" are questions we might correctly answer, and neither of these is a "Why" question. Further, we might be tempted to think that correctly answering these questions would qualify as instances of successful ethical explanation. Nevertheless, although it is possible to answer both of these questions, this fact fails to falsify my claim, that correctly answering a certain sort of "Why" question is required in order for ethical explanation to succeed.

The second sort of question fails to falsify my claim, because, although we might provide an answer to this question in the form of a causal explanation for how Adolf Hitler became a bad man, such an explanation would fail to be an ethical explanation. The reason for this is that the corresponding object of explanation that this question wants

---

[9] One exception to this ordinary pattern is when a person attempts to say why an act has a given moral feature despite knowing that the given act does not have that feature. For example, suppose that Richard Nixon's bugging the offices of his political opponents was morally wrong and that he knew this to be wrong. In that case he might still attempt to say why his acts were morally acceptable as part of an attempt to deceive the public. In doing so he would not be attempting to explain why his acts were morally acceptable, though he would of course be purporting to be engaging in such an attempt.

7

demystified – "Adolf Hitler became a bad man" – is not an object appropriate for ethical explanation. Such an object is non-ethical, for although it includes a normative component ("bad"), this object is incapable of playing a functional role in originating an agent's motivation to act.[10] It is important to recognize that ethical explanations (and not just the ethical explanatory sentences that are related to these explanations) are always non-causal explanations that have to do with *why* some object of ethical explanation obtains. This is why the first sort of question also fails to falsify my claim. Although the explanation given in response to this question would be an ethical explanation, such an explanation, once given, would answer an ethical "Why" question, namely "Why was Adolf Hitler a bad man?"[11] Because ethical explanations always have to do with *why* some object of ethical explanation obtains, there will always be some corresponding "Why" question these explanations answer, even if we are able to phrase these inquiries in some way that does not make use of a "Why." Since any instance of successful ethical explanation will at least answer one of these corresponding "Why" questions, my claim stands: *Correctly answering* a certain sort of "Why" question, a question about why some object of ethical explanation obtains, is required in order for ethical explanation to succeed.[12]

Because ethical explanatory sentences answer these "Why" questions, we often make use of them when we engage in ethical explanation. But ethical explanation can be engaged in without making use of an ethical explanatory sentence, and the use of an ethical explanatory sentence does not always suffice to engage in ethical explanation. I can engage in ethical explanation without making use of an ethical explanatory sentence by, for instance, drawing a diagram or constructing a parable. And I can fail to engage in ethical explanation while making use of an ethical explanatory sentence by, for instance,

---

[10] Although "Adolf Hitler is a bad man," "Adolf Hitler was a bad man," or "Adolf Hitler will be a bad man," each express objects capable of playing a functional role in originating an agent's motivation to act, "Adolf Hitler became a bad man," can only play a subsidiary or enabling role in motivating an agent to act. This is because this last expression does not properly ground any moral attitude in itself. Only the former expressions do.

[11] Note that, as phrased, this question could correspond to either question we're discussing, "In virtue of what qualities was Adolf Hitler a bad man?" or "How did Adolf Hitler become a bad man?" It is only the answer to the "Why" question corresponding to the first question that is an ethical explanation.

[12] By "correctly answering" I simply mean providing the correct answer to. It is certainly possible to provide the answer to questions not made explicit. For example, on learning some new piece of information, people sometimes say, "That answers a lot of questions." In these cases, none of the questions that have been answered by this new piece of information have been made explicit.

uttering an ethical explanatory sentence from a foreign language without understanding it's meaning.[13] It is clear that what is required to engage in ethical explanation is not the expression of an ethical explanatory sentence, but is instead an attempt on the part of the agent to provide an answer to an ethical "Why" question, a question about why some object of ethical explanation obtains.[14] If correctly answering this sort of "Why" question is required in order for ethical explanation to succeed, attempting to provide a correct answer to this sort of "Why" question is required in order to engage in ethical explanation at all. Just as we cannot explain why a given act is wrong (for instance) without correctly answering the question of why that act is wrong, we cannot *attempt to explain* why an act is wrong without *attempting to provide a correct answer to* the question of why that act is wrong. All of this is important because it means that any account of ethical explanation, the process by which we attempt to explain ethical features, must include an account of what it is an agent does in attempting to provide a correct answer to an ethical "Why" question. Because attempting to provide a correct answer to an ethical "Why" question is (at least) part of what it is to engage in ethical explanation, any account of ethical explanation that cannot explain what an agent does in making such an attempt will be incomplete.[15] It is beyond the scope of this paper to venture a full answer to the question of what it is to attempt to provide a correct answer to an ethical "Why" question (and to do so would embroil us in unnecessary controversy). Nevertheless, at least this much is certain: If attempting to provide a correct answer to an ethical "Why" question is required in order to engage in ethical explanation, attempting to provide *an answer* to an ethical "Why" question is required in order to engage in ethical explanation. Just as we cannot attempt to explain why a given act is wrong (for instance) without attempting to provide a

---

[13] For another example, see footnote 9.

[14] An agent may attempt to express an answer to a "Why" question without attempting to answer said "Why" question if the agent instead is attempting to answer a question correspondent to said "Why" question i.e. attempting to answer the question "In virtue of what qualities was Adolf Hitler a bad man" instead of the question "Why was Adolf Hitler a bad man?"

[15] What else might be required to engage in ethical explanation? The answer to this question will turn on 1) one's views about what constitutes a correct answer to an ethical "Why" question, in particular whether there are good-making features of explanations that are not built into the notion of a correct answer to a "Why" question e.g. whether the absence of irrelevancies in an explanation is required for it to count as a correct answer and 2) Whether one thinks that pursuit of all or some these additional good-making features is required in order to engage in ethical explanation e.g. a prankster delivering a correct answer to an ethical "Why" question riddled with irrelevancies might be seeking to obscure an object of explanation rather than seeking to illuminate it.

correct answer to the question of why that act is wrong, so we cannot attempt to provide a *correct answer* to the question of why an act is wrong without attempting to provide *an answer* to the question of why that act is wrong.

