Cognitivism, Expressivism and Hybrid Expressivism

*How to Solve The Moral Problem*

Marcus Bøhn

Thesis presented for the degree of

MASTER OF PHILOSOPHY

Supervised by

Associate Professor Sebastian Watzl

Department of Philosophy, Classics, History of Art and Ideas

Faculty of Humanities

UNIVERSITY OF OSLO

Spring 2016
Cognitivism, Expressivism and Hybrid Expressivism

How to Solve The Moral Problem
Cognitivism, Expressivism and Hybrid Expressivism

*How to Solve The Moral Problem*

Marcus Bøhn
Abstract

This thesis investigates whether an Expressivism or Hybrid Expressivism account is better suited to solve “The moral problem” than cognitivist theories. “The moral problem” is the unified account of 1) the practicality requirement – a satisfactory account of moral psychology and motivation and 2) a satisfactory metaphysical and epistemological account of the objective features of morality, the objectivity requirement. I will argue that Expressivism accommodates motivational issues and the seemingly sentimental base of moral judgements better than cognitivism, and also that the theory of moral evolution seem to favour anti-realism and noncognitivism. I argue that hybrid versions of expressivism may be seen as advancing the quasi-realist project by allowing moral sentences to express belief-like states of mind in addition to desire-like states of mind, in which the belief component have propositional contents. There still remain serious unresolved issues for any expressivist – the Frege-Geach problem in particular.
Acknowledgements

First I would like to thank my supervisor, Sebastian Watzl, who has offered continuous support, insightful comments, objections and suggestions.

I would also like to thank the other members of our informal metaethics reading group – Conrad Bakka, Ainar Petersen Miyata, Martin Ravneberg, Øyvind Sætre Strøm, Sebastian Watzl and, sometimes, Caj Strandberg – for countless hours of metaethical discussion.

Oslo, juni 2016
Marcus Bøhn
# Table of contents

## 1. Cognitivism versus Noncognitivism

1.1. Introduction – Metaethics and the Natural Sciences 12  
1.2. Introducing Cognitivism and Noncognitivism 15  
  1.2.1. Cognitivism 17  
  1.2.2. Noncognitivism 18  
1.3. Early noncognitivism – Emotivism 18  
  1.3.1. Speaker Subjectivism 20  
1.4. Moral motivation 22  
  1.4.1. The Humean Theory of Motivation 23  
  1.4.2. The Amoralist 25

## 2. Cognitivism and The Moral Problem

2.1. Moral Realism 31  
2.2. Supervenience 33  
2.3. Cornell Realism 35  
  2.3.1. The Open Question Argument 35  
  2.3.2 The Revised Open Question Argument 37  
  2.3.3. Hare’s Cannibal Objection 38  
2.4. Smith’s Solution to The Moral Problem 39  
  2.4.1. A Counterexample 42

## 3 Expressivism

3.1 Why Expressivism? 45  
  3.1.1. The parity thesis 47  
  3.1.2. Minimalism about moral truth 48  
  3.1.3. Projectivism and Quasi-Realism 49  
  3.1.4. The Subjectivism objection 51  
3.2. Expressivism and Moral Evolution 52  
  3.2.1. The Evolution of Morality 52  
  3.2.2. The No-Missing Link Account of Moral Evolution 55  
  3.2.3 Debunking arguments 57
4 Hybrid expressivism

4.1. Hybrid Cognitivism versus Hybrid Expressivism

4.1.1. Hybrid Cognitivism

4.1.2. Hybrid Expressivism

4.2. The Frege-Geach Problem

4.3. Concluding remarks – Results and Unresolved Issues

4.3.1. Disagreement

4.3.2. The Moral Problem

4.3.3 Summary

References
Chapter 1

Cognitivism versus Noncognitivism

1.1. Introduction – Metaethics and the natural sciences

The most important metaethical question has been to settle whether there is an objective moral\(^1\) reality, if there is something that is genuinely right or wrong, good or bad, that amounts to something that is more than personal opinions or cultural oddities. If I believe that it is wrong to commit murder, could there be some state of affairs, properties or features of the world, independent of my personal feelings or thoughts on the matter, that could make this belief true? Another question follows immediately, that if a moral belief could be true, what kind of truth would it be? Could something be true for me, and something else be true for an indigenous tribe of Amazonian warriors, or would one of us be mistaken? Since the time of ancient Greece philosophers have battled these questions. Is there one unified story of objective moral truths, like Plato thought, or is what is true a matter of perspective, and different for different persons, like Protagoras thought?

The question concerning the nature of morality has, in turn, generated many associated questions and different approaches in the attempt to find an answer. One important challenge is to fit one’s ethical and metaethical theories into a bigger scientific picture of the world. By “naturalism” or “a naturalistic world view” I shall simply mean a world view based on generally uncontroversial scientific beliefs shared by the majority of the scientific community working in the fields of the natural sciences\(^2\), and where, to quote Frank Jackson, the world’s

---

\(^1\) The scope of this thesis will be the field of morality, and in agreement with most metaethicists, I will assume morality to be a sub-field of normativity – the field of “oughts” and “shoulds”. There are vast areas of language, situations and actions that are normative, but not moral. Consider for instance what we call “instrumental normativity”, which describes the conditions of means to reach some end, concerning every little possibility of instrumental reasoning, of the form “If S wants to obtain A, S should do B”. For example, if I want to get to work on time, I should get up before 7 o’clock.

\(^2\) This formulation is based on a definition of “nature” by Moore, as “that which is the subject-matter of the natural sciences”. (Moore 1903: 40)
“physical nature exhausts all its nature.” (Jackson 1998: 11) If we were to believe in the existence of moral properties and entities at odds with accepted theories of the natural sciences, it could weaken our beliefs, and we should consider revising our ethical theories to make them materially respectable. There are several interesting nonnaturalist metaethical positions that are challenging to reconcile with a naturalistic world, but they fall outside the scope of this thesis. ³

The continuously increasing importance of the natural sciences has both moved us towards holding moral metaphysical views that easily integrate into a larger naturalistic world view, but also raised questions about our ability to correctly grasp or access potential moral aspects of our world. For a long time there was a significant gap between the methodology of metaethics and the methodology of the natural sciences. Most metaethical inquiry started with metaphysics, which in earlier times meant fewer scientific experiments and more thinking. Since the time of Plato, metaethicists have constructed grand theories about the nature of objective morality, often based on little evidence other than metaphysical contemplation. However, since the beginning of the last century or so we have seen a new methodological dawn, in which the complexity and preciseness of the methods of the natural sciences have influenced how we do metaphysics.

Metaethical theories can be divided into those that attempt to accommodate naturalism, *metaphysical naturalism* and those that do not – nonnaturalistic theories. But they can also be divided into theories that claim that the methods applied in the natural sciences can be used to discover moral truth, *methodological naturalism*, and those that claim that we must use other methods for moral inquiries:

We can distinguish two broad trends in contemporary moral theory depending upon how “the problem of placing ethics” is identified and faced, and the implications drawn. The first starts out with the idea that the “problem” is a product not of ethics, but of the wrong-headed notion of seeking to understand the objectivity of moral judgements on the model of the objectivity of empirical science. […] The second broad trend in response to “the problem of placing ethics” accepts the challenge of showing that moral judgements are factual in the paradigm sense afforded by empirical or theoretical judgements in the natural sciences. (Darwell, Gibbard and Railton 1997: 9)

³ But I will discuss some important objections nonnaturalists have raised against noncognitivists. Prominent nonnaturalistic forms of realism have been defended by philosophers such as G. E. Moore (1903), John McDowell (1998), David Wiggins (1998, 2006), David Parfit (2011), and David Enoch (2011).
This difference of outlook could be described as having different methodological approaches to ethics, motivated by whether one believes morality to be an “an area of genuine and objective inquiry” (ibid.) comparable to the natural sciences, or one believes morality to be inaccessible by the methods of the natural sciences, and that moral inquiries demand other tools than what is demanded for inquiries in the natural sciences.

One of the important questions that have dominated the metaethical debate since the time of the Vienna circle and the logical positivists, is how the study of the language we use to express our moral convictions and judgements could help us understand the nature of morality itself. This has often been called the linguistic turn, and has largely reformed the methodology of metaethics. Developments within the philosophy of language have for some become an integrated part of the methodological naturalist approach, while these developments have led others to depart from the naturalist approach.

1.2. Introducing Cognitivism and Noncognitivism

We can roughly divide the metaethical landscape into two camps, initially based on two different theories of moral psychology and language. One side follows in the footsteps of Hume, and believes moral thought to be “the slave of the passions”, where moral judgements are grounded in our emotions, feelings or sentiments. The other side holds that moral thought seeks to capture something objectively true about right and wrong, where moral judgement is not based on emotions or feelings, but on faculties of reason. The way we understand the sentence “It is wrong to commit murder”, what kind of mental state it expresses, and whether this expression of a mental state relates to some external aspect of the world, would roughly place the proposed set of answers on one of two metaethical branches – noncognitivism or cognitivism. Although many metaethicists currently rely heavily on the analysis of moral language to develop their respective theories, we could say that many cognitivists are also methodological naturalists with regard to the discovery of moral truths, where most

---

4 Hume, David (1739: SB 415).
5 When focusing on the psychological aspects of these views, they are often labelled moral sentimentalism, where Adam Smith was another early inspirational source. (See Slote 2010).
noncognitivists would believe this to be futile, as there are no moral truths, or if there might be (as many later noncognitivists hold), they are not apt for scientific discovery.

The contemporary landscape of metaethics consists of a myriad of different, similar and overlapping views belonging to the cognitivist and noncognitivist camps. These two theoretical clusters could best be understood as different combined theories of moral language, psychology, metaphysics and epistemology. Their differences become clear in how they understand sentences like “It is wrong to commit murder”, “It is good to give to famine relief” or “Abortion is wrong”, what kind of mental states these sentences express, how they could be true, and how we could obtain knowledge of this truth.

We could illustrate the typical build-up of contemporary metaethical theories by expanding the four theoretical categories with a few common interrelated questions:

**Moral language**
What is the nature and structure of moral language?
What is the meaning of a moral sentence like “It is wrong to commit murder”?
How do sentences function to express moral judgements?

**Moral Psychology**
What is the nature of moral judgements?
How do moral judgements seem to have motivating force?

**Moral metaphysics**
Does morality exist in any objective sense?
If there are moral facts and truths, how do they relate to natural facts?
If there are moral facts, do they exist independent of our opinions of them?

**Moral epistemology**
If there are moral facts, how can we obtain knowledge of them?
How do we explain that people make different moral judgements and disagree about what is morally right?
Although there has been constant development in how these questions are answered, and considerable disagreement at all times about the answers, we can offer some unified claims about each metaethical camp, which identifies the most important feature of each group of cognitivist and noncognitivist theories.

1.2.1. Cognitivism

If I believe that it is wrong to commit murder, cognitivists hold that I am in a cognitive state of mind, a similar state of mind I am in if I believe that grass is green. I have an ordinary belief in both cases. The sentence “It is wrong to commit murder” is used to express a moral belief or moral judgement. And just as our moral beliefs work in the same way as ordinary non-moral beliefs, the moral sentence “It is wrong to commit murder” works in the same way as the non-moral sentence “grass is green”. They express beliefs about a subject matter. This is the one claim that all cognitivists share, often called “descriptivism”: Our moral language functions like ordinary descriptive language, in so far as moral sentences express beliefs that are fundamentally similar to ordinary beliefs. Since ordinary beliefs can be true or false, cognitivists therefore think that moral beliefs are truth-apt as well. Many cognitivists are also moral realists. That is, they hold that moral beliefs, like non-moral beliefs, are often true, which assumes the existence of moral properties and facts. Most realists also hold moral facts to be mind-independent, which means that they are true or false independent of what our beliefs about them are. There are also positions that could be considered realist, where moral facts are dependent on human psychology or practical reasoning, but still independent of any one person’s actual moral beliefs (Smith 1994: 187; Korsgaard 1996: 160; Foot 2001: 18; Copp 2007: 110). As we shall see, realists do not always agree on what kind of entities moral properties and facts are, or how they relate to our judgements about them or the language we use to describe them. Some realists believe moral properties can be reduced to non-moral natural properties. (Railton 1986: 165; Jackson 1998: 146–147; Schroeder 2007: 198; Finlay 2014: 5) Other realists believe moral properties are irreducible, and are a special species of natural properties. (Sturgeon 2006 [1985]: 126; Boyd 1986: 106; Brink 1989: 160) If a cognitivist is an anti-realist, and does not believe in moral truths, he is an error-theorist, and holds that moral judgements and discourse aim to represent moral facts, but they do not exist. (Mackie 1977; Joyce 2006)
1.2.2. Noncognitivism

If I believe that it is wrong to commit murder, noncognitivists, on the other hand, hold that I am in a conative state of mind. In this mental state I do not really believe anything in an ordinary descriptive way, even though the language I use to express my moral beliefs makes it seem so. Instead, to make a moral judgement is to be in an emotional or desire-like mental state. Noncognitivists disagree, however, about what sort of mental state this is. It could be a state of disapproval towards murder (Ayer 1936: 111; Stevenson 1944: 16; Blackburn 1984: 16, 1993: 18, 1998: 66), a state of commanding or prescribing not to murder (Ayer 1936: 111; Hare 1952: 155), a state of accepting a norm that murder is wrong (Gibbard 1990: 7, 2003: 7), or a state of planning not to murder (Gibbard 2003: 195). This judgement does not have belief-like content similar to descriptive and non-moral beliefs, like the belief that grass is green. The sentence “It is wrong to commit murder” is simply the expression of a sentimentally or emotionally-centred attitude of disapproval towards committing murder, and the sentence itself does not express anything belief-like akin to ordinary descriptive sentences. This means that moral language does not function like ordinary descriptive language, and that having a moral belief is fundamentally different from having an ordinary descriptive belief.

Most noncognitivists are also anti-realists. That is, they hold that there are no objective moral facts, independent of our moral sentiments and emotional apparatus. Thus, moral sentences like “It is wrong to commit murder” cannot be true or false, since there are no moral facts about murder. However, since noncognitivism is primarily a view about moral psychology and language, it does not entail metaethical anti-realism, and as we shall see, expressivists are a type of noncognitivist who do not altogether dispense of moral facts.

