Expressing Moral Belief

Sebastian Hengst

A thesis presented for the degree of
Doctor of Philosophy

Faculty of Philosophy,
Philosophy of Science and Religious Studies
LMU Munich
Germany
2022
Disputation: 16.02.2022

Supervisor: Prof. Dr. Dr. Hannes Leitgeb
Second supervisor: Prof. Dr. Christine Tiefensee

Expressing Moral Belief

Sebastian Hengst

FOR *Jenny, Rosalie, Elisa & Philine*

Abstract

It is astonishing that we humans are able to have, act on and express moral beliefs. This dissertation aims to provide a better philosophical understanding of why and how this is possible especially when we assume metaethical expressivism. Metaethical expressivism is the combination of expressivism and noncognitivism. Expressivism is the view that the meaning of a sentence is explained by the mental state it is conventionally used to express. Noncognitivism is the view that the mental state expressed by a moral sentence is a noncognitive, desire-like state.

The central question of this dissertation is this:

> *How can metaethical expressivists explain that we are able to express moral beliefs?*

The shortest answer this dissertation will give is the following:

> *by becoming 'dispositional expressivists'.*

Dispositional expressivists hold that moral beliefs are dispositions to be in certain mental states that resemble desires. More precisely, a central thesis of this study is that to hold a moral belief is to be disposed to be in a desire-like, noncognitive state where this state has some specific structure.

This dissertation arrives at and justifies this thesis by making a number of steps. It starts with preparing the ground by developing a general theory of what it is for a sentence to *express* a mental state. Roughly, this theory states that for a sentence to express a mental state is for that sentence to be conventionally used to perform a speech act which counts as sincere only if the user or speaker is in that mental state. This theory of expressing satisfies all the constraints that must be satisfied for it to be compatible with the purposes of metaethical expressivism.

The next step is to provide a solution to the most notorious semantic problem confronting metaethical expressivism, the problem of explaining how sentences containing moral words can have compositional meaning and stand in logical relations also known as the *Frege-Geach* or embedding problem. Building on the research of Mark Schroeder (2008a), a new expressivist semantics, called 'Attitude Semantics', will be developed for a first-order

language. The motivating idea underlying attitude semantics is that expressivists can avoid the problematic feature of so-called 'biforcation' if they treat the structures of the mental states expressed by moral sentences as well as that expressed by non-moral sentences as being isomorphic. In order to give those structures a philosophically interesting interpretation, it will be suggested that we treat sentences as expressing dispositional states. As a toy example we will treat the sentence 'murder is wrong' as expressing the disposition to be for blaming for murder.

After dealing with the Frege-Geach problem, I turn to the metaethical expressivists' project of quasi-realism. Two major goals of this project are to explain how moral sentence can be *truth-apt* and how the states expressed by them can be regarded as *moral beliefs*. In order to establish those goals I critically examine a number of theories of truth-aptness by providing counterexamples to them. I then present my own theory which says that a sentence is truth-apt iff and because there is a conventional way to use it to express a belief. What is new about this theory is that it treats expressing a belief as being explanatorily more basic than being truth-apt. Belief, according to this theory, explains truth-aptness, not vice versa. This has the important consequence that it shows that what I call the 'orthodox strategy' of quasi-realists is mistaken and needs to be inverted: expressivists must 'earn the right' to moral belief before they can establish that moral sentences are truth-apt, and not the other way round.

The second part of this dissertation is then concerned with the question of what it is to hold a moral belief and how they can figure in rational reasoning. In short: it is concerned with earning the right to moral belief. After discussing the major problems confronting a noncognitivist theory of belief, most importantly what I call the 'tightrope problem', I argue for a dispositional form of noncognitivism about moral belief that combines elements from Eric Schwitzgebel's dispositionalism about belief and Sebastian Köhler's 'conceptual role expressivism'.[1] This dispositional noncognitivism perfectly matches and justifies the dispositionalist interpretation of the expressivist semantics developed before.