So, what is it that an agent does in attempting to provide an answer to an ethical "Why" question? A natural thing to say is that an agent attempts to answer an ethical "Why" question by attempting to express an ethical explanatory sentence. After all, ethical explanatory sentences *do* answer ethical "Why" questions. Nevertheless, this answer cannot be correct, for we have already seen that agents can attempt to answer these ethical "Why" questions without attempting to express ethical explanatory sentences, and that they can attempt to express ethical explanatory sentences without attempting to answer ethical "Why" questions.[16] Still, something important can be gleaned from the fact that agents often attempt to express ethical explanatory sentences in their attempts to answer ethical "Why" questions. And that is that attempting to express an ethical explanatory sentence is often *a way of doing something else*, that itself constitutes an agent's attempt to provide an answer to an ethical "Why" question. What is this something else? A couple of possibilities present themselves. First, we might think that an agent attempts to answer an ethical "Why" question by attempting to express a proposition of some sort. For example, perhaps agents attempt to answer ethical "Why" questions by attempting to express one of the propositions that ethical explanatory sentences ordinarily express (A proposition of the form "X *because* Y"). Another possibility is that agents attempt to answer an ethical "Why" question by attempting to express a thought of some sort. For example, perhaps agents attempt to answer ethical "Why" questions by attempting to express an ethical explanatory thought, the kind of thought that ethical explanatory sentences ordinarily express (A thought of the form "X *because* Y"). These two possibilities may not be the only ones, though they do seem to me to be the two most plausible candidates. Nevertheless, although there may be a variety of possible answers to the question "What is it that an agent does in attempting to provide an answer to an ethical 'Why' question?" the non-cognitivist can answer in only one way. The non-cognitivist must claim that an agent attempts to provide an answer to an ethical

---

[16] They can do the former by (for instance) constructing a parable or drawing a diagram, and the latter by (for instance) attempting to express an ethical explanatory sentence from a foreign language without understanding it's meaning.

"Why" question by attempting to express an explanatory thought, the sort of thought that is ordinarily expressed by an ethical explanatory sentence. We can see this by focusing on the second thesis that the non-cognitivist accepts, that "all sentences (including ethical sentences) can be explained solely in terms of the thoughts those sentences express." Because the non-cognitivist is committed to explaining ethical explanatory sentences *solely* in terms of the thoughts these sentences express, and these sentences answer ethical "Why" questions, the non-cognitivist must say that ethical explanatory sentences answer these ethical "Why" questions in virtue of the thoughts they express – explanatory thoughts. This means, crucially, that the non-cognitivist must claim that what it is for an agent to attempt to provide an answer to an ethical "Why" question (to engage in ethical explanation) is to attempt to express one of these explanatory thoughts. All of this is to say, that for the non-cognitivist, expressing explanatory thoughts is *how* one engages in ethical explanation. This means that an account of these thoughts is, as I stated earlier, required for a non-cognitivist account of ethical explanation. Since we can test whether the non-cognitivist can account for these thoughts by testing whether they can account for the meanings of ethical explanatory sentences, we can now clearly see that an appropriate way of testing whether non-cognitivism can account for ethical explanation is by testing whether non-cognitivism can account for the meanings of these sentences.

If explanatory thoughts are so important for ethical explanation and ethical explanatory sentences comparably less so, one might wonder why I don't propose to test whether the non-cognitivist can account for ethical explanation more directly, by focusing on the explanatory thoughts themselves. To be clear, my interest in ethical explanatory sentences *is* a product of my interest in the explanatory thoughts these sentences express. And much of what I have said and will say *does* focus directly on these explanatory thoughts. Nevertheless, I stand by my proposed test, for two reasons.

Firstly, let me reiterate, because non-cognitivism explains sentence meaning in terms of the thoughts these sentences express, testing for whether non-cognitivism can account for the meanings of ethical explanatory sentences *is* sufficient for testing whether non-cognitivism can account for explanatory thoughts in ethics. Secondly, and most importantly, conceiving of our test as a test about the meanings of ethical explanatory sentences, allows us to locate the challenge for non-cognitivism, the challenge of giving

an account of ethical explanation, in a familiar place. We can see that the challenge non-cognitivism faces in giving such an account is simply a particular instance of the challenge of answering the classic embedding problem; we can see it as the challenge non-cognitivism faces in accounting for the meanings of the complex ethical sentences that express the explanatory thoughts necessary for ethical explanation. Seeing the challenge in this light allows us to better understand some of the difficulties the non-cognitivist faces in giving an account of ethical explanation. But even more importantly, seeing the challenge in this light will allow us to better understand *how* the non-cognitivist must answer it. For these reasons, the test I've proposed is not just appropriate. It is also useful. For thinking of this test as embodying this particular challenge will serve to structure our attempt to answer this paper's question, the question of whether non-cognitivism can account for ethical explanation.

Chapter 2 – Responding to the Embedding Problem

2.1 The Embedding Problem

I have argued that we can think of the challenge non-cognitivism faces with respect to giving an account of ethical explanation as the challenge non-cognitivism faces in accounting for the meanings ethical explanatory sentences. This is the challenge of accounting for meanings of the ethical sentences that express the explanatory thoughts necessary for ethical explanation, and it is a particular instance of the classic embedding problem. In this section, I describe that problem. This section mainly serves as background for Gibbard's attempted solution to the problem, which I present in the following section. That attempted solution will itself allow us to see in more detail the resources the non-cognitivist has available to address the particular instance of the embedding problem we're interested in.

The embedding problem, also known as the Frege-Geach problem, is a classic problem non-cognitivism faces in accounting for the meanings of complex sentences that embed simple, predicative ethical sentences.[17] The non-cognitivist faces a problem because simple ethical sentences are capable of embedding in all of the same contexts as simple non-ethical sentences and appear to function the same as non-ethical sentences once embedded. But according to the non-cognitivist, ethical sentences express thoughts that motivate while non-ethical sentences generally express thoughts that represent. So it is hard to see why these distinct types of sentences should function the same.

Let me give a concrete example. Like the non-ethical sentence "Chairs support persons," the ethical sentence "Lying is wrong," is capable of being embedded in a premise of an argument, *and* of playing a role in that argument's validity. Consider the following two arguments:

(1) Lying is wrong.
(2) If lying is wrong, it's wrong to get your brother to lie.
So: (3) It's wrong to get your brother to lie.

---

[17]Some of the earliest discussion of this problem is in Geach (1965) and Searle (1962).

(4) Chairs support persons.

(5) If chairs support persons, chairs support mice.

So: (6) Chairs support mice.

One way of making sense of the second argument's validity it by claiming that the phrase "Chairs support persons," has the same meaning as it appears in (4) and in the antecedent of (5). This allows one to see this argument as an instance of modus ponens. But the non-cognitivist cannot say this about the first argument. This is because (among other things) the non-cognitivist must explain the meaning of "Lying is wrong," in (1) as expressing some thought that plays some functional role in originating and agent's motivation to act (presumably not to lie). But the phrase "Lying is wrong" in (2) does not play this same role in motivating the agent to act since I can accept (2) without being motivated in the same way as (1). This means the meaning of the phrase "Lying is wrong," must differ from (1) to (2). And since this is the case, the non-cognitivist cannot explain the validity of the first argument by seeing it as an instance of modus ponens (since the antecedent of the conditional claim in modus ponens must match the other premise). They must explain the appropriateness of the transition from (1) & (2) to (3) in some other way. And they must explain the meaning of (2) as expressing some thought or thoughts that does not include the ethical thought that (1) expresses.[18]

Although the non-cognitivist has trouble explaining the meanings of conditional claims that embed simple ethical sentences like (2), and of explaining how these sentences can play a role in making an argument valid, these are not the only sentences the non-cognitivist has difficulty explaining. In *Being For*, Mark Schroeder emphasizes the extent of the non-cognitivist's problem nicely:[19]

In fact, and this cannot be emphasized enough, *every* natural-language construction that admits of descriptive predicates admits of moral predicates, and seems to function in precisely the same way: tense; conditionals; every kind of modal–alethic, epistemic, or deontic; qualifiers like 'yesterday'; generics and

---

[18] Indeed, (2) may not express an ethical thought at all.
[19] Schroeder (2008) p. 5

habituals; complement-taking verbs like 'proved that' and 'wonders whether'; infinitive-taking verbs of every class, including 'expects to', 'wants to', and 'compels to'; binary quantifiers like 'many' and 'most'; and more. It is crucially important to understand that the embedding problem for non-cognitivism is not simply a problem about the validity of *modus ponens*, or even simply about logic. *Every* construction in natural languages seems to work equally well no matter whether normative or descriptive language is involved, and to yield complex sentences with the same semantic properties. So any view on which moral language works in some way that is deeply different from descriptive language seems extraordinarily unlikely to be true–at least if it is a hypothesis about the language that we actually speak.