1.3. Early noncognitivism – Emotivism

The first noncognitivists laid the modern foundation for the existing division in metaethics, between those who claim that sentiment is at the centre of moral practice and discourse, and the cognitivists who appeal to reason. The early noncognitivists were heavily inspired by logical positivism. They held that only sentences that could be verified logically or

---

6 To avoid confusion with the concepts of cognitivist theories I will hereafter not refer to moral judgements as “moral beliefs” when talking about ordinary noncognitivist theories.

7 It should be noted that there is no connection between moral anti-realism and scientific anti-realism. It is very common to be realist about unobservable physical particles but anti-realist about unobservable moral facts.
empirically were meaningful, and metaphysical or moral sentences, which were unverifiable, were meaningless – or nonsensical. A. J. Ayer developed one of the first extensive noncognitivist accounts, called emotivism, and he held that moral sentences “are simply expressions of emotion which can be neither true nor false.” (Ayer 1936: 104). Sentences are the investigational objects according to the verificational account, which can lead to knowledge when verified, but no knowledge can be obtained concerning the nature of morality, as moral sentences have no “literal” or “factual” meaning. “They are unverifiable for the same reason as a cry of pain or a word of command is unverifiable – because they do not express genuine propositions.” (Ayer 1936: 112) A sentence like ‘It is wrong to commit murder’ has two roles according to Ayer, 1) to express feelings and 2) to issue commands:

Thus the sentence ‘It is your duty to tell the truth’ may be regarded both as the expression of a certain sort of ethical feeling about truthfulness and as the expression of the command ‘Tell the truth’. The sentence ‘You ought to tell the truth’ also involves the command ‘Tell the truth’, but here the tone of the command is less emphatic. In the sentence ‘It is good to tell the truth’ the command has become little more than a suggestion. And thus the ‘meaning’ of the word ‘good’, in its ethical usage, is differentiated from that of the word ‘duty’ or the word ‘ought’. In fact we may define the meaning of the various ethical words in terms both of the different feelings they are ordinarily taken to express, and also the different responses which they are calculated to provoke. (Ayer 1936: 111)

Although moral terms and sentences have no literal or factual meaning, Ayer has developed a theory of meaning here with regards to the functional role of moral language. When we make a moral judgement about something, for instance that it is wrong to commit murder, this judgement is principally something emotional, which we can express by the sentence “It is wrong to commit murder”. When so expressed the meaning of this sentence has two components:

1) The feeling it expresses.
2) What is commanded by it.

Consider the following case. Smith witnesses Jones being murdered in the street. Smith is terrified, confused and overwhelmed by emotions of fear, shock and the strong conviction that a grave injustice has befallen Jones, and he yells “It is wrong to commit murder”. According to Ayer’s theory, this sentence does two things:
1) It expresses Smith’s emotions concerning the committal of murder, which are clearly negative.
2) It issues a command about murder, namely, not to do it.

To be fair to Smith, his spontaneous mental reaction is probably based on a specific and context-dependent array of emotions evoked by the particular circumstance of the murder of Jones, but the generic form he gave the sentence that expresses his judgement, probably due to shock and distress, makes it possible for us to capture his feelings towards murder more generally, that it is wrong to murder *simpliciter*. But Smith’s utterance functions not only to communicate that he disapproves of murder, but that everyone who hears him should do so as well.

This special double character of moral terms, that they not only express our own feelings, but also arouse feelings in others, provides us with the understanding of moral sentences as given in the imperative mode. Some sort of imperative or command mode was a common feature of all the early noncognitivist theories, and the theories of meaning of moral terms, developed by Ayer, Stevenson and Hare all apply functional definitions tied to the *usage* of the term. Charles Stevenson (1944) and R. M. Hare (1952) worked in the continuation of Ayer’s theory, and made important contributions to the development of early noncognitivism, where we can interpret each of them as having expanded on one of the two components of Ayer’s account – Stevenson on the emotional element, and Hare on the command element. All three define the meaning of moral words in terms of their function, and we should therefore understand their theories as *speech act* theories, according to Mark Schroeder. (2010a: 74) This is an important difference between early and later noncognitivists, where the latter find the performative features of moral sentences less interesting than their expressive features, and makes it easier to enter into debate with cognitivists, who find the performative features of moral sentences less interesting than their descriptive features.

### 1.3.1. Speaker Subjectivism

As Ayer admits himself, emotivism is a form of *speaker subjectivism* (1936: 113), which in its noncognitivist version holds that my moral judgements are only dependent on my own emotions, and can therefore not be true in any objective sense. “It is wrong to commit murder” only tells us that I am in some desire-like state of disapproval of committing
murder. He discusses an old objection against all forms of speaker subjectivism, raised by
Moore, that “if all ethical statements were simply statements about the speaker’s feelings, it
would be impossible to argue about questions of value.” (ibid.) Ayer’s response is that we do
not really argue about questions of value, for in every case of apparent moral disagreement,
“the dispute is not really about a question of value, but about a question of fact.” (1936: 114)
Ayer finds it unconvincing that we would argue that someone has the wrong feeling towards a
situation. We should instead focus on identifying incorrect and untrue beliefs about the facts
of the circumstance. Ayer’s hope is that if we convince our opponent to agree with us about
the natural facts, he would adopt the same moral attitude towards the matter. (ibid.) He lists
some typical factual error sources which could distort our beliefs:

1) Misconceiving the agent’s motive
2) Misjudging the effects of the action
3) Misjudging the action’s probable effects in view of the agent’s knowledge
4) Failure to take the agent’s special circumstances into account
(1936: 115)

In some cases, disagreement will evaporate if one of the participants changes his feelings
towards something, as a result of detection of error and reform in their factual beliefs.

Consider the following case. Smith states “In a recession, it is right that the central bank
lowers the policy interest rates to increase income”, while Jones states “In a recession, it is
wrong that the central bank lowers policy interest rates to increase income”. Let us assume
that it becomes clear after some discussion that Jones has misunderstood the effects of
increased income, mistakenly believing it to cause higher saving rates and decreased
spending, and thus deflation (a fall in aggregate price levels). Since Jones has often heard talk
of the danger of deflation during a recession, this originally made him feel that lowering the
interest rates would be wrong. But as soon as Smith explains to him the fundamental
dynamics of macroeconomics, Jones comes to understand that increased income would lead
to more spending, and thus increased inflation (a rise in aggregate price levels). As Jones
comes to realise his factual errors, and adopts similar beliefs to those of Smith, he also comes

---

8 There is a parallel cognitivist version of speaker subjectivism, a form of simple relativism, where my moral
judgements are only dependent on my own beliefs, dating back to the sophist Protagoras (ca. 490–420 B.C.). “It
is wrong to commit murder” is true if I believe that it is wrong to commit murder.
to feel that lowering the interest rates is right. We can identify Jones’ erroneous beliefs as a type 2 factual error, where he failed to realise that the central bank’s action would not lead to a dangerous level of deflation. And as his feelings towards lowering the interest rates change, they too conform to those of Smith, with whom he now shares all factual beliefs about the situation, and is in perfect moral agreement.

In other cases, it is harder to see how all moral disagreement would be solved by appealing to the factual features of a situation. If Smith claims “Abortion is right”, and Jones claims “Abortion is wrong”, it is reasonable to think that this disagreement could also persist after a careful examination of all factual matters concerning the issue. Let us imagine that Smith and Jones agree that human life is inviolable, but Smith does not believe that a foetus is a human being, while Jones does. On a noncognitivist reading, there must be different attitudes towards what constitutes human life at work that is the cause of their disagreement. If Smith thinks it is right to hold a scientific standard of developed consciousness in post-embryonic foetuses as a minimum requirement for the attribution “human life”, and Jones thinks it is right to hold that every embryo in virtue of its disposition to develop into a human being, must be attributed with human life, it is difficult to see how their disagreement could be resolved by appealing to facts. Many cognitivists have tried to resolve the problem of disagreement in similar ways as Ayer, where it seems moral disagreement would disappear if we reach agreement on all natural facts. Later noncognitivists embrace the approach Ayer was reluctant to take, and consider moral disagreement to be a *clash of attitudes*.\(^9\) This approach may explain genuine moral disagreement, but faces serious problems of identifying the precise content of the disagreement, if desire-like states of mind cannot be analysed in descriptive terms. We will return to this issue in section chapter 3.

### 1.4. Moral Motivation

The most novel aspects of noncognitivism, like Ayer’s emotivist account, is how it revolutionises moral psychology and language, by placing emotions, or desires, and not beliefs, at the centre. This in turn has also enabled the development of new theories of the

---

\(^9\) Stevenson might be considered the earliest proponent of a sort of hybrid view, and held that moral disagreement consists in both a “disagreement in belief” and a “disagreement in attitude”. (Stevenson 1944: 11)
practical qualities of moral judgements, of why we seem to be motivated to act in accordance with what we judge to be right. The difference between a desire-based and a belief-based account of moral judgements could be described as a difference in *directions of fit*. Desires have a mind-to-world direction of fit, seeking to get the world to conform to them, and beliefs have a world-to-mind direction of fit, seeking to conform to the world. (Anscombe 1957: 56; Searle 1985: 3; Smith 1994: 112). If moral judgement is a desire-like mental state it would then be directed towards fulfilling what it is about.

1.4.1. The Humean Theory of Motivation

This practical, or action-guiding, character of desires is the essence of the *Humean theory of motivation*. According to Hume, “reason alone can never produce any action” and “is, and ought only to be the slave of the passions” (Hume 1739: SB 415). A belief is not enough to motivate, and a desire must be present, where belief and desires are distinct existences. This theory is strongly tied to the thesis of *motivational internalism*, a thesis all noncognitivists adhere to in one way or the other, which says that the emotional base of our judgements generates motivation for action, and so our judgements are intrinsically motivating and action-guiding. By claiming that desires are the central components of moral judgements, where the very nature of desire is to change the world to its designs, the noncognitivist can easily explain how we are prompted to action in accordance with our judgements: Being in a desire-like state of mind entails that this state is simply a state of being motivated to change the world to fit that desire. We may define internalism in the following way.

**Motivational internalism:**

There is a necessary connection between moral judgements and motivation.

Internalists disagree about the strength of this connection, but agree that we are always motivated to some extent by what we morally judge is the right course of action, absent certain mental inhibitions like irrationality, apathy, depression, or other emotional dysfunction. If I judge that it is wrong to commit murder, and I am fully rational, normally strong-willed and psychologically healthy, I am necessarily motivated not to commit murder, to some extent. This form of internalism is also called judgement internalism. Another important form internalism relevant for our project is the thesis of *Reasons Internalism*, which holds that there is a necessary connection between an agent’s reasons for action and motivation. (Björnsson et al. 2015: 3)
**Reasons internalism**

Reasons internalism is the view that there is a necessary connection between an agent’s reasons for action and motivation.

In following Bernard Williams we can divide reasons into *internal* and *external* reasons:

Sentences of the forms “A has a reason to \( \phi \)” or “There is a reason for A to \( \phi \)” (where “\( \phi \)” stands in for some verb of action) seem on the face of it to have two different sorts of interpretation. On the first, the truth of the sentence implies, very roughly, that A has some motive which will be served or furthered by his \( \phi \)-ing, and if this turns out not to be so the sentence is false: there is a condition relating to the agent’s aims, and if this is not satisfied it is not true to say, on this interpretation, that he has a reason to \( \phi \). On the second interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive. I shall call the first the “internal”, the second the “external,” interpretation.

(Williams 1979: 101)

Internal reasons are reasons that concern “the agent’s *subjective motivational set*” (Ibid.: 102), which we can call *motivating reasons*. On a naturalist account, these reasons are reducible to psychological facts about an agent’s preferences and desires. For instance, I have a motivational reason to give to famine relief if I desire to give to famine relief. External reasons are objective reasons, and on a naturalist account they are reducible to mind-independent normative facts about what we ought to do. Reasons capture what we ought to do in both moral and non-moral situations. Following Mark Schroeder, we could define normativity\(^{10}\) in terms of reasons: “What it is to be normative, is to be analysed in terms of reasons.” (Schroeder 2007: 81) For instance, if it is a normative fact that I should get up at 7 o’clock in order to catch the train, I have a normative reason to get up at 7’clock. Subjective reasons are often called “explanatory reasons”, because they explain why I want to do something, by pointing to my desires or preferences. According to a simple noncognitivism view, I would then always have a motivating reason to do what I judge to be morally right, because I would be in a desire-like state of mind. Objective or normative reasons are called justificatory reasons, as they justify why I should do something, by pointing to normative facts about what I ought to do. Williams does not believe in external moral reasons, or in any objective “morality system” for that matter (Williams 1985: 174), and leaving non-moral

---

\(^{10}\) We consider morality to be a subfield of normativity.
normative reasons aside, early noncognitivists would not believe in objective moral reasons, either. Later noncognitivists, like moral realists, would not deny the existence of objective moral reasons, as they do not deny the existence of moral facts. We will return to moral reasons in the next section in regards to Smith’s solution to the moral problem, but will here concentrate on the relation between moral judgements and motivation.