We can see that the crux of the embedding problem is that the non-cognitivist claims that ethical sentences and non-ethical sentences are importantly different– one sort of sentence expresses thoughts that motivate, while the other sort of sentence expresses thoughts that (typically) represent. Yet the view yields no obvious story about why it is that simple ethical and non-ethical sentences should embed in all of the same contexts and function so similarly once embedded. Indeed, non-cognitivism yields no obvious story about the meanings of the complex sentences that embed simple ethical sentences at all! Because of this, the non-cognitivist faces the extraordinary challenge of making sense of the similar behavior of embedded ethical and non-ethical sentences in these contexts while simultaneously making sense of these sentences' divergent meanings at these contexts.

It is no exaggeration to say that the embedding problem has come to be seen as the greatest challenge for non-cognitivism generally. Nevertheless our focus in this paper is not on this general challenge, but rather on a specific instance of it: Can the non-cognitivist account for the meanings of ethical explanatory sentences? Our focus on this specific instance of the challenge is, we've seen, a product of our interest in ethical explanation, and of our interest in answering a key question concerning ethical explanation, namely: Can non-cognitivism account for ethical explanation?

Clearly, although the answer to this question is interesting in its own right, an affirmative answer to this question would not give us grounds for claiming that the non-cognitivist could solve the embedding problem generally. It would at most give us grounds for thinking that this particular instance of the embedding problem involving ethical explanatory sentences could be solved. Nevertheless, I think that whether or not the non-cognitivist can solve this particular instance of the embedding problem is not of trivial significance with regard to the general embedding problem; I think that focusing on this particular instance of the embedding problem is instructive with regard to the problem more generally. Part of my reason for thinking this is that considering how to answer particular instances of general problems is almost always instructive in clarifying the nature of the problem and of the prospects for a general solution. But another reason is that in this case I am aiming to show that this particular instance of the embedding problem *cannot* be solved. If this is correct, than this *would* answer the question of whether the non-cognitivist can solve the general embedding problem, and so focusing on the details of this particular instance of the embedding problem is relevant to the problem more generally. In any case, as we'll see, this particular instance of the embedding problem admits of no obvious solution. For that reason alone, it is worthwhile for the non-cognitivist to investigate this particular instance of the embedding problem, to see if they can do any better than I can at solving it for them. If in the course of examining the difficulties of this particular case a solution is found, the non-cognitivist will be only that much closer to finding a solution more generally.

2.2 Gibbard's Solution to the Embedding Problem

There have been many attempted solutions to the embedding problem over the years, though Gibbard's is certainly one of the best at *clearly* illustrating the general resources the non-cognitivist has available for solving this problem. Gibbard attempts to solve the embedding problem for non-cognitivism generally by developing what he calls a "device of automation" for answering questions about the meanings of complex ethical sentences. Since, as we've seen, the challenge for the non-cognitivist concerning the meanings of ethical explanatory sentences is just a particular instance of the embedding problem, Gibbard's "device of automation" is all too relevant for our purposes. In this

section I detail Gibbard's solution to the embedding problem with the hope of providing a clearer picture of *how* the non-cognitivist must go about solving it.

Non-cognitivists will have solved the embedding problem when they have offered a principled way of accounting for the meanings of sentences that explains how ethical sentences and non-ethical sentences can function the same way in various contexts while expressing different sorts of thoughts. Gibbard's solution involves the basic thought he takes ethical sentences to express – plans. *Plans*, as Gibbard conceives them, are decisions about what to do in particular scenarios, including counterfactual scenarios. An ethical sentence like "Murder is wrong," Gibbard claims, expresses a plan not to murder, for any agent, in any possible situation, where this is an available action. So construed, my plans include decisions for myself, about what I will do in a given situation, but they also include plans for others, plans for what to do in another person's shoes.[20] These plans for what to do in a person's shoes can be very general, applying to all possible agents in all possible situations – as the plan expressed by the sentence "Murder is wrong," is – or they can be highly specific, applying to just a single agent in a single situation.[21]

This characterization of my plans as plans for others as well as myself, Gibbard claims, enables him to explain how ethical sentences and non-ethical sentences could function the same despite the fact that the sort of thoughts expressed by each sort of sentence are very different. Both sorts of sentences function the same way in a variety of contexts, Gibbard claims, because of the possibility of disagreement with both sorts of sentences. "People can agree or disagree in belief," Gibbard says, "and they can, in a sense, agree or disagree in plan."[22] It is this capacity for disagreement, the capacity for ethical disagreement as well as for factual disagreement, that allows both ethical and non-ethical sentences to embed and function similarly in the same sorts of contexts. Gibbard illustrates how this is supposed to work generally, by illustrating how it works in the context of the following argument:

---

[20] Gibbard (2003) pp. 48-49
[21] Gibbard (2003) p. 54 claims I can plan for what to do in the case where I am Caesar at the Rubicon.
[22] Gibbard (2003) p. 48

(1) Either packing is now the thing to do, or by now it's too late to catch the train anyway.

(2) It's not even now too late to catch the train.

(So) Packing is now the thing to do.

This argument, Gibbard claims, results in an ethical conclusion, and includes a premise, (1), that connects a statement of plan "Packing is now the thing to do," to a statement of fact "It's too late to catch the train anyway," via a disjunction. We can understand the meanings of each of these sentences, including the sentence that mixes a statement of fact and a statement of plan expressivistically, by characterizing the states of mind these sentences are used to express. The first premise expresses a state of mind that *rules out* both deciding (ethically) that it's not the case that "Packing is now the thing to do", and deciding (factually) that it's not the case that "It's not even now too late to catch the train." The second premise *rules out* deciding (factually) that it's not the case that "It's not even now too late to catch the train." The combined effect on the agent that accepts both (1) and (2) Gibbard claims, is that the agent is forced to *rule out* deciding (ethically) that it's not the case that "Packing is now the thing do," to conclude, in other words, that "Packing is now the thing to do."