1.4.2. The Amoralist

Important objections to motivational internalism are rooted in situations where moral judgements seem not to be followed by the appropriate motivation, for instance in cases where persons are simply reporting other people’s views about a moral matter, “in which a person is simply paying lip-service to a convention” (Hare 1952: 125), or are making an ironic point and putting moral terms in the “‘inverted comma’ use” (ibid.: 124). Hume wrote about the “sensible knave”, who “in particular incidents, may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any breach in the social union or confederacy.” (Hume 1777: 232) The sensible knave is a man who would do what he knows to be morally wrong if he could get away with it. Such a person could be called amoral, and the case of the amoralist is one of the most serious challenges to internalism. The amoralist is someone who judges something to be morally right, but is not motivated to act accordingly. David Brink has developed an account of the amoralist as a rebuttal of motivational internalism, and writes the following about the amoralist:

We can imagine someone who regards what we take to be moral demands as moral demands – and not simply as conventional demands – and yet remains unmoved. (Brink 1987: 49)

It is possible to imagine that there could be a person who is fully rational and strong-willed, who does not regard moral judgements as conventions or inverted comma cases, but still is not motivated by his moral judgements. As the amoralist case is proposed as a counter-example to motivational internalism, the internalist would naturally first try to find faults with his moral judgements, which could explain why the assumed necessary connection between judgement and motivation does not occur. One of the more extensive replies to Brink is developed by Michael Smith (1994). Smith responds that what the amoralist is doing, is something along the lines of making “inverted comma” type judgements in Hare’s sense, “where they [amoralists] try to make moral judgements, but fails.” (Smith 1994: 68) Smith sees the amoralist as failing to use moral terms properly, so he fails to make proper moral
judgements because of a lack of mastery of moral terms. An analogy to the mastery of colour terms explains the reason for failure:

The problem can be brought out by reflecting on the case of someone, blind from birth, who has a reliable method of using colour terms. We might imagine that she has been hooked up to a machine from birth that allows her to feel, through her skin, when an object has the appropriate surface reflectance properties. (Smith 1994: 69)

But the ability to have visual experience is a necessary condition for the mastery of colour terms, so this blind person does not really make colour judgements, in the same way the amoralist fails to make moral judgement, according to Smith.

Smith believes the amoralist lacks the mastery of moral terms in the same way. I believe Smith’s argument fails to refute the amoralist case, because Smith fails to recognise the possibility of an amoralist making genuine moral judgements. To explain why, let us consider two versions of the amoralist case – a weak version and a strong version.

The weak amoralist case:
The amoralist seemingly makes moral judgements, not just conventional reports about the moral judgements of others, but is not motivated by them.

The strong amoralist case:
The amoralist makes moral judgements, not just conventional reports about the moral judgements of others, but is not motivated by them.

The weak amoralist case represents the reading of Brink, and I agree with Smith that his argument refutes this version, because the weak amoralist fails to make real moral judgements, even though it seems like he makes real ones. The analogy to colour terms also refutes a similar version of the colour case, where the blind girl only seemingly makes colour judgements, but not really. But according to Brink the amoralist makes genuine moral judgements, not just seemingly:

Why should we assume that the person who asks ‘Why should I be moral?’ is using moral language in inverted commas or is mistaken about what morality requires? Why can’t someone have correctly identified his moral obligations and still wonder whether these obligations give him good reason for
action? The thought that someone might possibly not have good reason to act on his moral obligations
give his moral obligations need not force us to withdraw our ascription of obligation. (Brink 1989: 59)

This is the strong amoralist case, where the amoralist really does recognise moral facts, not
just seemingly, and makes real moral judgements based on those facts, guaranteed by the
realist claim of how moral beliefs are formed by the moral facts they aim to capture. So in
order to refute this case we should grant Brink the possibility of an amoralist making real
moral judgements, and look for other reasons to explain the lack of motivation.

If we consider Smith’s argument as an argument against the strong amoralist case, he begs the
question when setting up his argument to refute it. He assumes that it is an open question
whether he makes real or apparent moral judgements, where the real challenge concerns
refuting a case where this is a closed matter, and that the judgements are real. We could
understand Smith’s argument as bearing the form of a conditional proof with a straight-
forward modus ponens form:

(Premise) 1. If the amoralist lacks mastery of moral terms, he only seemingly makes moral
judgements.
(Premise) 2. The amoralist lacks mastery of moral terms.
(Conclusion) 3. The amoralist only seemingly makes moral judgements.

But the conclusion is false, since Smith fails to appreciate the conditions of the strong
amoralist claim. If we grant that Smith has established convincingly that the amoralist lacks
mastery of moral terms, so it must be the conditional in premise 1 that is false, since we know
that no false conclusion can follow from true premises in a valid argument, which modus
ponens is. Given the conditions of the strong amoralist case, it cannot follow from the lack of
mastery of moral terms that the amoralist only seemingly makes moral judgements.
Symbolized and with truth values Smith’s arguments reads:

(Premise) 1. P ⊃ Q – False
(Premise) 2. P – True
(Conclusion) 3. Q – False
The possibility of a strong amoralist may be improbable, and it may be enough to show that the conditions of the challenge are false, and no such amoralist exists. Many noncognitivists would hold that the strong amoralist is an impossibility, since the lack of motivation signifies the lack of a desire-like state of mind, which implies the lack of any moral judgements at all. But if we are to take the challenge seriously, we must look outside the amoralist’s moral judgements, as they are flawless, by the conditions of the case. I believe Smith is on the right track by focusing on the amoralist’s lack of mastery of moral terms, or lack of moral competence. Given the absence of irrationality and weak will, it is reasonable to think there might be some other psychological defect present, which distorts the way his cognitive or emotional faculties relate to moral judgements, thus failing to generate motivation. If the internalist is not able to resolve amoralist issues satisfactorily, it might jeopardise his project.

The denial of moral internalism is moral externalism, a claim that is defended by many cognitivists, like Brink. And unlike the internalist, it escapes the problems of the amoralist, but unlike the internalist, it faces the graver issue of explaining how our moral judgements motivate us. One of the attractive traits of all noncognitivist theories is how well they can accommodate motivational issues, issues that have been a long-standing problem for moral realists. We can formulate this problem in the form of a challenge that must be answered by any plausible cognitivist theory.

**The motivational challenge:**
A cognitivist theory must explain how we can be motivated to action by our moral judgements, when these are belief-like and not desire-like states of mind.

One philosopher who has advanced the realist project considerably in this respect is Michael Smith. According to Hume, motivation requires the presence of a desire and a means-end belief of how to fulfil that desire. Smith acknowledges the importance of the Humean theory of motivation, and develops his own form of internalism, which he calls the “practicality requirement of morality”. (1994: 6)

But as a realist Smith must also account for the objective features of morality, and the real challenge for the realist lies in making a unified account of both the objectivity and
practicality requirements of morality. Smith calls this *the moral problem*\(^{11}\), and we will review different cognitivist solutions to this problem in the next chapter.

\(^{11}\) Sayre-McCord (1997) points out that “the moral problem” should be called “the metaethical problem”, as it only concern metaethical issues.
Chapter 2

Cognitivism and The Moral Problem

2.1. Moral Realism

I will now consider some general aspects of cognitivist theories, especially what is required of them to meet the objectivity requirement. Then we will discuss two versions of non-reductive realism, Cornell realism and Michael Smith’s theory, and see how successful their solutions can be in solving The Moral Problem. In the previous section I outlined the motivational challenge, and how failure to answer this challenge would be a serious problem for any metaethical views.

As noted in section 1.2, most cognitivist views incorporate some form of moral realism, which holds that there is an objective moral reality independent of what we think about it. The moral problem for the realist would then first be to explain how there could be external moral facts, detached from human reasoning and sentiments, that still have such a relation to our beliefs about them, that we often are able to form true beliefs about them. Secondly, the realist must explain how descriptive beliefs about a detached realm of moral facts are able to motivate us to act in accordance with these facts. The different features of moral truth can be described by a variety of moral properties, relations and facts. We can formalise metaphysical moral realism as a positive thesis with one central claim.

**Moral realism:**

There are mind-independent moral facts.\(^{12}\)

---

\(^{12}\) If we had subtracted the attribution of mind-independence from this claim, we would also include all forms of cognitivist speaker subjectivism and many forms of moral constructivism (e.g. Neo-Kantianism versions), which also could be considered forms of moral realism. To complicate the matter there are less constructivist forms of moral realism that rely on moral facts which could be considered mind-dependent, as mentioned in section 1.2, for instance in the moral functionalism of Jackson (1998: 156–157), Railton’s ideal advisor theory (1986; 2003) or in Smith’s theory of normative reasons (1994: 187), where moral facts are based on the convergence of people’s shared beliefs and desires.
But moral realists also need an account of how we can obtain knowledge of moral facts, of how our moral judgements and language are able to capture and illuminate them. How else can we know that they are there? Moral realism has been defined numerous times and very differently over the recent decades: The most common starting point can roughly be described as a semantic thesis with the conjoined claims of 1) cognitivism and 2) a success theory, that some moral claims are true because they correspond to moral facts. (Sayre-McCord 1986: 8; Finlay 2007: 82; Miller 2009: 125) Without a plausible epistemological account linking the psychological and metaphysical accounts, Smith’s objectivity requirement would not be met, and the whole realist project would be undermined. Most moral realists develop this story along the following lines: moral language has a descriptive, fact-stating grammatical structure, which enables moral judgements to aim to capture, or track, moral facts. Moral facts can give us reasons to act in accordance with them, and they can motivate us to do so if our judgements are able to track them. We express these judgements, e. g. that it is wrong to commit murder, through moral sentences like “It is wrong to commit murder”. Here the moral term “wrong” picks out the property of wrongness from the external realm of moral properties, and attributes it to the committing of murder, and the sentence is true or false depending on whether or not it is a fact that it is wrong to commit murder. Moral realists usually combine a correspondence theory of truth with a truth-conditional semantic theory (Davidson 1967). Moral sentences are truth apt because they express (moral) propositions capable of truth and falsity. Propositions are a useful abstract tool of analysis, letting us determine whether the representational contents of sentences and beliefs are similar to the facts we wish to examine. The truth-value of these propositions depends on the moral facts they stand in a truth-making relation to – that is, correspond to. Moral facts, therefore, are the truth conditions of moral propositions, and moral propositions, in turn, derive their meaning from their truth conditions – moral facts. Many metaethicists, and especially expressivists, prefer a minimalist to a correspondence theory of moral truth, as we will see in chapter 2.

13 See Keith Simmons and Simon Blackburn (1999).
14 The role of propositions in moral sentences will be examined closer in the next chapter.
2.2. Supervenience

Naturalism about morality, as we have described it, implies that any moral properties must be “fully made up”\(^\text{15}\) by natural properties. Some believe moral properties are identical to, or reducible to, natural properties, and others that moral properties, though natural, are irreducible to some other natural non-moral properties. Given naturalism, there is an important conceptual constraint on this property-relationship, construed to show in which way moral properties are dependent upon natural properties. We will call this constraint the *moral supervenience thesis*.\(^\text{16}\)

**Moral Supervenience:**

Moral properties\(^\text{17}\) supervene on natural properties if, and only if, for any two possible worlds, \(w_1\) and \(w_2\), if all natural properties in \(w_1\) and \(w_2\) are identical, all moral properties in \(w_1\) and \(w_2\) must also be identical.\(^\text{18}\)

Any plausible form of moral naturalism, whether cognitivist or noncognitivist, would have to ascribe to some variant of this supervenience thesis. Michael Smith wrote that everyone believes in moral supervenience (Smith 1994: 21). This is a truth with some modifications, for even though most metaethicists, of any inclination, would accept some form of supervenience thesis, there are those who do not embrace it wholeheartedly, or not at all. Many are found among nonnaturalist realists, like Moore (1903: 41), McDowell (1998: 202) and Wiggins (2006: 380). For anti-realists the supervenience relation is easy to explain, as it is an empty one, because there are no moral facts. Some realists have suggested a broader understanding of the supervenience thesis, which in Sturgeon’s simple formulation, reads “that we cannot

---

\(^{15}\) The description “fully made up” might seem wide-ranging, but is meant to neutrally capture the many ways the supervenience relation can be instantiated (e. g. by different forms identity-, reduction- or constitution-claims).

\(^{16}\) The notion of supervenience was first introduced by Hare (1952: 131, 145).

\(^{17}\) We assume that supervenience governs not only the relationship between moral and natural properties, but between any moral aspect of reality, such as features, facts, events, circumstances, etc., and their natural counterparts.

\(^{18}\) It might be debatable whether certain abstract entities could be considered natural, e. g. propositions, relations or functions. So the thesis is most tenably formulated as a relation between moral and non-moral properties or facts, although practically every non-moral property and fact is to be considered a natural property. For our purposes, I will assume that every mental or abstract construct or entity is a part of the natural world.
suppose two objects to differ evaluatively without taking them to differ in some natural respect as well.” (Sturgeon 2009: 58). As we move on, we should expect any naturalist realist theory to have a reasonable account of the relationship between moral and non-moral properties.\(^{19}\)

We could distinguish between two broad naturalist realist traditions, one reductive and one non-reductive, with respect to the relationship between the moral and natural aspects of reality. How these two realist traditions describe the supervenience thesis this relationship relies on, and how they account for the discovery of moral truths, form the basis of how they propose to meet the objectivity requirement. We remember that The Moral Problem is the problem of combining the objective features of morality with the practical, or action-guiding, character of morality.

**The objectivity requirement**

An account of the objective features of morality.

**The practicality requirement**

An account of the motivating force of moral judgements.

The objectivity requirement could be reinterpreted as a requirement to answer two objections raised by John Mackie against moral realism, called the argument from relativity and the argument from queerness. In the first case the realist needs to explain the diversity of contrary moral beliefs held by different individuals and cultures, and at different times. In the second case the realist needs to explain the nature of moral properties and facts, how they can be both objective and part of a plausible naturalistic world view. (Mackie 1977: 36–38)\(^{20}\)

\(^{19}\) Earlier nonnaturalist intuitionists like McDowell (1998) and Wiggins (1998) do not support the supervenience thesis, but later nonnaturalists like Shafer-Landau (2003) and Enoch (2011) support weaker nonreductive forms of supervenience, where the supervenience relation consists of some form of constitution, and their accounts bear a strong resemblance to naturalism. Shafer-Landau has suggested that the difference could be best understood as epistemological, as a nonnaturalist could claim that moral facts exist in a naturalist sense, without being identical to or reducible to non-moral natural facts (2003: 108, 115).