This explanation of the "force," so to speak, of the argument, as derived from the pressure to move from certain states of mind to other states of mind that follow from these initial states of mind, Gibbard applies to purely factual arguments as well. This pressure is the pressure to be consistent in one's decisions about *how things are* as well as in decisions about *what to do*, the pressure to follow through on these decisions to their proper conclusions. We can understand these decisions, including decisions that mix fact with plan (as the decision that (1) does), by thinking of them in terms of the possible *fact-prac worlds*, worlds whose content includes a statement of *how things are* as well as a statement of *what to do*, that are *allowed* or *ruled out* by these decisions.[23] We have seen that a world is *ruled out* by a decision if that decision rejects that world as a statement of *how things are* or *what to do*. For example, the decided state of mind corresponding to premise (1) of the argument above, "Either packing is now the thing to do, or by now it's

---

[23] Gibbard (2003) p. 47

18

too late to catch the train anyway," rules out all worlds in which the thing to do is not to now pack and in which it is the case that it is not now too late to catch the train anyway. In contrast, a decision *allows* a world just in case it does not by itself reject that world. For example, premise (1) allows both worlds in which "Grass is green," and worlds in which it's not the case that "Grass is green," because it does not itself rule out either of these sets of worlds.

Fact-prac worlds form an important part of Gibbard's attempt to offer a principled solution to the meanings of ethical sentences in the variety of contexts they appear. We can conceive of a maximal contingency plan, what Gibbard calls a *hyperplan*, a plan for every situation one might conceivably be in. These plans will be *complete* Gibbard claims, if for each situation and each alternative available in these situations the agent either rules out that alternative or rules out ruling out that alternative, which Gibbard says amounts to *permitting* that alternative. And we then we can think of an agent who accepts a hyperplan, an agent with cognitive (and non-cognitive) abilities far beyond ours, who is fully decided on all matters of what to do and how things are. This agent would be in a state of mind Gibbard calls a *hyperstate*. This apparatus of decided states, ascending from simple decisions of fact or plan like "Grass is green," or "Murder is wrong," all the way up to the complete decision of fact and plan that is the hyperstate form Gibbard's "device of automation," for answering the embedding problem, and for explaining the meanings of sentences.[24] Gibbard explains:[25]

> We can represent the meaning of a claim by asking in which such hyperstates a
> person would agree with it and in which she would disagree. Hyperstates, then,
> make for a canonical way to tally agreement and disagreement in the realm of
> judgments that bear on what to do.

Here's how this process is supposed to work. We start with a sentence whose meaning we want to explain. We then consider which hyperstates – states that can themselves be thought of as expressions of particular fact-prac worlds – would agree or disagree with

---

[24] Gibbard (2003) p. 83
[25] Gibbard (2003) pp. 140-141

the sentence in question. And then we use this matrix of agreement and disagreement at hyperstates to characterize the meaning of the sentence we're trying to understand. Suppose for example that we want to understand the meaning of the sentence "Grass is green." We'll start by tallying agreement and disagreement at hyperstates. The hyperstates that agree with this sentence will be all those that represent that grass is green, and the hyperstates that disagree with this sentence will be all those that represent that it's not the case that grass is green. We can now use this matrix of agreement and disagreement to help us toward the following explanation: The sentence "Grass is green," expresses the mental state that represents that grass is green. And using a similar process we can be led to the following explanation of the sentence "Murder is wrong": The sentence "Murder is wrong," expresses the mental state of planning not to murder for the case of being any possible agent in any possible situation.

Gibbard's attempted solution is particularly helpful for understanding the resources the non-cognitivist has for addressing the embedding problem. Because the non-cognitivist explains the meanings of sentences in terms of the thoughts they express, any solution to the embedding problem will require specifying the sort of thoughts expressed by each and every complex sentence that embeds a simple ethical sentence. Gibbard's apparatus of decided states that themselves are explained with reference to the fact-prac worlds they allow, rule out, or permit, allows us to come close to modeling each and every distinct thought the non-cognitivist might appeal to in explaining the meanings of these sentences. And this fact allows us to come close to appreciating every distinct meaning that the non-cognitivist might offer for each of these sentences. This illustration of fact-prac worlds thus delivers to us a very rich starting point for conceiving of the different kinds of thoughts the non-cognitivist might appeal to in order to explain the meanings of ethical explanatory sentences.

Chapter 3 – Non-Cognitivism Cannot Account for Ethical Explanation

3.1 Basic Problem Cases

Now that we have at least a basic understanding of how the non-cognitivist must respond to the embedding problem, and of the general resources available to them in presenting a solution to this problem, I want to move on to describing a set of basic problem cases that, if existing, would undermine such a solution. The force of these basic problem cases does not rest on any of the specific details of Gibbard's solution to the problem, but rather rests on the limitations of non-cognitivist resources generally. Nevertheless, the specific details of Gibbard's solution should help us in understanding how these cases are supposed to work. I will first describe these cases, before going on to argue for the conditional claim that *if* there exist basic problem cases, cases having the features I describe, the non-cognitivist cannot account for the meanings of ethical explanatory sentences.

Basic problem cases, I claim, are cases on which no possible non-cognitivist account of the meanings of the ethical explanatory sentences that appear in these cases could be satisfactory. I am thinking of cases like the following:

> *Socrates & Euthyphro* – Socrates and Euthyphro are two distinct agents who find themselves in astonishing agreement. Happily, their thoughts about the natural world, and all their non-ethical thoughts for that matter, are in complete alignment. In addition, their non-explanatory ethical thoughts are exactly the same; most importantly, they both concur that "One ought always do what all the gods love," and that "One ought always refrain from doing what all the gods hate." Nevertheless, there is one thing they do disagree about. Euthyphro thinks that "One ought always do what the gods love *because* the gods love it," while Socrates denies this. He's not entirely sure what does make certain actions the ones one ought to do, but he's definitely sure Euthyphro's answer is off the mark.

I claim this case has the following features: It consists of linguistically competent agents who agree about how things are non-ethically and on all ethical matters of what they

ought to do, yet disagree, ethically, about *why* they ought to do it. Importantly, I claim, the agents' agreement in this case about how things are (or how things might be) and what they ought to do extends to all possible worlds. Cases that include these features are *basic problem cases*. Whether this particular case does include all of these features, or whether any case could, we will examine in more detail in the following section. What I aim to defend here is the conditional claim that, *if* there exist basic problem cases, cases that include the features just enumerated, non-cognitivism cannot offer a satisfactory account of the meanings of the ethical explanatory sentences that appear in these cases. What I aim to defend, in other words, is that basic problem cases really are a problem for a non-cognitivist account of ethical explanation.

Assume that *Socrates & Euthyphro* is a basic problem case. If so, I claim, we can conclude the following:

> (1) Socrates and Euthyphro do not disagree ethically.
> (2) Socrates and Euthyphro do not disagree non-ethically.
> So: (3) Non-Cognitivism must explain Euthyphro's ethical explanatory claim as consisting of thoughts Socrates agrees with.
> But: (4) Socrates disagrees with Euthyphro's ethical explanatory claim while making no linguistic error.
> So: (5) Non-Cognitivism cannot explain Euthyphro's ethical explanatory claim.
> So: (6) Non-Cognitivism cannot explain ethical explanatory sentences.