\(^{20}\) As mentioned in section 1.2 cognitivism is also compatible with a scepticist tradition that is anti-realist, which holds that moral judgements aim to represent moral facts, but fail because there is no moral truth. This theory is often called error-theory, and a failure to answer the objectivity requirement adequately has led many subscribe to this view. The two most prominent representatives are Mackie (1977) and Joyce (2006). Error-theorists often
2.3. Cornell Realism

One very influential way to answer the objectivity requirement has been developed by the so-called Cornell realists\(^\text{21}\), where, in addition to David Brink, Richard Boyd and Nicholas Sturgeon are the most important contributors. Boyd has developed a semantic theory of moral terms based on the causal theory of reference developed by Putnam (1975) and Kripke (1980). According to Boyd, a moral term like “good” has a synthetic definition, that is, a definition that refers to \textit{a posteriori} (empirically detectable) properties that reveals the essence of the moral terms. There is a family of natural properties that causally regulates the use of the term “good”. (Boyd 1997 [1988]: 117) The idea is that moral terms have certain natural, functional properties that are similar to natural kind terms. The term “good” is a rigid designator, which picks out the same natural property across all possible worlds.\(^\text{22}\)

Epistemologically speaking, we can say that through our linguistic competence of natural languages our use of the term “good” \textit{tracks} the natural properties they refer to and are regulated by, in the same way as the term “water” tracks and refers to H\(^2\)O. The identification of moral terms and natural properties are thus necessary truths, but they are not analytical or conceptual truths and discovered \textit{a posteriori}.

2.3.1. The Open Question Argument

Cornell realism established itself as an answer to an argument by G. E. Moore, called \textit{The Open Question Argument}, which intended to show that any attempt to define moral terms in natural terms would lead to a fallacy – \textit{the naturalistic fallacy} – because moral properties are not reducible to natural properties. Additionally, Moore argued that a definition of a moral term, like “good”, would have to be analytical, that is, involve an identity relation. Therefore, any definition of “Good” in natural terms would have to be a conceptual truth, where the meaning of the \textit{definiens} and \textit{definiendum} are synonyms, and a closed matter not open for further investigation. This argument is related to what is commonly called “the paradox of appeal to evolutionary arguments in undermining the objectivity of morality. This is a line of argumentation that also has a strong appeal to noncognitivists, which we will see in the next chapter.

\(^{21}\) The name illustrates their affiliation with the Cornell University of Chicago.

\(^{22}\) We can say that a “Natural kind” is “a natural grouping of objects which mirrors the natural structure of the world” (my definition). Examples of natural kind terms could be “water”, “gold”, “horse” etc. Natural kind terms are “rigid designators”. Something is “a rigid designator if in every possible world it designates the same object”. (Kripke 1980: 48)
“Analysis”, which is the problem for conceptual analysis of rendering results that should be analytic, but at the same time correct and informative. (Smith 1994: 37; Jackson 1998: 140) According to Moore, it seems that for any natural property N, we can always ask, “but is it good?”, so the question of how to define “good” by reference to natural properties is always open:

But whoever will attentively consider with himself what is actually before his mind when he asks the question ‘Is pleasure (or whatever it may be) after all good?’ can easily satisfy himself that he is not merely wondering whether pleasure is pleasure. (Moore 1903: 16)

The question is open because we can never be certain whether any proposed natural property N, for instance, pleasure, is also good. “‘Good’, then, is indefinable.” (ibid.)

Many metaethicists hold that Moore’s argument is flawed, and does not involve a fallacy. William Frankena objected that Moore begs the question against naturalism, by assuming that “good” refers to an indefinable non-natural property, which can never be defined in naturalistic terms. He should have instead called it the definist fallacy, the fallacy of confusing one property falsely with another. (Frankena 1939 [2006: 52]) In Moore’s case we could call this the informal fallacy of confusing a sui generis indefinable moral property with a natural property. Even though there is disagreement about the soundness of Moore’s argument, many believe it raises a serious question – the difficulty of reducing moral properties and facts to natural properties and facts.

For a long time, the open question argument seemed to be fatal for naturalist moral realism, relying on conceptual analysis\(^\text{23}\), and the only open course for the realist seemed to be non-naturalism or some form of nihilism. While Cornell realists agree with Moore that moral properties are non-reducible and sui generis, they hold that they are natural properties. Sturgeon argues that we find evidence for natural moral facts simply by examining our observations about the world, and moral facts figure in our best explanations of moral situations and beliefs. For instance, consider Hitler’s moral character or the anti-slavery movement of the 18\(^{th}\) and 19\(^{th}\) century. The reason why Hitler killed a lot of innocent people

\(^{23}\) Following Jackson, we can define “conceptual analysis” as “the very business of addressing when and whether a story told in one vocabulary is made true by one told in some allegedly more fundamental vocabulary.” (1998: 28) Conceptual analysis and the traditional analytic-synthetic devide is perhaps most famously criticised by
was because he was morally depraved, and the reason why the anti-slavery movement was so strong in the 18th and 19th century was because slavery is morally wrong. Appeal to moral facts and properties are part of the best explanations in these two instances, and in countless more from our everyday life and world history. (Sturgeon 1985 [2006]: 130)

2.3.2. The Revised Open Question Argument
The Cornell realists avoid Moore’s charges against reductive (or definitional) naturalism, but there is a modified version of Moore’s argument specifically directed at their Putnam-Kripke semantics that poses greater challenges. The argument is of a version of Putnam’s (1975) Twin Earth thought experiment developed by Mark Timmons and Terry Horgan. Putnam’s argument is meant to illustrate that natural kind terms are rigid designators. Putnam asks us to imagine a planet called Twin Earth, where everything is similar to our Earth, except that the term “water” does not refer to the liquid, transparent, tasteless substance with the molecular structure H₂O, but instead a seemingly identical liquid, transparent tasteless substance with a very complex molecular structure, abbreviated to XYZ. According to Putnam our intuitions tell us that people from Twin Earth and people from earth do not mean the same thing when using the term “water”. For a person from Earth, “water” rigidly designates H₂O, which he would still refer to if we imagine that he was transferred to Twin Earth and engaged in conversation there.

If moral terms are rigid designators, the term “good” should refer to the same natural property across all possible worlds. Timmons and Horgan ask us to imagine that a person from Earth and a person from Moral Twin Earth discuss the nature of moral terms such as “good” and “right”. It turns out that the person from Earth uses “good” and “right” to refer to certain natural consequentialist properties that are captured by a specific consequentialist theory, and the person from Moral Twin Earth uses “good” and “right” to refer to certain natural non-consequentialist properties, captured by a specific deontological theory. But the consequentialist properties from Earth and the non-consequentialist properties from Moral Twin Earth are so similar that the consequentialist theory of Earth and the deontological theory of Moral Twin Earth operate similarly in the moral discourse of the two planets. The reason why moral terms pick out different natural properties on Earth and Moral Twin Earth could be because of small species-wide psychological differences, for instance that persons
from Twin Earth are more susceptible to feelings of guilt and less susceptible to feelings of sympathy. (Timmons and Horgan 1992 [2006: 188–190]; Timmons 1999: 59–67)

Let us further imagine that a competent language user from Earth visits Moral Twin Earth, and engages in moral discourse with an equally competent person from Moral Twin Earth. The person from Earth discovers that the person from Moral Twin Earth use “good” and “right” to refer to the deontological properties, and not consequentialist properties. Here we should think that their intuitions would be the same as in Putnam’s Twin Earth example, and that moral terms from Earth and Moral Twin Earth have different meaning. Instead the intuitive response would be that the terms “good” and “right” has the same meaning on Earth and Moral Twin Earth, and the fact that they refer to different moral properties (and different normative theories) constitute non-trivial differences in moral beliefs – and genuine moral disagreement. According to Timmons and Horgan this shows us that moral terms are not rigid designators, and not causally regulated by the same moral properties across all possible worlds. In other words, it is an open question what natural properties terms like “good” and “right” refer to. (ibid.)

2.3.3. Hare’s Cannibal Objection

If instead of Earth and Twin Earth we imagine two different cultures on Earth, this thought experiment would generate the same result. A much-discussed related argument by Hare illustrates this difference. Imagine a western missionary visits a cannibal island and discovers that the word “good” has the same broad role in the language of both cultures as “the most general adjective of commendation”. (Hare 1952: 148) The difference is, however, that the missionary applies “good” to people who are “meek and gentle” and are “doing no murder”, while the cannibals apply “good” to people who are “bold and burly and collect more scalps than the average”. (ibid.) Obviously, “good” does not mean the same in the two languages, and the difference in meaning seem to concern something more than a mere difference in the descriptive meaning of “good”. The different meanings of “good” suggest the same here as in the Moral Twin Earth case, that there is a fundamental difference in moral belief. If we assume the truth of naturalist realism (of any kind), it seems unreasonable to think that the difference in moral belief between the cannibal and the missionary is caused by contradictory

---

24 A modern reader would find Hare’s example rather racist, but since it is a much-discussed example well suited to our purposes, I ask that we disregard this aspect of it.
moral facts – either “murdering people for scalps is good”, or it is not – *tertium non datur*. So for either the cannibal or the missionary, or both, their use of “good” is not able to track the natural properties instantiating goodness. At least one of them is mistaken. Both of them are making moral judgements, widely shared by, and deeply anchored in, the moral sentiments of members of each culture. This suggests that the difference in moral judgements is based on a difference in other attitudes than fact-tracking descriptive belief, perhaps a difference in feelings such as guilt, sympathy or honour, which strengthens the case for noncognitivism. We could construct an infinite number of similar cases, not only based on the variations of moral beliefs through history, but also on the conflicting moral beliefs held by different cultures and individuals at any given time.

Both the revised open question argument and Hare’s argument are examples of arguments from moral diversity (cf. Mackie 1977: 36). While the revised open-question argument is directed at non-reductive accounts, and poses a potentially fatal challenge to a Cornell realist solution to the objectivity requirement, Hare’s argument is an objection to any moral realist theory.

**2.4. Smith’s Solution to The Moral Problem**

We will proceed to show how Michael Smith’s solution to the moral problem offers both an account of the objectivity requirement not affected by any version of the open question argument, and also a unified solution to the two requirements and The Moral Problem.

At this point there could be good reason to distinguish objectivism and subjectivism regarding morality from mind-independence and mind-dependence in the metaethical debate. Different forms of moral constructivism, particularly Neo-Kantian accounts, could be considered both cognitivist, and perhaps realist, but not mind-independent, with regards to their metaphysical or ontological moral claims. Neo-Kantians normally think that moral objectivity is established by accounts of rationality, autonomy and moral reasoning (cf. Korsgaard 1996: 160; Putnam 2004: 18, 51).

Michael Smith has responded to some of the objections to the Humean theory of motivation from the realist camp as a mistaken mix-up of explanatory and justificatory reasons for action.
While motivating reasons explain actions, only normative reasons can justify them, and account for what it is right to do. (Smith 1994: 97) A solution to The Moral Problem must contain a satisfactory account of both types of reasons. Smith’s solution combines a Humean theory of Motivation with an anti-Humean theory of reasons, that is, motivational internalism combined with a theory of normative reasons.

Smith holds that motivational internalism is a conceptual truth because of the nature of desires, specifically their mind-to-world direction of fit. So if you make the moral judgment that it is right to φ in circumstance C, given certain rationality constraints, you are motivated to φ in circumstance C. This is the answer to the practicality requirement, which explains the action-guiding nature of moral judgements.

Making a moral judgements involves both a desire and an end-means belief about how to fulfil that desire. Having a desire is, inter alia, having the goal of bringing the world in tune with that desire. (ibid.: 113) Our normative reasons are formed by rational deliberation about our long-term interests. Smith applies Williams’ (1979) definition of practical rationality:

i) the agent must have no false beliefs
ii) the agent must have all relevant true beliefs
iii) the agent must deliberate correctly
(Smith 1994: 156)

Desires are dispositional states, which encompass certain ends, or long-term interests. Our desires and goals are thus formed through rational deliberation about what to do, which gives us our normative reasons. In other words, our rational long-term desires together with a means-end belief about how to satisfy those desires constitutes our normative reasons (ibid.: 131):

Rational deliberation $\rightarrow$ desire + means-belief $\rightarrow$ normative reasons

Our normative reasons are formed through a deliberation process dependent upon counterfactual analysis, commonly called an “ideal advisor” model. “Because what we have normative reason to do is a matter of what we would desire that we do if we were fully rational.” (ibid.: 180) Given rational deliberation, we should desire that which we would
desire if we were fully rational. Our actual desires, of course, do not always correspond to our ideal desires – what we would desire if we were fully rational and deliberated correctly. A failure to desire what we have normative reason to do could be the result of a number of things, for instance weak will, imperfect information about all relevant facts, or other psychological defects.

A vital part of our moral deliberation process is the analysis of moral concepts. If I ask myself what the right thing to do is, I must be able to define the concept of “right” by reference to certain platitudes, according to Smith. We can understand platitudes as rational requirements of a certain type that work as constraints on our deliberation process. Some of the most important platitudes are “platitudes concerning practicality, supervenience, objectivity, substance and procedure.” Smith calls the method of analysis we should apply a “non-reductive summary style dispositional analysis” (ibid.: 185) This is a method of specifying which moral concepts apply to a given situation, and how they relate to what we should desire. The analysis contains both a conceptual claim concerning our concept of “right” and a substantial claim concerning how the concept of “right” is instantiated in the world. This theory of analysis would avoid a potential problem for what Smith calls network style analysis, that is reductive realist theories based on conceptual analysis. These theories are susceptible to a variation of the paradox of analysis, called the permutation problem, a difficulty to differentiate between identical natural properties, where these are supposed to be unique. (ibid.: 50, 54)

Reasons are not always agent-neutral, and we should also consider whether something is desirable generally (de dicto), or just for me in a given circumstance (de se). (ibid.: 169) If I wonder whether I should give to famine relief, my deliberation process would involve asking both whether it is right to give to famine relief on a general basis, and whether it is right for me to give to famine relief in this circumstance.

Moral facts about what is right and good are facts about our shared normative reasons (ibid.: 39), and “they are themselves simply requirements of rationality or reason.” (ibid. 62)

25 The best known example of this type of reductive theory is the so called “Canberra plan” (named after its affiliation of the Australian National University, Canberra), where Frank Jackson and David Lewis are prominent figures. The idea is to use conceptual analysis to define a moral property “m” in virtue of a non-moral property “n”, and quantify the process. (cf. Jackson 1998:140)
According to Smith, the objective, or shared, content of normative reasons is caused by a "convergence in our desires under conditions of full rationality". (ibid. 187). This is Smith’s answer to the objectivity requirement.

The solution to The Moral Problem
If I make the moral judgement that it is right to φ, I am, ceteris paribus, motivated to φ. (practicality requirement) Judging that it is right to φ involves having the desire to φ. If rational deliberation reveals that I should desire to φ, this (ideal) desire together with a belief concerning how to φ constitute a normative reason to φ (bridging argument). Given rational deliberation, my moral judgement that it is right to φ captures the moral fact that it is right to φ, which is a fact about people’s shared normative reasons to φ, where shared normative reasons to φ are the result of a convergence in people’s desires to φ (objectivity requirement).

2.4.1. A Counterexample
I believe Smith’s solution may be troubling both with regard to the practicality and the objectivity requirement. In order to satisfy both requirements he has given desires a much more prominent position in both his theory of moral judgements and his theory of moral facts than previous realists have done. He makes moral judgements dependent upon desires through their role in shaping normative reasons, and by implication, moral facts. Smith’s theory of moral facts must then be understood as mind-dependent.26 It may be impossible to satisfy the objectivity requirement without postulating mind-independent facts about morality, and I believe Smith has failed to do so. I will now construct a counter-example to Smith’s theory of moral facts, in the form of a reduction ad absurdum argument based on Hare’s cannibal example.

We assume moral facts are objective. Let us assume rational deliberation causes all fully rational members of the society on Cannibal Island to believe that it is right to murder and collect scalps, and that they should have the desire to murder and collect scalps. Then there is a convergence of the desire to murder and collect scalps, and there are shared normative reasons among all rational cannibals to murder and collect scalps. Moral facts are constituted

26 For a similar observation regarding Smith’s moral metaphysics see FitzPatrick 2008: 166.
by shared desire-dependent reasons, so it is a moral fact that it is right to murder and collect scalps. Let us then assume that an identical rational deliberative process causes the missionary and all other fully rational members of western society to believe that it is wrong to murder, and that they should have the desire not to murder. Then there is a convergence of the desire not to murder, and there are shared normative reasons among all rational westerners to not murder. Moral facts are constituted by shared desire-dependent reasons, so it is a moral fact that it is right not to murder. Then it is a moral fact that it is right to murder and collect scalps, and it is a moral fact that it is not right to murder and collect scalps, which is a contradiction. We must conclude that Smith’s proposed solution to The Moral Problem implies that moral facts are not objective, and fails to offer an objective account of moral realism regarding moral facts.

We have now considered different moral realist theories and, and have seen there are considerable difficulties involved in finding a solution to the moral problem through a unified account that satisfies both the practicality requirement and the objectivity requirement. We seem to be unable to find plausible naturalist realist accounts of objectivity among traditional cognitivists. And the most promising cognitivist theories, like Smith’s, also seem to have unattractive features for a naturalist realist theory. Smith’s mind-dependent account of moral facts seems to display an inter-cultural relativity between isolated societies. If Smith were to revise or specify the criteria for convergence in desires, by adding platitudes to regulate the development of ideal desires in a way that could avoid too much cultural diversity, he could of course escape our reduction-argument. This would, however, make his account more constructivist and rationalist, and less realist. These difficulties can be seen as an argument for a noncognitivist approach to metaethics. We will revert to noncognitivism, particularly expressivism and its hybrid version, in the next chapters to see if a more plausible account of morality could be found there
Chapter 3

Expressivism and Hybrid Expressivism

3.1. Why expressivism?

Perhaps the most attractive feature of expressivism is how it combines a general theory of scientific naturalism with a simple Humean account of motivational internalism, that is, how to think that it is wrong to commit murder, namely to disapprove of murder, motivates us and gives us reason not to commit murder. To have a moral belief is to be in some desire-like state, and being in a desire-like state entails motivation.

I find that the most potent argument for expressivism is the combined force of three key metaethical considerations, which amount to one negative epistemological claim, one positive claim regarding moral psychology and one practical claim.

1) If there are objective moral facts, it is no reliable way to obtain knowledge of them.
2) Moral judgements are based on emotions, shaped by evolution and culture.
3) Motivational internalism is true.

When comparing different theories of any given phenomenon, wielding Occam’s razor would lead us to place considerable weight on the more simple, parsimonious and clear-cut version. I believe many realists would also agree with 1), even though they have at least developed partial accounts, which when expanded and refined, may give us the tools we need to finally locate the missing moral facts. The Cornell realists might refer their causal theory of the reference of moral terms (Brink 1989; Boyd 1988), and the Canberra Planners might mention their infinite disjunctive morality function (Jackson 1998). Note that 1) is a purely epistemological claim, as it remains neutral with regards to the ontological status of moral facts and properties. But a corollary of 1) is that moral facts are inaccessible, which could reasonably be regarded as weakening any metaphysically realist claims about the status of moral facts. If it is impossible to know whether it is wrong to commit murder, is seems highly speculative to assert it as an objective fact.
Expressivism in metaethics is both a theory about moral metaphysics and a theory about moral language. With regard to moral language we can consider “expressivism” to be a “meta-semantic theory” rather than a “semantic theory”, as our primary concern is the question of how moral sentences get their meaning, not the content of their meaning. (cf. Ridge: 2014: 8) The main challenge for all noncognitivists is to give an account of the meaning of moral sentences not based on a truth-conditional theory of meaning. Expressivists like Simon Blackburn and Allan Gibbard believe that the meaning of moral sentences is given by the contents of moral thought. And the nature of moral thought and reasoning concerns engaging our moral feelings and sentiments, where these are desire-like states of mind and not belief-like states of mind:

That words express judgments will, of course, be accepted by almost anyone. The controversial elements of the norm-expressivistic analysis are these: first, its account of what it is to make a normative judgment, and second, its claim that the meaning of normative terms is to be given by saying what judgments normative statements express — what states of mind they express. (Gibbard 1990: 84)

A moral sentence like “it is wrong to commit murder” functions to express some emotional attitude towards murder. As we saw in chapter 1, noncognitivists disagree on which sort of attitude or state of mind this might be. Blackburn believes it to be a state of disapproval towards murder (Blackburn 1984: 16, 1993: 18, 1998: 66), and Gibbard believes it to be a state of accepting a norm that murder is wrong (Gibbard 1990: 7, 2003: 7), or a state of planning not to murder (Gibbard 2003). Blackburn describes the process of forming attitudes of approval or disapproval as a process where a set of inputs yields certain outputs. For instance, if Jones observes someone torturing the cat, he could form certain assumptions about the torturer’s intentions, the cat’s pain, and through other factors of the circumstance (input), come to harbour a strong attitude of disapproval towards torturing the cat (output). (Blackburn 1998: 5) This account is often referred to as response-dependent, because our sensibilities respond to certain inputs. Gibbard’s account of the state of accepting a norm could be considered similar to Blackburn’s input-out account, where the process is governed by principles of rationality.
3.1.1. The parity thesis

Expressivists are committed to what Mark Schroeder calls “the parity thesis”, which is a claim concerning the expression relation for moral sentences. (Schroeder 2008b: 89; 2014: 280)

**The parity thesis**

Moral sentences have the same relation to noncognitive attitudes as ordinary descriptive sentences have to ordinary propositional beliefs.

This claim entails the “expression relation” to be the same for the expressivist and the cognitivist, even though the content of the expression is different. If Smith is an expressivist and Jones is a moral realist, and they both utter the sentence “murder is wrong”, this sentence relates to the content of the moral judgement it expresses in the same way for both Smith and Jones, but the content expressed is different. For the realist it expresses propositional contents with truth conditions, while for the expressivist it expresses a noncognitive attitude.

Expressivists are confronted with a serious problem concerning the analysis of moral sentences, called “the Frege-Geach problem”. The problem consists in giving an account of the meaning of complex moral sentences, when these contain logical operators and express desire-like states of mind. The expressivist will need to show how the meaning of a complex sentence is derived from the meaning of its parts, in which he needs to account for the meaning of the simple parts of the sentence. The question is how the attitudes expressed by the simple sentence parts are inter-related, and how we are to understand their meaning when they are part of a complex. The problem arises through the use of logical operators on sentences. For instance, let us assume that the sentence “it is wrong to commit murder” expresses an attitude of disapproval towards murder. What kind of attitude does the negation of this sentence express? A disapproval of not-murdering, an approval of murdering or non-committal towards murdering? We will return to this problem in the next chapter.

---

27 The problem is named after Peter Geach and Gottlob Frege, who Geach (1965) draws inspiration from in his original formulation of the problem.
3.1.2. Minimalism about Moral Truth

On the metaphysical level most noncognitivists hold that moral sentences like “it is wrong to commit murder” cannot be true in any substantial, objective way, independent of human psychology. But many expressivists support a minimalist or deflationist theory of moral truth, where a minimal version of moral truth is guaranteed by the deflationist theory of minimal truth (cf. Schroeder unpublished a). When combining his expressivist theory of moral language and thought with a deflationist or minimal theory of truth, Blackburn seeks to avoid conflict with the view that there may be moral truths, without adopting full-blown realism. The general theory of truth minimalism states that truth is not something that exists in any substantial sense (cf. Paul Horwich 1990), and that the sentence “it is true that grass is green” asserts the same as the sentence “grass is green”, and nothing more substantial is expressed in the first sentence. When applied to a moral sentence, “it is true that murder is wrong” asserts the same as “murder is wrong”, and this might very well be a moral truth, but not in any ordinary substantive conception of truth, where it is believed that asserting “p is true” contains nothing more than simply asserting “p”. The truth predicate is used merely to agree or disagree. By this move Blackburn wants to show that for the expressivist it is not possible to say something universally true about thin moral concepts like “good” or “right”, but still allows for the possibility of moral judgements being true in the first order sense, in given circumstances. “[T]here is no error in our ordinary ways of thought and our ordinary commitments and passions”. (Blackburn 1993: 158) So the sentence “it is wrong for Smith to murder Jones”, which expresses my disapproval of Smith murdering Jones, may be true. One objection to expressivism is that the theory threatens to collapse into a form of moral realism if it incorporates a theory of minimalistic moral truth. (Dreier 2004) Blackburn and Gibbard seem unconcerned about this possible collapse, because whether moral facts exist or not is less important to expressivism. Its focus is the nature of moral thought and language. In his earlier writing, Gibbard held that there were no moral facts, but now he seems to be an agnostic about them. (Gibbard 2003: x)

An objection to any form of realist account of morality, whether minimal or robust, is what is often called “Harman’s challenge”: moral facts do not figure in our best explanations of the world. (Harman 2000: 178) On this view, the fact that Hitler killed a lot of innocent people is not best explained by a fact that he was morally depraved, but by a combination of other psychological factors, e.g. a lack of sympathy or other emotional defects, megalomania, irrationality etc. Sturgeon’s view that moral facts do figure in our best explanations of the
world, accounted for in section 2.3.1., is developed as a response to Harman’s challenge. No matter who turns out to be correct, Harman or Sturgeon, the expressivist seems to be able to accommodate the result either way. If there are no moral facts, he can do without them, and if they exist, they can be explained by the deflationist account. In essence, expressivism can be summed up by the combination of (1) the parity thesis and (2) minimalism about moral truth.

3.1.3. Projectivism and Quasi-Realism

The parity thesis is closely linked to the expressivist programme of “projectivism” and “quasi-realism”. These are constructive accounts of how to utilise the apparent descriptive features of moral reality. “Projectivism” is a way of transforming the contents of moral judgements into a descriptive moral claim, by “projecting” our emotions onto the external world. For instance, if Jones observes someone torturing the cat, he could from certain assumptions about the torturer’s intentions, the cat’s pain, and other factors of the circumstance, come to harbour a strong attitude of disapproval towards torturing the cat. This attitude is then projected onto his experience of the world in the form of a moral judgement that it is wrong to torture the cat. Here the property of “wrongness” is attributed to the act of torturing the cat. Projectivism could be considered a part of the “theory of quasi-realism”.

Blackburn’s programme of quasi-realism is built around the parity thesis. Moral discourse behaves as if it is descriptivist towards representational beliefs, which allows the expressivist to talk about morality in a descriptivist manner, although the account differs on the level of content and meaning, because there are no such descriptive moral beliefs according to this view. This is why it is often labelled “quasi-realist”. (Blackburn 1993: 167). The semantic structure of moral judgements expressed by sentences like “it is wrong to commit murder” and “it is right to give to famine relief” may involve nothing more than a “boo” and a “hooray” operator, thus making them equivalent to “boo to murder” and “hooray to giving to famine relief”. (Blackburn 1984: 193) But because of the parity thesis we can use the “surface grammar” of moral sentences to construct a quasi-realist theory of moral truth, where we introduce “quasi-propositions” to analyse moral sentences as if they had truth-conditions.

In effect, quasi-realism is trying to earn our right to talk of moral truth, while recognising fully the subjective sources of our judgements, inside our own attitudes, needs, desires, and natures. (Blackburn 1984: 197)
Let us assume Jones is torn between two different attitudes towards a given subject matter, that “Ovid is better than Tacitus” and that “Tacitus is better than Ovid”, perhaps caused by a division in literary taste. One the one hand he values passionate and emotionally-engaging texts, but on the other hand he also values philosophical reflection concerning life and the moderation of the passions. Careful consideration of each sensibility and the relevant facts, leads Jones to judge them to be equally good. Quasi-realism works as an “evaluative system” where we can contemplate and discuss what attitudes we should adopt in a descriptive manner. (Blackburn 1984: 201) Blackburn and Gibbard’s expressivist theory could be summed up by the combination of (1) quasi-realism (2) minimalism about moral truth.

Quasi-realism, like moral truth minimalism, could be charged with the possibility of collapsing into realism. Sharon Street has claimed that this would constitute a dilemma for the quasi-realist, if the conjunction of the quasi-realist programme’s success and the existence of moral facts obtain.

Quasi-realism either does or does not achieve its stated aim of mimicking realism. On the one horn, if quasi-realism fails to achieve this aim, then the position fails by its own lights. On the other horn, if quasi-realism succeeds in meeting this aim, then it faces a different problem – a problem which we might characterise most generally as failing to distinguish itself from realism.