Allow me to explain how each of these conclusions (and so the argument composed of these conclusions) follows from the assumption that *Socrates & Euthyphro* is a basic problem case. First, recall that on *non-cognitivism*, 1) ethical thoughts can be explained solely on the basis of the functional role these thoughts play in motivating an agent to act, and 2) all sentences (including ethical sentences) can be explained solely in terms of the thoughts those sentences express. This means that…

(1) Socrates and Euthyphro do not disagree ethically. Consider that Socrates and Euthyphro are fully agreed concerning what they ought to do, including what they ought to do in all possible worlds. And since non-cognitivism explains ethical thoughts,

22

including thoughts about what one ought to do, on the basis these thoughts play in motivating an agent to act, this agreement about what they ought to do suffices for agreement about what to do. Indeed, one of Gibbard's most famous slogans is "Thinking what one ought to do *is* thinking what to do." Socrates and Euthyphro have identical motivational profiles, since they agree about what to do in every possible case. But since they agree on what to do in every possible case, they do not disagree ethically. For if they had an ethical disagreement they would have a motivational disagreement. But they do not. Additionally…

(2) Socrates and Euthyphro do not disagree non-ethically. For Socrates and Euthyphro's only disagreement is an ethical explanatory disagreement. Socrates and Euthyphro agree about all non-ethical matters of fact. And as we've seen above, Socrates and Euthyphro have no ethical disagreements. This means they have no disagreements, and *this* means that…

(3) Non-Cognitivism must explain Euthyphro's ethical explanatory claim as consisting of thoughts Socrates agrees with. Whatever the meaning of Euthyphro's explanatory claim, it cannot include any thoughts that Socrates disagrees with, because, as we've seen, Socrates and Euthyphro have no disagreements. Except that…

(4) Socrates disagrees with Euthyphro's ethical explanatory claim while making no linguistic error. Socrates does not misunderstand that Euthyphro's explanatory claim merely consists of claims he already accepts. He understands that Euthyphro's ethical explanatory claim says something *different from* and *in addition to* the ethical and non-ethical claims he and Euthyphro already agree on. And he is correct in this understanding. And this means that…

(5) Non-Cognitivism cannot explain Euthyphro's ethical explanatory claim. For, as we've just seen, the only thoughts the non-cognitivist can appeal to in order to explain Euthyphro's ethical explanatory claim are thoughts Socrates already agrees with. But no account in this vein can square with Socrates' disagreement. And this means …

(6) Non-Cognitivism cannot explain ethical explanatory sentences. For the non-cognitivist is after a general account of the meanings of ethical explanatory sentences, and if they cannot account for the meaning of a given ethical explanatory claim and its

attendant sentence, then they cannot provide us with a general account of the meanings of ethical explanatory sentences.

If this argument is correct, than *if* there exist basic problem cases, the non-cognitivist cannot offer a satisfactory account of the meaning of ethical explanatory sentences. Since, as I demonstrated in section 1, a failure to account for the meanings of these sentences would result in a failure to account for ethical explanation, we can also say that *if* there exist basic problem cases, than non-cognitivism cannot account for ethical explanation.

3.2 Gibbard's Account of Ethical Explanatory Sentences

In addition to offering a general solution to the embedding problem, Gibbard also offers a particular account of the meanings of ethical explanatory sentences. Because his particular account is instructive for demonstrating why all non-cognitivist accounts of the meanings of ethical explanatory sentences result in basic problem cases, in this section I take up his particular account. Ultimately I will use the details of his account to construct a new kind of basic problem case, before we move on to the final section of the paper to examine possible non-cognitivist responses to these cases.

Gibbard claims that the thought expressed by an ethical explanatory sentence is a particular hyperplan, a plan for what to do in every possible scenario. He illustrates this by drawing a picture of a particular hyperdecided agent, Hedda, who, in addition to being fully decided about *what to do*, is also fully decided about *how things are*:

> "Hedda, our egoistic hedonist, thinks that, in any possible situation, all and only acts that maximize one's hedonic prospects are okay to do. In this sense, she thinks that maximizing one's hedonic prospects and being okay to do are coextensional. She also thinks that an act's being okay to do depends, explanatorily, on its maximizing one's hedonic prospects. (The explanations in this case are not purely causal explanations; they are explanations of why to do

24

things or not.) In her view, then, maximizing one's hedonic prospects *constitutes* being okay to do in a way that roughly parallels the case of H_2O and water."[26]

This passage is puzzling for two reasons. Firstly, Gibbard claims that Hedda "also" accepts an explanatory claim, which seems to imply that this claim is somehow *different from* and *in addition to* her hyperplan. Secondly, Gibbard's claim about constitution in ethics paralleling constitution in science looks suspiciously like a factual claim. But, non-cognitivism is supposed to be able to explain ethical thought without appeal to facts. Fortunately, Gibbard later clarifies both puzzles in a way that is consistent with his overall commitments. Accepting Hedda's view that maximizing one's hedonic prospects constitutes being okay to do, Gibbard claims, just amounts to thinking the following, which he identifies with the constitution claim – that there is a prosaically factual property, the property of being egohedonic, such that for any act open in any possible situation *s*, act *a* is okay to do in *s*, just in case *a* in *s* has the property of being egohedonic.[27] And this claim is precisely a statement of Hedda's hyperplan. Just so there is no doubt on this point, Gibbard later puts it in the following way:

> "If the property of being egohedonic realizes the concept of being okay, that is a matter of how to live. Accepting this claim consists in accepting the hedonistic egoist's hyperplan."[28]

We can see that accepting an explanatory claim is not *different from* or *in addition to* accepting a hyperplan, and the parallel with constitution in science is indeed rough.

Gibbard's view that ethical explanatory sentences express hyperplans results in two very odd consequences. Firstly, Gibbard must deny that any ethical explanatory sentence that expresses a plan *short* of a hyperplan is genuinely explanatory. This means that a sentence like "Lying is wrong because lying is willful deception," is not in fact explanatory, on Gibbard's view. Secondly, this commits Gibbard to the claim that the

---

[26] Gibbard (2003) p. 95 Down the page Gibbard says, "I'll focus on necessary coextensionality, and then just note informally that a kind of explanatory dependence seems to obtain."
[27] Gibbard (2003) p. 96
[28] Gibbard (2003) p. 116

meaning of the sentence "Right acts are right because they maximize one's hedonic prospects, is the same as the meaning of the sentence "Right acts are those that maximize one's hedonic prospects." In other words, Gibbard must claim that every explanatory sentence in ethics *means the same thing* as some other ordinary predicative ethical sentence. This alone is enough to falsify Gibbard's account, but it is worth pointing out that his account also results in a particularly interesting basic problem case:

> *Hyperdecider Disagreement*: Hedda is hyperdecided. She has a view about *how things are* in terms of non-ethical matters of fact, and she has a view about *what to do* in response to these facts. She plans to always maximize one's hedonic prospects. She also accepts an explanatory claim – "Right acts are right *because* they maximize one's hedonic prospects." Hyper is also hyperdecided and agrees with Hedda's hyperplan, her plan for what to do in every possible situation. But he denies the explanatory claim she accepts, thinking "It is not the case that right acts are right because they maximize one's hedonic prospects." Instead he thinks that each explanation of an act's being the thing to do is hyper-particular and requires mention of every non-ethical feature the act has. As it turns out, Hyper claims, Hedda's hyperplan is correct. But her explanatory claim is not. Zeus too is hyperdecided, and agrees with Hedda's hyperplan. But he disagrees with both Hyper's and Hedda's explanatory claims. In fact, he disagrees with all explanatory claims and is very puzzled by their purpose. He adopts a slogan about ethics: "Thinking what one ought to do *is* thinking what to do." Once you've decided what to do, Zeus claims, there's nothing more to decided ethically. This talk of explanation adds nothing.

3.3 Demonstrating that Non-Cognitivism Results in Basic Problem Cases

Non-Cognitivism, I have claimed, results in basic problem cases with respect to an account of ethical explanatory sentences. These cases feature linguistically competent agents who agree about how things are non-ethically and on all ethical matters of what they ought to do, yet disagree, ethically, about *why* they ought to do it. Furthermore, this agreement extends to all possible worlds. I have already argued that if there exist basic

problem cases non-cognitivism cannot account for ethical explanatory sentences. In this section I aim to demonstrate that *Hyperdecider Disagreement* is a basic problem case, though what I say should translate to *Socrates & Euthyphro* as well. I will proceed by defending each premise of the argument in 3.3 and considering whether the non-cognitivist can deny this premise. Although I will stick to talking of an agent's plans for reasons of simplicity, what I say should translate straightforwardly to any other kind of motivational thought the non-cognitivist might want to replace plans with. By the end of the argument, it should be clear that non-cognitivism cannot account for ethical explanatory sentences and so ethical explanation.

      (1) Hedda, Hyper, and Zeus do not disagree ethically.

      In order for the non-cognitivist to deny this premise, he will need to claim that Hedda, Hyper, and Zeus do not share all of the same plans. He will need to claim that Hedda and Hyper accept some plan over and above the plans of Zeus, and he will need to claim that Hedda and Hyper disagree with one another ethically with regard to the plan each of them accepts over and above the plans of Zeus. Now, it is clear that each of these agents plans to maximize one's hedonic prospects in any possible situation. So what additional sort of plan might Hedda and Hyper accept? One attractive option, is that both Hedda and Hyper accept some sort of higher-order plan that Zeus does not, perhaps a plan to plan in a certain way, a plan to be guided by different sorts of features. It is the acceptance of this additional sort of plan, one might think, that constitutes the difference between Hedda and Hyper on the one hand, and Zeus on the other; it is the acceptance of this additional sort of plan that constitutes accepting a claim about *why* to do something on the one hand, rather than merely accepting claims about *what* to do on the other.

      Now, in order for this to be a plausible answer, Hedda, Hyper, and Zeus are going to have to each act differently from one another in at least some situations (or some possible situations). This is because, according to non-cognitivism, ethical thoughts are explained in terms of the functional role these thoughts play in motivating an agent to act. If these three agents do not act differently from one another in *any* cases, this must mean that their motivational profiles are the same. And if this is the case there will simply be no distinct ethical thoughts to appeal to in order to explain the content of the additional

higher-order plans that Hedda and Hyper are supposed to accept; there will simply be no distinct ethical thoughts to appeal to in order to explain the difference in plans between each agent.

One sort of obvious difference between each agent is a difference about what each agent says in various cases. For instance, Zeus will not be inclined to say anything about *why* a particular act is the thing to do, while Hedda and Hyper will say different things concerning why a particular act is the thing to do. Nevertheless, it is crucial to recognize that this sort of difference, a *mere* difference about what to say in various cases, cannot itself be grounds for a difference in plan. For one thing, Gibbard himself is explicit about *mere* verbal disputes being the result of confusions rather than being true ethical disagreements. Speaking of two agents who agree about what to do in terms of lived practical conclusions, but *appear* to disagree about the rightness of what they do *only* in terms of what they say, Gibbard says this:[29]

> I myself think there's not much difference between them: they agree remarkably in thinking their way to decisions, and disagree only on what to say about it and on the words with which to think about it. Their disagreement is verbal; they disagree on what words to mouth. They have no serious difference between them on what to do and why.

The non-cognitivist cannot make mere verbal disputes into genuine ethical disagreements in these sorts of cases, because it is clear that when two agents disagree about an ethical explanatory claim, this disagreement is about something more than what words to use (at least from the perspectives of the agents). These are disagreements about *why* to do a certain thing. Because of this, in order to claim that ethical explanatory disagreement is legitimate ethical disagreement (and not mere verbal disagreement), the non-cognitivist must find something more for this dispute to be about than simply what to say.

One could be forgiven for thinking that the source of the difficulty here must be that up to this point the non-cognitivist has adopted too restricted an understanding of what distinguishes different plans that has rendered him unable to make the crucial

---

[29] Gibbard (2003) pp. 12-13

distinctions between plans he needs in order to explain the content of ethical explanatory claims. After all, Gibbard's "device of automation" for accounting for sentence meaning – the apparatus of decided states understood by reference to ruled out fact-prac worlds – only distinguishes plans intensionally, by the extension of the properties these plans concern at possible worlds. As Gibbard says:[30]

> … a plan can distinguish between situations only in terms of the prosaically factual properties of those acts. If two acts in two possible situations differ in no prosaically factual way, a plan can't distinguish them, permitting one and ruling out the other.

This understanding of plans as being distinguished solely by the properties they concern *does* unnecessarily limit the content of plans. Plans are, after all, thoughts, and thoughts can be different even if these thoughts concern the same things (and thus have the same intension). For example, my thoughts "Water is H_2O," and "Water is water," are different, even though the things these thoughts concern (water and H_2O) are precisely the same. Similarly, my plan to drink H_2O and my nephew's plan to drink water are different plans even though the properties that these plans concern are *necessarily* the same. What is needed to account for the content of ethical explanatory plans, it might be thought, is to distinguish plans hyperintensionally, thereby allowing that necessarily cointensive plans may nevertheless be different. The thought that this is what is needed to solve the non-cognitivist's problem of the content of ethical explanatory claims is especially appealing, because explanatory claims are known to be hyperintensional. For example, although the sentences "That snow is white is true because snow is white" and "Snow is white because snow is white" are co-intensive (they both have the same truth value in every possible world), these sentences do not mean the same thing, and only the first sentence is capable of being genuinely explanatory.[31]

Making these new distinctions does allow us to make some progress on behalf of the non-cognitivist. For instance, the non-cognitivist can now explain Hedda and Hyper's