(Street 2011: 3)

Gibbard responds to Street’s objection by referring to the quasi-realism deliberation process, which is very different from the realist’s. Quasi-realists do not begin by considering the moral aspects of reality, but by considering which attitudes occur in a given circumstance and what constraints and pressure apply to it. (Gibbard 2011: 43) Again, I do not believe this possible collapse poses any greater threat, and I believe Street is mistaken in believing there is a dilemma. Although quasi-realism works to mimic realism, both the starting point, as Gibbard points out, and the process, are too different for it to be confused with realism if the programme is successful. Even if the relations that obtain between our moral attitudes, moral sentences and non-moral facts were be able to mimic the realist account perfectly, the quasi-realist would still lack an adequate epistemological account of fact detection. If some minimal version of truth for first order moral discourse is correct, it seems much more likely that it will go unnoticed to the expressivist. Blackburn’s “quietist” theory could also be understood as a response to the charge of collapsing into realism. “Quietism” is an attitude towards truth, “which urges that at some particular point the debate is not a real one, and that we are only
offered, for instance, metaphors and images from which we can profit as we please”. (Blackburn 1984: 146).

Whether moral facts exist or not, it would be useful to have an available terminology to deal with moral issues that seek to specify our first order commitments in more general terms. Through quasi-realism, moral truth can be discussed as an objective phenomenon independent of human psychology, regardless of whether there are such truths or not. For the expressivist, the notion of deflationist/minimalist realism and the broader project of quasi-realism are closely linked. By embracing first order deflationist-realism the expressivist can explain the relevance of the overall project of quasi-realism – dealing with our actual moral commitments and beliefs. The success of the quasi-realist programme is, however, tied to the Frege-Geach problem. We need a way to account for the logical structure of moral sentences expressing desire-like states of mind.

3.1.4. The Subjectivism objection
Since the days of the early noncognitivists, one chief objection to all noncognitivist theories has been that they are essentially some form of subjectivist theories, not compatible with the view that we live in a world of moral truths. This objection has also been levelled at modern expressivists like Blackburn and Gibbard (e.g. Jackson and Pettit 1998). Although early noncognitivism is a form of subjectivism, this is not the case for expressivism. If to think that committing murder is wrong only implies that the speaker disapproves of committing murder, then expressivism could be nothing more than speaker-subjectivism. But the answer to this objection rests on the parity thesis. Moral language has the same relation to desire-like states as non-moral descriptive language has to ordinary propositional belief-like states. This implies that since “grass is green” means that grass is green, and not that the speaker believes that grass is green, the same must hold for the moral sentence-moral thought-relation, and disapproval of committing murder does not simply consist of the speaker’s disapproval of committing murder.

A related objection is raised by David Lewis, who accuses quasi-realism of being a type of “moral fictionalism”, where we only “make believe that moral realism is true, though it isn’t.” (Lewis 2005: 319) Moral fictionalism is the view that when we hold that moral judgements are systematically wrong and no moral facts exist, we still make-believe we are making real moral judgements and make-believe that moral facts exist. Some error theorists have
suggested that the best thing to do could be to embrace this make-believe to ensure peaceful and productive coexistence: “Moral judgement thus can function as a kind of social glue, bonding individuals together in a shared justificatory structure and providing a tool for solving many group coordination problems.” (Joyce 2006: 117) Blackburn answers, that for the moral fictionalist “the false content is integral to our practice, which must retreat to make-believe once this falsity is exposed. But the quasi-realist will dissent, because he will deny that a false content is integral to and explains our practice” (Blackburn 2005: 329). To clarify, for moral fictionalism there is error “in our first-order practices” (ibid. 331) but for the quasi-realist there is nothing “erroneous in moral practice per se”, (ibid.) where our first-order practices could be considered true, but only error in “the domain of the theorist”, the second-order realm of moral facts and properties. If we grant the expressivist truth in first order moral practices, which are true simpliciter, quasi-realism is not moral fictionalism.

3.2. Expressivism and Moral Evolution

Assuming naturalism and the general theory of evolution, and believe that human morality may have evolved from more rudimentary aspects of human social behaviour of cooperation and group behaviour, this strengthens the case for noncognitivist or cognitivist anti-realist (error theories). I will now argue how moral sentiments and moral language may have evolved through natural selection and group selection from certain types of pro-social behavioural traits like reciprocal altruism. I will then try to show how a noncognitivist theory of moral discourse and thought might exhibit a higher level of compatibility with this evolutionary story than cognitivist theories.

3.2.1. The Evolution of Morality

If we can show how it is likely that human morality has evolved from primate behaviour and psychology through natural selection, and has an emotional basis, this will in itself strengthen the case that morality is a function of human psychology, and not some other independent entity, which would weaken the claims of both moral realism, and also, as I will argue, cognitivism.

In their essay “‘Any Animal Whatever’: Darwinian Building Blocks of Morality in Monkeys
and Apes” (2000) Jessica Flack and Frans de Waal argue how behavioural parallels between non-human primates and human beings could be drawn in support of biological evolution of human morality. If morality has a biological origin, then there should be similarities in social or group behaviour between primates and humans. And a range of observations and animal experiments has shown that there are considerable similarities of cooperative behaviour between monkey, ape and humans, especially when it comes to “resolving, managing and preventing conflicts of interests within their groups” (Flack and de Waal 2000: 1).

Flack and de Waal name the basic social traits or properties that primates have and humans shared at an earlier evolutionary stage, Darwinian Building Blocks:

Any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man.

(Darwin 1982 [1871]: 71–72 quoted in Flack and de Waal: 1)

Thus, these building blocks would thus be the evolutionary basis that a system of morality would develop from, and include “reciprocity and food sharing, reconciliation, consolation, conflict intervention, and mediation, […] reciprocity, empathy, sympathy, and community concern, […] a sense of justice, […] and perhaps even the internalization of social norms” (Flack and de Waal: 3). This is a rather extensive list of character traits similar to those found in humans, but even though non-human primates “may not be exactly moral beings”, they do show “indications of a sense of social regularity that parallels the rules and regulations of human moral conduct” (ibid.).

One of the central building blocks is reciprocal behaviour, often observed in relation to resource distribution, where sharing is very common among primates. There are three typical explanations for food sharing: 1) the sharing-under pressure or tolerated theft hypothesis, where the possessor allows food to be taken and does not give it away, 2) the sharing-to-enhance-status hypothesis, where food becomes a means to bribe their way to an enhanced standing in the group, and 3) the reciprocity hypothesis, where sharing food is part of a mutual obligation system that includes both resource and favour (e. g. grooming or protection) exchange. (ibid.: 4–5) The reciprocity hypothesis is, however, held as the best explanation for food sharing, even though the reason for sharing might sometimes be a combination of the three (ibid.: 7).
There is also data in support of a calculated form of reciprocity, not with regards to food sharing, but observed with regards to conflict situations, where individuals discriminate between group members. Both chimpanzees and some macaque species have been observed to exhibit this behaviour. For instance, “if A intervenes in favour of B, B is more likely to intervene in favour of A” at the next encounter. (ibid.: 8)

Both conflict intervention (beneficial or aggressive), mediation and reconciliation by a third party are common occurrences among primates, and suggest a somewhat advanced grasp of the social scene. Flack and de Waal suggest that the most interesting conflict-resolving situations occur when an impartial third party tries to resolve the conflict without choosing sides, apparently only in order to restore peace. They think that this behaviour is an “indication that a rudimentary form of justice may exist in the social systems of non-human primates”. (ibid.: 12)

Both food sharing and conflict interventions, as well as other situations where empathy or sympathy are displayed, e. g. for consolation after losing a conflict, seem to suggest that some sort of community concern has evolved. (ibid.: 14)

Flack and de Waal summarise the Darwinian building blocks into “four ingredients of morality” (ibid.: 22): 1) sympathy related, 2) norm related, 3) reciprocity and 4) getting along.

According to this view, a rudimentary morality system evolved from a set of pro-social emotions also present in non-human primates, as “primate groups individuals are motivated to respond to others based on the emotional reactions they have to one another’s behaviour”, where “sympathy, based on empathy, seems to direct the emotional responses of some primates to others”. (ibid.: 20) I believe that evolutionists have built a strong case for how pro-social emotions, cooperative behaviour and group structure among primates and our early ancestors must be a part of certain “Darwinian building blocks” from which a morality system must evolve, and that could be taken as evidence for a common social and emotional starting point for all primates, and which constitute the evolutionary origin of human morality.

As the emotional range and minds of our ancestors expanded and developed, so did the moral sphere and its different forms of expression. As we have seen the early foundations of
morality lay in simple pro-social emotions of guilt, shame, resentment, empathy and sympathy, produced by natural group selection as a result of between-group competition, which resulted in different forms of altruistic behaviour being adapted to increase survival and reproductive success. Everything concerning morality was closely connected to emotions – and, for our early ancestors, being a moral creature might simply have consisted of having certain pro-social emotions.

3.2.2. The No-Missing Link Account of Moral Evolution

At the early stages of moral evolution morality consisted of simple emotions, with only unreflective content lacking in any linguistic expression, as described above, and if the cognitivist grants that morality is the result of an evolutionary process and did not suddenly spring into existence, he must also accept that in the beginning it consisted of these unreflective mental states of simple emotions, and only later developed into more complex moral sentiments or states of mind. I believe this creates large explanatory obstacles for the cognitivist, who would have to come up with a reliable story of how the moral outlook suddenly shifted from that of personal and interpersonal emotions to beliefs about an objective moral reality.

Even greater problems emerge for the cognitivist who is also a moral realist, who would hold that morality suddenly spawned or mutated into something fundamentally different, i.e. that it suddenly went from being constituted in some way by our own emotions into being non-human entities and emotionally-detached facts about nature – independent of the emotions that originally spawned them – almost like some divine entity extracted the moral essence from our emotions, and sprayed it onto the external world, thus creating detached moral facts. I find it hard to envision such a story. I find it so hard that I believe it is safe to assume that morality developed along a continuous line – continuous with respect to nature, and thus also continuous with respect to our emotions, from which it originated, and where it is constituted and from which it cannot be separated. I suggest that we call this the no missing link-account of moral evolution.

For realism to be true, given the theory of moral evolution, we would have to be able to offer an account of the link missing somewhere in our evolutionary past that would bridge the emotional–extra-emotional gap that would explain the evolutionary history of moral facts from something mind-dependent to something mind-independent.
There is an argument from the evolution of moral language that could both work as an argument in favour of cognitivist anti-realism (error theory), but also as an argument in favour of the expressivist view that moral language only appears to be descriptive. The objectivist- and descriptivist-like structure of moral language, which makes it appear like moral sentences are used to express beliefs about facts, like ordinary descriptive language, is often taken as an argument in favour of cognitivism. However, the general theory of evolution might provide another explanation independent of the theory of moral evolution. In the same way that language in general is believed to have developed in order to improve communication and cooperation, and thus higher linguistic competence were selected for, it would not be unreasonable to think that the adaptation to a moral language that appeared to treat moral sentences as expressing objective facts would be group beneficial and thus increase the inclusive fitness\(^{28}\) of the individual. Certain types of behaviour licenced by moral utterances would then have higher authority than requests and commands uttered in a non-moral language, and be useful tools for leaders in controlling social groups, as they would be commands whose authority would appear to lie outside the speaker who issued them. And they would function as threats of punishment should they be breached. (cf. Joyce 2006: 117)

So the appearance that moral sentences express beliefs about objective facts is fully compatible with the view that they really just express our mind-dependent conative attitudes, thoroughly grounded in our human psychology.

If we assume moral evolution, I believe the no missing link-account to be an intuitively strong argument against moral realism. I believe the evolutionary store of pro-social emotions that developed into strong moral feelings at the centre of our moral discourse, favours noncognitivism and other cognitivist theories (e. g. error theory). I also believe modern expressivists like Allan Gibbard is particularly advantageously positioned to explain why cognitivism is false even precisely on the level of moral language.

All cognitivists share the view that moral thought consists of cognitive belief-like states in some way separated from our emotions, and that the meaning of moral sentences is in some way given by their truth conditions, that is, their successful or unsuccessful reference to moral

\(^{28}\) Inclusive fitness describes an individual’s fitness by the contribution it makes both to his own survival and reproductive success and to that of the social group he is a part of. (cf. West, S. A., C. E. Mouden and Gardner 2011: 232)
facts, captured (or not) by the propositions they express. The new move by the expressivist – in contrast to older noncognitivists like Ayer, Hare and Stevenson, who defined the meaning of a moral sentence as the speech act it was supposed to express – is that they define the meaning of moral sentences as the content of moral thought, and no reference to the external world is needed or called upon to speak about morality. And it is precisely the invocation of something mind-independent and external which renders cognitivism implausible in any moral evolutionary respect.

Moral judgements according to Gibbard are not in themselves feelings, but “judgements of what moral feelings it is rational [29] to have” (1990: 6), and when we make these moral judgements the state of mind we are in is that of accepting a norm. For instance, “to think that compassion is good is to accept a norm that says to desire compassion” (Gibbard 2003: 7). The evolution of morality is then nothing more than the evolution of such internalised norms within a group, a group of groups, a community and within society. (cf. Flack and de Waal: 3).

Thinking about morality along Gibbard’s lines – as norms – makes it very simple and effective to explain moral judgements causally by their evolutionary story. Every step of the ladder of moral evolution represents a different level of evolved and internalised norms. At earlier evolutionary stages, before we could talk, perhaps before we could reason morally and accept norms rationally, our pro-social emotions and group behaviour based on cooperation, reciprocity and altruism could be described as expressions of social, and perhaps even moral norms. Gibbard would then say we were in their “grip” (cf. Gibbard 1990: 60), acting on internalised unreflective but psychologically committing norms, formed and shaped by social, cultural and biological evolution.

### 3.2.3. Debunking Arguments

There is a bundle of powerful arguments meant to favour moral anti-realism over realism from a moral evolutionary perspective, the so-called “evolutionary debunking arguments”. According to Guy Kahane we can describe the general form of most debunking arguments as follows:

---

29 Gibbard’s definition of calling something “rational” is “to endorse it in some way”. (Gibbard 1990: 6)
Causal premise. We believe that p, an evaluative proposition, because we have an intuition that p, and there is an evolutionary explanation of our intuition that p.

Epistemic premise. Evolution is not a truth-tracking process with respect to evaluative truth.

Therefore, we are not justified in believing that p. (Kahane 2011: 111)

The best-known version of a debunking argument is perhaps developed by Sharon Street (2006):

The evolutionary history of morality explains why we have the moral/evaluative beliefs we have (causal argument), but since these judgements do not track the truth of any detached/mind-independent moral facts, but have instead undergone a slow and continuous process of adaptation in response to the external circumstances of human life (epistemic argument), moral realism must be false (conclusion). That means we are left with an anti-realist form of moral cognitivism – scepticism, nihilism or a form of error theory.

Street’s debunking argument (2006) works to explicate how the moral evolutionary story is non-compatible with a story of an external mind-independent moral reality that we might have moral knowledge of. Guy Kahane argues how most debunking arguments assume an overall metaethical objectivism in order to make the argument work (2011: 112), but it remains to see whether some form of objectivist structure survives at all if it does.

In order to examine whether the attribution of cognitivism to morality is applicable from an evolutionary perspective I will step into the core of the object of the debunking argument. I have allready discussed how the gradual and continuous evolution and cultivation of our emotions into something we can call morality discredits realism, and also how it might discredit cognitivism. Now I will consider an objection to debunking arguments, that I believe fails as it succumbs to an inductive fallacy when applied to what Kahane calls the global evolutionary debunking argument (2011: 114), which is the argument that is meant to show

---

30 The inductive fallacy is an informal fallacy that would be committed if some property that is identified within an unrepresentative sample space of objects is attributed to the whole population.
how all moral judgements are shaped by our evolutionary history (e.g. by Street 2006), and not just applied to a specific area of morality or some given judgements.

There is an argument against evolutionary debunking arguments that focuses on the difference between explanation and justification, which roughly concerns the observation that the justification of some claim need not have a relevant relation to its causal explanation. This is then meant to show that even though our moral judgements may be causally explained by our evolutionary history, it does not follow that they cannot be justified without appeal to their own causal history.

As an example of an argument of the same form, Kahane mentions the recent speculation that Marx’s theory of alienation grew out of his own feeling of being alienated from others, caused by the fact that he suffered from hidradenitis suppurativa, a deforming skin-disease that can cause depression and self-loathing, which of course has nothing to do with the justifying reasons he submitted in favour of his view, that need to be examined to prove him wrong. (2011: 105)

This objection is a typical companion-in-guilt argument, which seeks to identify an argument with a certain formal argumentative structure in ethics with an argument concerning a case of an unrelated topic, but with a similar argumentative structure. If the argument fails for the unrelated case, that is taken as evidence that the argument concerning ethics must also be false. (cf. Lillehammer 2007) This may be a dubious strategy without necessary considerations concerning the relevant differences of the subject matters involved in the arguments – for instance anti-realism about ethics does not imply anti-realism about science.

This objection could hold for a local version of the evolutionary debunking argument: that is an argument concerning a single or some subset of moral judgements. However, I do not believe it is a threat to global versions of the debunking argument. The causal explanation of some moral judgement analysed in isolation may be irrelevant for its justification, but it seems that a generalisation of this argument would be a textbook example of the inductive fallacy. That the causal explanation of all (or most) moral judgements should be irrelevant to their justification, because they might be completely unrelated, would seem highly unlikely. As Street argues, if the realist denies that there is a relation between the causal evolutionary history of our moral judgements and moral facts, then if most of these judgements were to be
true it would be an incredible coincidence – “this would require a fluke of luck that’s […] extremely unlikely” (Street 2006: 122)

I have showed how moral evolution favours both anti-realism, supported by debunking arguments, but also how it supports a noncognitivism emotional account of morality, and how Gibbard’s theory of norm acceptance fits especially well with the theory of moral evolution. Evolution has shaped us into the beings we are today, with all our sentiments and propensities of acting certain ways that we label “good” or “right”. And the totality of these sorts of sentiments, pushing us towards action of the good or right kind, is what we label morality – something that has been formed, nurtured and sewn into the fabric of our genetic make-up by millennia of biological and cultural adaptation – moral norms developed and internalised by our ancestors, that has been past down to us by our genetic and cultural inheritance.

The combination of the theory of minimalist truth and the programme of quasi-realism offers a clear and simple unified naturalist account, which in addition to supporting a standard noncognitivist take on the practicality of morality, where the desire-like contents of moral judgements motivates intrinsically, offers an original account of the objectivity requirement through the combination of minimalism about moral truth and the quasi-realist programme. But without a solution to the Frege-Geach problem, expressivism will be an incomplete theory, as we the tools to make the formal analysis necessary to render satisfactory analyses of moral sentences.
Chapter 4

Hybrid Expressivism

4.1. Hybrid Cognitivism versus Hybrid Expressivism

In the last years various hybrid cognitivist and noncognitivist theories have been developed. On these views a moral sentence like “to commit murder is wrong” express or involve both a desire-like state and a belief-like state of mind to some extent. For the hybrid expressivist, introducing belief-like states that have propositional contents makes it possible to utilise a standard propositional approach to sentential analysis. With regard to a moral sentence like “it is wrong to commit murder”, it is the introduction of the proposition that it is wrong to commit murder which provides the expressivist with a new apparatus that may solve some old problems.

Hybrid expressivism might prove useful for advancing the expressivist project by offering new or improved solutions to some of the traditional problems of expressivism. In this part, we will consider three subjects of controversy where we might profit by applying the theory of hybrid expressivism: The Frege-Geach problem, in accounting logically for simple sentences as part of a complex moral sentence, The problem of moral disagreement, how to account for differences in moral judgements, The Moral Problem, particularly remaining problems concerning questions of practicality and motivational internalism, and a solution to the problem with the amoralist, and questions The hybrid expressivist might find improved solutions to these challenges by introducing the essential claim that moral sentences also express some sort of belief, or subspecies of belief, whose contents are propositions.

A number of both cognitivism and expressivist hybrid accounts have been proposed the last few years. The motivation for a hybrid cognitivist account could be described as the

---

31 Notable Hybrid expressivist accounts have been developed by Horgan and Timmons (2006), Ridge (2007, 2014), Boisvert (2008) and Schroeder (2009; 2010a; 2013b). A notable Hybrid cognitivist theory that we will consider has been developed by Copp (2007; 2011).
aspiration to satisfy the practicality requirement, by supplementing cognitivist realism with a more acceptable theory of motivation. The motivation for a hybrid cognitivist account could be described as the aspiration to satisfy the objectivity requirement, by supplementing expressivism with a more acceptable account of the objective features of morality. Most of the theories do not depart significantly from cognitivism or expressivism, so the main arguments for each camp largely remains as described in previous chapters. We will now briefly consider the two different branches of the theories.

Another motivation for any hybrid theory is a similarity between moral terms and pejorative terms, or “slur terms”. Hybrid theories are able to explain the double character of such terms better than single-attitude theories. Consider the German derogatory term “kraut”, derived from Sauerkraut. If Smith utters the sentence “Jones is a kraut”, he may express the descriptive belief that Jones is a simple, rural, uneducated man, but at the same time the sentence expresses an attitude of disapproval towards kraut, which is a desire-like state. This has inspired many to support the view that slur terms, are subject to “conventional implicatures”, which is the theory developed by H. P. Grice that certain sentences have a conventional meaning that implicates certain attitudes, in addition to their descriptive meaning. (Grice 1989: 25–26)

In the following two sections I will briefly discuss one cognitivist and two expressivist hybrid accounts before I move on to a more general discussion of how a hybrid account should be constructed to best answer the practicality and the objectivity requirement within the expressivist programme.

4.1.1. Hybrid Cognitivism
There seems to be significant similarities between slur terms and moral terms, which has lead to the development of several hybrid metaethical accounts holding that moral terms functions in a manner similar to slur terms. According to these account, if Jones utters the sentence “it is wrong to commit murder”, he expresses the belief that it is wrong to commit murder, but the sentence also expresses an attitude of disapproval towards murder. David Copp and Daniel Boisvert are two proponents of this similarity thesis.

In Copp’s view, moral beliefs are a species of ordinary beliefs, which involve desire-like attitudes:
My proposal was that a person who makes a moral assertion using a moral predicate normally thereby expresses a conative attitude in that she conventionally implicates that she has such an attitude. […] Conventional implicatures are determined, as Grice says, by what the audience must “assume the speaker to think in order to preserve the assumption that she is following the Cooperative Principle” – that she is saying something that is relevant in the context – and that she is otherwise following the “conversational maxims.”


In Copp’s view, someone who utters a moral sentences signals that the he has the attitude conventionally implicated by the sentence. However, having a moral belief does not entail having a desire-like attitude. (Copp 2007: 155) This implies that to utter a moral sentence that expresses a belief only signals to the audience that the speaker has an accompanying attitude, in virtue of the conventional meaning of the sentence. That the meaning of the sentence conveys the attitude to the speaker does not entail that the speaker possesses the attitude. So Copp’s claim to motivational internalism must be considered a weak, where it is a contingent matter whether the speaker of a moral sentence has the attitude implicated conventionally. If Smith states “it is wrong to commit murder”, Jones will assume that Smith disapproves of murdering, conventionally implicated by the sentence and according the cooperative principle, but he might have an opposite attitude towards murder, or no attitude at all. This is a weakness of Copp’s theory in the accommodation of the practicality requirement.

Copp refers to his theory as Realist-Expressivism, which can be described as the conjunction of cognitivism and internalism. His realist account bears some similarities to Smith’s, and can be described as a constructivist and rationalist, where “morality has the function of making society possible by providing rules governing our lives that, when they have currency in society, enable society to meet its needs.” (ibid.: 23)

4.1.2. Hybrid Expressivism

Boisvert’s view, called “Expressive-Assertivism”, is a conventional implicature theory, but with stronger relations between moral sentences and attitude-expressions than Copp’s theory. There is a necessary connection between the attitude conventionally implicated by the meaning of a moral sentence and the attitude possessed by the speaker. According to Boisvert, moral belief is, inter alia, having an attitude of approval or disapproval. (Boisvert 2008: 171) This implies that according to Boisvert’s view, the speaker is also in possession of the attitude
that is conventionally implicated by the utterance of a moral sentence. If Smith utters the
moral sentence “it is wrong to commit murder”, this entails Smith having the attitude of
disapproval towards murder conventionally implicated by the sentence. A challenge for
Boisvert’s theory is its strong internalist claim, which might make it difficult to solve certain
motivational issues, as the amoralist problem. One the other hand, the strong relation between
the moral belief and the attitude expressed by moral sentences may make solving the Frege-
Geach problem easier. (ibid.: 197) Boisvert believe that this relation guarantees that no
compositional theory of meaning for attitudes is needed when considering embedding issues,
because every time someone utters a moral sentence the speaker expresses an attitude. (ibid.:
188)

According to Michael Ridge’s hybrid theory, called ecumenical expressivism, normative
judgements, or moral judgements, are relational states, which can be realised in a number of
different ways. (Ridge 2014: 6) These states are hybrid states, comprised of a belief-desire
pair, which bear some special relation to one another, where “there will be no representational
content associated any given normative judgement.” (ibid. 7) This account of the desire-belief
is pair is perfectly compatible to a familiar expressivist account of moral language and
metaphysics, where the meaning of moral sentences is the content of moral judgements. As
mentioned in chapter 3, expressivism could be understood as a meta-semantic theory about
the nature of moral sentences and thought, and not a semantic theory about the specific
contents of sentences and judgements. (ibid.: 105) Understood in this way, Ridge is free to
combine a general expressivist theory of meaning with, for instance, a correspondence theory
of truth for first-order moral issues concerning specific moral sentences and judgements. This
compatibility need not mean a departure from a deflationist theory of truth, only that the
ecumenical expressivist is not committed to it. (Ridge 2014: 107)

Ridge emphasises the general metaphysical commitments of all expressivist theories, that it
an approach that the starting point and the main motivation for adopting any expressivist
account, are the affect-laden and practical aspects of morality. And it is precisely this
commitment to the desire-like nature of moral judgements, that makes the hybrid expressivist
“reject the cognitivist’s representationalist order of explanation.” (ibid. 101) Another feature
of Ridge’s theory is that his theory is compatible with a context-sensitivity account of both general normative words like “ought” and “should”, but also moral words like “good”.

Ridge suggests that the meaning of moral terms should be understood in terms of standards of practical deliberation, where these standards are fixed by the context of utterance. According to Ridge’s account, the sentence “it wrong to commit murder” would then express a relational state, with both a normative perspective disapproving of murder and a belief-state containing the proposition that it is wrong to commit murder. Whether the belief component has representational contents is a first order question.

An objection to hybrid expressivist theories is the possible collapse into realism. When moral sentences express a relational state in the form of a belief-desire pair, it seems to be an important point for any hybrid expressivist theory that the belief-element does not take primacy. I believe that this would not be a problem for simple, parsimonious uninflated and truth-uncommitted accounts, like the ecumenical expressivism of Ridge. As long as the hybrid expressivist can refrain from postulating objective moral facts, a first order instantiation of moral truth in specific circumstances under given conditions would mean the end of expressivism.

4.1.3. Moral Propositions

Mark Schroeder has developed an influential account of expressivism in which he introduced an general attitude of “being for”, as a sort of universal state of mind that could take other states as its object (e.g. desires, beliefs, intensions etc.). (Schroder 2008a) Part of the motivation for this move was to enable the expressivist to make logical operations on sentences expressing noncognitive attitudes, where this function is filled by propositions with regards to descriptive sentences. Schroeder has in his most recent writing retracted this position, and holds now that both pure and hybrid expressivist could understand moral sentences as expressing propositions. The introduction of propositions would also expand the expressivist theory of meaning, in which moral sentences not only would receive their meaning from the mental states they express, but also through the propositions they would...