---

[30] Gibbard (2003) p. 92

[31] The second sentence is incapable of being genuinely explanatory (at the very least) because no object of explanation can be explained by itself.

disagreement: Hyper, despite appearances, actually denies Hedda's plan to maximize one's hedonic prospects. Although Hyper *ends up* maximizing his hedonic prospects as a result of carrying out his plans, his plans (being hyper-particular) are actually something like "Do a_1 in circumstances $x_1, y_1, z_1...$" and "Do a_2 in circumstances $x_2, y_2, z_2 ...$" and so on. We can now see that Hedda and Hyper really are guided by different sorts of concepts. Our initial suggestion on behalf of the non-cognitivist – that Hedda and Hyper accepted different higher-order plans about what to be guided by – thus has the appearance of being vindicated. Unfortunately for the non-cognitivist, this is merely an appearance of vindication. For although these new distinctions have allowed us to distinguish Hedda's and Hyper's plans, we are still not in a position to distinguish Hedda and Zeus's plans. Indeed, we can now see that the difference between Hedda and Hyper, far from being a difference of higher-order plan, is simply a difference of first-order plan. Hedda accepts the plan, "Maximize one's hedonic prospects," while Hyper accepts a set of plans far more particular. But neither of these plans is anything like the higher-order plan imagined, the kind of plan which would allow us to distinguish between Hedda and Zeus.

All of this brings us to the core of the problem for the non-cognitivist, at least as far as this premise is concerned. The non-cognitivist cannot distinguish between Hedda's and Zeus's plans because the non-cognitivist cannot distinguish between an agent's plan *to* act in a certain way and the reason *why* an agent acts a certain way. The non-cognitivist cannot distinguish between thinking that one ought to maximize one's hedonic prospects, and thinking that those acts that one ought to do, one ought to do *because* they maximize one's hedonic prospects. The non-cognitivist cannot make sense of an ethical *why* at all, and this is because the non-cognitivist must explain *why* an agent acts a certain way in terms of *what* an agent does in acting.[32] Recall, once again, that the non-

---

[32] Gibbard (2003) pp. 188-191 explains an agent's acting for a reason in terms of what an agent does in acting: "What, then, is this purported state of mind, weighing factor R in favor of doing X? It is calculating what to do on a certain pattern, a pattern we could program a robot to mimic. Let the robot code aspects of its circumstances (factors), and code alternative movements that it is wired up to have emerge from its calculations (acts)…The robot, imagine, attaches number representations (call them 'indices'), positive and negative, to factor-act pairs. It then totals up the indices for each act, and performs the act with the highest resulting sum. If the robot is set up in this way, then the index it attaches to factor-act pair R,X then constitutes the degree to which it weighs factor R toward doing X. We ourselves can settle what to do in a like way, not toting up numbers explicitly, but proceeding as if we did. When we do, say I, we are *weighing*

cognitivist explains ethical thoughts in terms of the role these thoughts play in motivating an agent to act. This means that ethical thoughts are both simultaneously the *why* of action and the *what*. Thoughts about what to do serve as the reason why an agent acted and they simultaneously determine what an agent is to do.

Consider an agent who decides not to lie. She may decide not to lie for a variety of reasons, but let us suppose she decides not to lie *because* lying is willful deception. How can the non-cognitivist understand this decision? Presumably, the non-cognitivist will say that she had a plan not to willfully deceive, she recognized that lying is willful deception, and that she therefore decided not to lie for the reason that lying is willful deception. But notice what is going on here: the *because* of the agent's reason is understood as a plan *to* do something, a plan not to willfully deceive. The agent will say that her reason for not lying is the fact that lying is willful deception; this fact explains why the act is wrong. Her thoughts here are many. She may think "That act is a lie," "Lying is willful deception," "Willful deception is wrong," and so "I won't lie." But what is the *ethical* content of all of these thoughts? Is it not exhausted by claims about *what* she will do, namely *that* she will not willfully deceive and *that* she will not lie? The ethical content *is* exhausted by these thoughts. Do, or do not. For the non-cognitivist there is no ethical why. Once Hedda and Zeus agree about *what* to do in all possible cases, there is simply nothing *ethical* for them to disagree about. Because of this, the non-cognitivist cannot explain their disagreement in this case as one of ethical disagreement. Hedda and Zeus, on our amended first premise, do not disagree ethically.

(2) Hedda, Hyper, and Zeus do not disagree non-ethically.

We have just established that Hedda and Zeus, at the very least, do not disagree ethically. But perhaps Hedda's explanatory claim, in addition expressing an ethical thought Zeus agrees with, also expresses some non-ethical thought Zeus disagrees with and that Hyper might disagree with as well.[33] In that case the non-cognitivist could

considerations. Regarding features of our situation as reasons to do one act as opposed to another, my theory is, consists in such weighing."

[33] It seems implausible to me to suggest that ethical explanatory claims do not express *any* ethical thoughts. After all, accepting one of these claims commits one to accepting the ethical claim embedded in the antecedent, and ethical explanatory claims do appear capable of playing a functional role in motivating the

account for the disagreement between them as a non-ethical disagreement, and could account for the meanings of ethical explanatory sentences by (in part) citing the non-ethical thoughts these sentences express. The non-cognitivist can deny this second premise, in order words, by finding some non-ethical thought that ethical explanatory sentences might plausibly express.

What sort of non-ethical thought might plausibly account for Hedda, Hyper, and Zeus' disagreement? I confess that I am at a loss as to what sort of non-ethical thought the non-cognitivist might possibly appeal to explain this disagreement. Admittedly, this inability does not amount to an argument. But there are perhaps a few things I can say to motivate this premise. Firstly, in *every other case* where there is a disagreement that concerns a "Why" question, this disagreement is a disagreement that concerns the domain of the object of explanation this question seeks to demystify. For instance, disagreements about answers to "Why" questions where the object of explanation is scientific, are scientific disagreements. Disagreements about answers to "Why" questions where the object of explanation is historical, are historical disagreements. And so on. In the absence of some compelling reason for thinking that ethical "Why" questions behave differently, I think it is reasonable to assume that they behave the same.

Secondly, both Gibbard and Blackburn have historically treated disagreements about answers to ethical "Why" questions as ethical disagreements. We have already seen that Gibbard thinks that the disagreements in these cases are disagreements concerning maximally specific plans (or hyperstates). Here is Blackburn:[34]

> If you say, for instance, "If an act creates happiness then it is good," I will understand you well enough: this is voicing a certain standard, and acknowledging that standard means being disposed to value things on the basis that they create happiness. And values we already have under control.

We can see that this passage implicitly commits Blackburn to understanding disagreements about answers to ethical "Why" questions as ethical disagreements, since

---

agent to act. Nevertheless, if one still has doubts what I say in this section should apply straightforwardly to the claim that ethical explanatory sentences express only non-ethical thoughts.