32 Many metaethicists agree that both normative and moral terms are context-sensitive or relative in some sense. Confer Björnson and Finlay, who argue that moral and normative terms are relative to a context of utterance both with regard to information and a standard of assessment (2010), and MacFarlane, who argues moral and normative words are relative to both a context of use and a context of assessment. To pursue this issue further would, however, lay outside the scope of this theses.
express. (Schroeder 2015: 2) If ordinary expressivists would take moral sentences to express propositions, this could, among other thing, help them to solve the Frege-Geach Problem. One possible way to introduce proposition for pure expressivists would be through the quasi-realist program. Hybrid expressivists can take moral sentences to express propositions because in is an integral part of their theory that moral sentences also express a belief component, where the meaning of a belief is the meaning of the proposition it expresses.

According to the standard semantic theory propositions could be considered abstract entities whose two main roles are 1) to be the objects of cognitive mental states and 2) the fundamental bearers of truth and falsity (Lycan 2008: 69–70). In as far as propositions are bearers of truth and falsity, they may have a relational role to descriptive content. Ordinary non-moral beliefs, like the belief in the proposition that grass is green, may be representational beliefs, and have relations to both propositions and representational content (facts about the world).33

Simplicity is a virtue of any theory, and we should aspire to make our hybrid views as simple as possible. The simplest form of hybrid expressivism could be conceived as only making one essential claim combined with normal expressivism, and this is a claim applied to moral semantics: a moral sentence expresses a belief component in addition to the attitude component, whose contents is propositions. This would make it descriptive not just in an as if way in terms of surface grammar, as it is regarded by normal expressivists/quasi-realists. We could treat moral propositions as a semantic construct applied for the purpose of analysis, and, as Ridge claims, the belief component would not need to have representational contents and be truth-apt in any non-deflationary way. Moral belief-states would have the same structure as non-moral representational belief-states, but we could regard their apparent representational content only as content of moral thought (cf. Schroeder 2013b). In whether they actually also have relations to representational contents, or facts about the world, we could join a quasi-realist like Gibbard and remain uncommitted. (Gibbard 2003: 18)

33 There are other theories of propositions, among them a view supportet by Ridge, that propositions are “cognitive event types” of prediction, that are essentially representational. (Ridge 2014: 126–127)
The Frege-Geach problem is the problem of how the meaning of complex moral sentences is determined by the meaning of its parts, when all levels of the sentence lack propositional contents suitable for conceptual or logical analysis. The problem is the problem of accounting for the meaning of moral sentences expressing desire-like attitudes, in which simple attitudinal moral sentences are embedded in complex moral sentences, and arises from the difficulty of treating desire-like attitudes semantically and logically when they appear as expressions of complex sentences involving logical operators like negations, quantifiers and inference-operators. This is not a problem for cognitivist theories, and this might not be a problem for hybrid expressivist theories if moral sentences were also understood to express propositional belief states able to figure formally and consistently in analyses. (Schroeder 2013a: 411) Caj Strandberg has made the case that there embedding problem can be generalised to apply to all metaethical accounts which involves desire-like states of mind, like approval or disapproval. The problem applies to accounts in which the meaning of moral sentences consists in the mental state of disapproval or approval expressed by the sentence. (Strandberg 2015b: 8)

This difficulty was first addressed by Peter Geach (1965), and an often-used example based on his original paper might better illustrate the problem. Consider the following argument:

a) It is wrong to tell lies.
b) If it is wrong to tell lies, then it is wrong to get your little brother to tell lies.
c) Therefore, it is wrong to get your little brother to tell lies.

The problem for the pure expressivist is a problem of analysis. Since the moral judgements expressed by sentences a) and b) consist of desire-like attitudes, the meaning of the simple sentence “It is wrong to tell lies” must be different when it appears alone and asserted, than when it is embedded in the complex sentence “If it is wrong to tell lies, then it is wrong to get your little brother to tell lies”, in an unasserted context. The problem for the expressivist consists in accounting for the different meanings of these two sentences, when the attitudes towards lying expressed by both are similar attitudes of disapproval towards lying. Since they both express attitudes of disapproval without descriptive content, there is no simple way of showing how the meaning of the complex sentence might be derived from the meaning of the
simple sentence. For the cognitivist there is no problem, as each sentence expresses propositional content, and the meaning of the complex sentence derives logically from the meaning of its parts, where one part contains the same proposition and has the same meaning as the simple sentence. The expressivist has no satisfactory tool for deriving the meaning of the complex sentence through the meaning of its parts, and is unable to account for the relation between the two attitudes expressed by the sentences. By allowing the sentences to also express a belief-like state, the hybrid expressivist would allow for a logical analysis of propositions in the same way as the cognitivist. There still remains the problem of showing how the different states expressed by each sentence relate to each other, but if it is possible to develop a hybrid expressivist account that shows how the meaning of the belief-like state and the desire-like state stand in relation to each other, there might be a solution to the problem.

Consider the negation problem for pure expressivists. This example is based on an account developed by Unwin (1999, 2001), refined by Schroeder (2015: 6). The problem involves accounting for what kind of attitudes the negation of the following sentences would generate:

1) Sophia believes that stealing is wrong.
2) Sophia does not believe that stealing is wrong.
3) Sophia believes that stealing is not wrong.
4) Sophia believes that not stealing is wrong.

On the expressivist account, to believe that stealing is wrong is to disapprove of stealing, and the negation of 1) – 4) gives us:

1*) Sophia disapproves of stealing.
2*) Sophia does not disapprove of stealing.
3*) ???
4*) Sophia disapproves of not stealing.

As we see, the negation of “3)” produces no comprehensible attitude, because attitudes have less structure than descriptive belief. (Schroeder 2013a: 411)

Ridge’s proposed hybrid solution to the Frege-Geach problem is simply to disregard the noncognitive attitudinal aspects of the belief-desire pair expressed by moral sentences:
Because logical complexity is built into the content of the representational component of moral judgements, there is no reason to worry that the account cannot properly distinguish logical errors from other sorts of errors. For now someone who accepts contradictory normative judgements will come out as having contradictory beliefs in a perfectly robust and representational sense of “belief”. There is no pressure to embrace a “logic of attitudes,” nor is there any need to explain logical validity or logical inconsistency in terms of being able to satisfy all of your commitments. (Ridge 2014: 144)

If the hybrid expressivist were able to treat sentences like “If it wrong to tell lies, then it is wrong to get your little brother to tell lies.” solely as descriptive belief, involving normal logical sentential operators, I believe Ridge’s solution could work, because the problem would disappear. But being expressivist, whether hybrid or pure, implies taking the attitudinal content of moral judgment as its primal moral component, which entails the expressivist theory of meaning. Although the meaning of moral judgement could be determined to by representational contents and truth-conditions on hybrid expressivist accounts, it should at least to some extent be determined by the content of moral thought and judgement for any expressivist. Strandberg has made an argument that supports this intuition about expressivist meaning, where he claims that in virtue of their expressivist account of meaning, hybrid expressivists must account for logical inference and validity with regard to moral sentences by showing how each relevant attitude relates to the sentences. (Strandberg 2015a: 94)

We have found that there is no simple solution to the Frege-Geach problem for the expressivist, by appeal exclusively to the belief-component expressed by moral sentences. As we saw in the negation case, performing logical operation on attitudes produces uncertain results, as attitudes, like the attitude of disapproval, seem to have less structure than descriptive beliefs.
4.3. Concluding remarks – results and unresolved issues

We will now briefly discuss how some remaining problems concerning the issue of disagreement and the solution to the moral problem may be improved upon by hybrid expressivism.

4.3.1. Disagreement

It is a problem that people seem to genuinely disagree about moral issues in cases where there is agreement on all relevant natural facts. As we saw in section 1.5, this is a particular problem for versions of moral realism that are susceptible to the open question argument. This has led many to believe that expressivism explains deep moral disagreement better on the psychological level than moral realism, where a clash of attitudes of some desire-like states of mind would have more explanatory force than conflicting belief-like states. And thus, if we assume perfect information of all relevant natural facts in a given moral circumstance, at least one participant does not need to be irrational in explaining moral disagreement from an expressivist point of view, where it is a result of a difference of noncognitive mental states. We will also be able to meet the cognitivist charge that moral disagreement presupposes belief-like states, and that differences of moral judgement based on desire-like states cannot constitute real disagreement.

Still maintaining that disagreement in moral judgements fundamentally consists of a clash of attitudes, the hybrid expressivist could claim that this difference of opinion would be analysable in the same way as for ordinary descriptive disagreement. If we also allow moral sentences to express a belief-like state of mind determined by propositional content, it could be possible to locate actual disagreement between different desire-like attitudes, assuming the parity thesis, by analysing disagreement in the propositional contents of their related belief-like states. However, a satisfactory account of disagreement presupposes a solution to the Frege-Geach problem, or the development of a theory of attitudinal logical operations. A solution to this problem would be an account of genuine moral disagreement, which surpasses the objection from the cognitivist camp that the expressivist cannot account for the relationship between conflicting attitudes, where the shape and structure of the individual attitudes depart from the shape and structure of descriptive beliefs.
4.3.2. The Moral Problem

The Moral Problem has long been a considerable challenge for both cognitivist and noncognitivist theories. As we saw in chapter 1, it concerns the difficulty of combining an account of objective moral truths (independent of our different states of minds) with the fact that making moral judgements, such as that it is wrong to commit murder, has a genuine motivational force that compels us to action that does not lead to murder. This problem originated from Hume, who, mildly reformulated, held that some desire-like attitude must be present in motivating us to action, and not just a belief-like state of mind. The problem presented the cognitivist in particular with some difficulty in explaining how the belief that it is wrong to commit murder should at the same time give us motivating reasons not to commit murder, when these beliefs refer to some external realm of moral facts, and not to our own feelings about murder.

The expressivist solution seen from the Humean perspective is simply the unified story of how our disapproval of committing murder is an emotional state of mind with an intrinsic motivational force – and, like all aspects of our human psychology, explainable in naturalist terms. An old problem for ordinary expressivists and other motivational internalists, however, is the possibility of the amoralist. As we saw in section 1.4 it concerns the question of making a moral judgement that something is right or good, but having no motivation to act upon it (without being irrational). A common response to the amoralist challenge is to argue that the amoralist really is being amoral, and does not make genuine moral judgements, due to some defect or other in what we could call his moral compass or empathic abilities, commonly found in different variations of psychopaths (Hare 1952: 124–125, Smith 1994: 67, Blackburn 1998: 208–209, Timmons 1999: 53). The problem for the internalist is presented by cases in which it is possible to make a moral judgement without being irrational and without having a corresponding motivation for action, as I argued in chapter 1.4.2. As Brink puts it, as long as the challenge of the amoralist is intelligible, we should take it seriously (1989: 59). If motivational externalism turns out to be true, that it is possible for a moral judgement to lack all motivational force, this could be a situation more plausibly defendable by a hybrid view than for pure expressivism. Perhaps a hybrid view with a cognitive belief component could account for this situation in a similar way that a cognitivist would, where the desire component might be disregarded in some sense that parallels the lack of motivation. But as Strandberg demonstrated with regard to the Frege-Geach problem, even for the hybrid expressivist uttering a moral sentence implies the presence of a desire-like attitude.
(Strandberg 2015a: 94) We must conclude that all versions of expressivism appear to have problems accounting for the amoralist.

4.3.3. Summary
In chapter 1 we started out by outlining the different methodological approaches to metaethical inquiry, and reviewed the different general observations concerning the moral aspects of reality, and how these observations have been taken to support of different metaethical views. We assumed that any metaethical theory should be compatible with a naturalist worldview. Moral and metaethical considerations have historically placed philosophers in one of two metaethical camps – cognitivism and noncognitivism. We then discussed the distinctive features and properties of each theory-camp, to be able to draw some conclusions regarding the structure and content of the relevant theories. We found that early versions of noncognitivism were the first modern metaethical accounts that seriously committed to emotional aspects of moral practice and discourse.

This emotional basis seems to motivate our choices and influence our moral judgements in a crucial manner, which can be formulated as the thesis of motivational internalism. These considerations concerning the desire-like nature of our moral judgements, specifically how what we judge to be right could motivate us to action, was then formulated as a practical requirement that needs to be accounted for by any respectable metaethical theory. We considered an important challenge to motivational internalism in explaining the case of the amoralist. The first chapter was concluded by introducing another requirement metaethical theories need to account for, the objectivity requirement. Any respectable theory should be able to offer a unified account of the practical and action-guiding character of moral judgements and the objective features of moral discourse and practice – The Moral Problem.

In chapter two we proceeded to examine how two different type of realist theories sought to solve the moral problem. We argued that both reductive and non-reductive accounts of moral realism were susceptible to variations of the open question argument. We showed that Michael Smith’s solution to the moral problem was vulnerable to certain arguments from moral diversity, and concluded that there was considerable difficulty involved in developing a moral realism solution to the moral problem.
In chapter three we discussed the expressivist approach to metaethics, where the combined theory of quasi-realism and truth-minimalism could be considered as unified account of both the practical and the objective features of morality. We examined some of the most serious objections to expressivism, and found that moral evolution as an isolated strong argument in favour of anti-realism in general and expressivism in particular. Although expressivism accommodates the practicality argument as an integral part of its theoretical structure, its quasi-realist answer to the objectivity requirement is a form of constructivism that we hoped to see improved.

In chapter four we have seen how hybrid cognitivist and expressivist accounts expands on the theoretical base of their respective parent-theories, where moral sentences express both a desire-like component and a belief-like component. We focused on the simple hybrid expressivist theory, as the theory best suited to evolve the expressivist programme. An obvious advantage for a hybrid expressivist theory is its ability to account for the descriptive and objective features of moral language and apply propositions as part of its semantic theory. This could be considered an advancement of the quasi-realist program, as it allows the hybrid expressivist to employ a truth conditional-semantics in the analysis of moral sentences. Despite these developments we did not manage to find a solution to the Frege-Geach problem by appeal to the propositional belief-component of the hybrid expressivist theory, but there might be reason to believe that the employment of propositions could lead to a more applicable account of logical operations on sentences expressing attitudes as part of their expressed states.
References