[34] Blackburn (1998) p. 67

he understands answers to ethical "Why" questions as expressing ethical thoughts. Once again, this point is hardly conclusive, and I am certainly not proposing that we disallow philosophers from altering their positions when difficulties with their views emerge. Nevertheless, both of these points sustain the following thought – that understanding these disagreements as non-ethical disagreements is ad hoc, and therefore requires some justification. In the absence of such justification, we ought to accept this premise. If this is correct, this means that…

(3) Non-Cognitivism must explain Hedda's ethical explanatory claim as consisting of thoughts that, at the very least, Zeus agrees with.

Unfortunately…

(4) Hyper, Hedda, and Zeus disagree with one another while making no linguistic error.

Each of these agents understands that ethical explanatory claims are something *different from* and *in addition to* the ethical and non-ethical claims that each of them already agree on. Because this is the last premise of the argument that the non-cognitivist can plausibly disagree with, the non-cognitivist must claim that at least one of these agents, makes some linguistic error. Since we have already seen that the non-cognitivist can make sense of the dispute between Hedda and Hyper, it is clear that he must claim that either Zeus or Hedda is the one who errs in this case. But the problem for the non-cognitivist is that no such charge can be sustained, since such a charge must be accompanied by an explanation of exactly what error Zeus or Hedda is guilty of making.

On the one hand, the non-cognitivist must claim that Zeus makes some error. This is because Zeus' claim is that ethical explanatory sentences are meaningless, but we do in fact use such sentences in our actual language. Since non-cognitivism aims to account for the meanings of sentences in our actual language (rather than be an error theory on this language) the non-cognitivist must say that Zeus is the one who is confused here. This means that Zeus' mistake must be that he is already committed to some explanatory claim, because that explanatory claim expresses thoughts that he already holds. But the non-cognitivist is not able to give a distinct account of these explanatory claims (as we've seen) because the non-cognitivist is unable to distinguish the ethical thought *that* something is wrong from the distinct ethical thought *why* something is wrong. This

supports Zeus' claim that (on non-cognitivism) these ethical explanatory thoughts are meaningless because they appear to suggest a distinction between these different sorts of ethical thoughts when there in fact is none. And this means that no charge of error on Zeus' part can be sustained.

This means that…

(5) Non-Cognitivism cannot explain Hedda's or Hyper's ethical explanatory claim. For, as we've just seen, the only thoughts the non-cognitivist can appeal to in order to explain Hedda and Hyper's ethical explanatory disagreement are thoughts that an agent (like Zeus) might accept even in the absence of acceptance of *any* ethical explanatory claim. But no account in this vein can square with the group's genuine ethical explanatory disagreement. And this means …

(6) Non-Cognitivism cannot explain ethical explanatory sentences. For the non-cognitivist is after a general account of the meanings of ethical explanatory sentences, and if they cannot account for the meaning of a given ethical explanatory claim and its attendant sentence, then they cannot provide us with a general account of the meanings of ethical explanatory sentences.

Because the non-cognitivist cannot account for ethical explanatory sentences, we can also now say that they cannot account for the explanatory thoughts expressed by these sentences, since the non-cognitivist must account for the meanings of these sentences solely in terms of the thoughts these sentences express. And since, as we've seen in section 1.2, ethical explanation (the process by which we attempt to explain ethical features) requires these explanatory thoughts, we can see that the non-cognitivist cannot account for ethical explanation as well. We are finally in a position to answer the initial question of this paper: Non-Cognitivism cannot account for ethical explanation.

REFERENCES

Ayer, A.K., (1952), *Language, Truth, and Logic* (New York: Dover Publications)

Barnes, W. H F (1933), "A Suggestion About Value," *Analysis*, 1: 45-46.

Blackburn, Simon (1998), *Ruling Passions* (Oxford: Clarendon Press).

Bratman, Michael (2006) "'Thinking How to Live' and the Restriction Problem," *Philosophy and Phenomenological Research,* Vol. 72, No. 3, pp. 707-713.

Correia, Fabrice and Schnieder, Benjamin, eds. (2012), *Metaphysical Grounding* (Cambridge: Cambridge University Press).

Díez, José and Khalifa, Kareem and Leuridan, Bert (2013), "General Theories of Explanation: Buyer Beware," *Synthese* 190, (3), 379-396.

Dreier, Jamie (2004), "Meta-ethics and the Problem of Creeping Minimalism," *Philosophical Perspectives*, 18, (1), 23-44.

Dreier, Jamie (2006), "Disagreeing (About) What to Do: Negation and Completeness in Gibbard's Norm-Expressivism," *Philosophy and Phenomenological Research*, Vol. 72, No. 3, pp. 714-721.

Fine, Kit (2012), "Guide to Ground", in Correia and Schnieder, eds. (2012), 37-80.

Geach, P.T. (1965) "Assertion" *Philosophical Review*. 74: 449-465.

Gibbard, Allan (2003), *Thinking How to Live* (Cambridge: Harvard University Press).

Gibbard, Allan (2006), "Précis of 'Thinking How to Live'," *Philosophy and Phenomenological Research*, Vol. 72, No. 3, pp. 687-698.

Gibbard, Allan (2006), "Reply to Critics," *Philosophy and Phenomenological Research*, Vol. 72, No. 3, pp. 729-744.

Haidt, Jonathan,Fredrik Bjorklund, and Scott Murphy (2000), "Moral Dumbfounding: When Intuition Finds No Reason." Unpublished manuscript.

Haidt, Jonathan (2012), *The Righteous Mind: Why Good People Are Divided By Politics and Religion*, (New York, Pantheon Books)

Hare, R. M., (1952), *The Language of Morals*, (Oxford: Clarendon)

Mackie, J.L. (1977) *Ethics: inventing right and wrong*. (Hardmondsworth, Penguin)

Nickel, Bernhard (2010), "How General Do Theories of Explanation Need to Be?", *Noûs* 44, (2), 305-328.

Rosen, Gideon "Metaphysical Dependence: Grounding and Reduction", in *Modality: Metaphysics, Logic, and Epistemology*, ed. B. Hale and A. Hoffman (Oxford:

Oxford University Press), 109-36.

Scanlon, T. M. (1998) *What We Owe to Each Other* (Cambridge: Harvard University Press)

Scanlon, T. M. (2006) "Reasons and Decisions," *Philosophy and Phenomenological Research*, Vol. 72, No. 3, pp.722-728.

Schroeder, Mark (2008), *Being For: Evaluating the Semantic Program of Non-Cognitivism* (Oxford: Clarendon Press)

Searle, J., (1962), "Meaning and Speech Acts," *Philosophical review*, 71: 423-32.

Stevenson, C., (1937), "The Emotive Meaning of Ethical Terms," *Mind*, 46:14-31.

Strevens, Michael (2008), *Depth: An Account of Scientific Explanation* (Cambridge: Harvard University Press).

Zangwill, Nick (2011), "Non-Cognitivism and Consistency," *Zeitschrift für philosophische Forschung,* 65, (4), 466-484.

Väyrynen, Pekka (2013), "Grounding and Normative Explanation," *Aristotelian Society Supplementary Volume* 87, (1), 155-178.