In order to fully earn the right to moral belief, however, metaethical expressivists must also deal with several epistemological objections concerning the rationality of moral belief. The first objection I deal with comes from Derek Baker (2018) and is directed against Mark Schroeder's specific development of expressivism. I deal with this objection because my own theory bears close resemblance to Schroeder's theory and so it might be worried that it confronts the same issues. Fortunately, we will see that Baker's objection rest on a number of mistaken assumptions and so neither applies to Schroeder's expressivism nor to the version which I defend in this study.

---

[1]See (Schwitzgebel, 2002), (Schwitzgebel, 2013), and (Köhler, 2017).

The other epistemological objection to metaethical expressivism I deal with is Cian Dorr's famous *wishful-thinking objection* from (2002). In contrast to previous replies to this objection I attack it by directly questioning the central assumption on which it rests, namely that reasoners can never be justified in basing a factual belief on a desire-like state. I argue that this is false, and that under certain conditions reasoning from desire-like mental states to factual beliefs is justified. My argument shows that even if holding a moral belief is being in a desire-like noncognitive state, basing factual belief on a moral belief can give one an epistemic reason for the factual belief. Hence, even if moral thoughts are wishes, moral thinking is not wishful thinking.

I take my arguments to earn the dispositional expressivist the right to moral belief. By the previously defended theory of truth-aptness in terms of belief, this allows the dispositional expressivists to conclude that moral sentences are also truth-apt. This finishes the quasi-realist project of this dissertation.

In the last part of this dissertation I switch gears and present two unnoted problems for metaethical expressivists having to do with explaining moral motivation. Both arguments are surprisingly simple. The first one is this: if motivation requires a belief as well as a desire (Humeanism), and moral judgments motivate by themselves (internalism), then moral judgments cannot be desires only. If this argument is sound, it shows that the famous 'motivation argument' not only entails the falsity of cognitivism, but also the falsity of noncognitivism. The second argument is the following: moral judgments when combined with suitable external desires motivate to act (externalism), but if motivation requires belief as well as desire (Humeanism), then moral judgments cannot be desires only. I discuss several ways in which noncognitivists might want to deal with those problems, but leave it as an open question whether those solutions will ultimately be viable.

Overall, this dissertation makes serious contributions to the research program of metaethical expressivism. It makes proposals for how to solve a number of notorious problems, such as the Frege-Geach problem in semantics and the wishful-thinking problem in epistemology. Besides this, however, it also raises two new problems for metaethical expressivism in the philosophy of action. For those reasons this dissertation will be of interest for anyone who wants to defend or criticize expressivism in metaethics, and more generally, anyone who wants to better understand our astonishing ability to express moral beliefs.

Zusammenfassung

Es ist bemerkenswert, dass wir Menschen in der Lage sind, moralische Überzeugungen zu haben, gemäß ihnen zu handeln und sie zum Ausdruck zu bringen. Diese Dissertation zielt darauf ab, ein besseres philosophisches Verständnis davon zu erhalten, warum und wie diese Dinge möglich sind, insbesondere wenn wir annehmen, der metaethische Expressivismus sei wahr. Der metaethische Expressivismus ist die Kombination des Expressivismus mit dem Noncognitivismus. Expressivismus wird dabei verstanden als die Auffassung, dass die Bedeutung eines Satzes erklärt wird durch den mentalen Zustand, welcher durch seine konventionelle Verwendung ausgedrückt wird. Der Nonkognitivismus ist die Auffassung, dass die mentalen Zustände, welche durch moralische Sätze ausgedrückt werden nonkognitive, wunschartige Zustände sind.

Die zentrale Frage dieser Dissertation ist:

> *Wie können metaethische Expressivistys[2] erklären, dass wir in der Lage sind moralische Überzeugungen auszudrücken?*

Die kürzestes Antwort, die diese Dissertation geben wird, lautet:

> indem sie den *dispositionalen Expressivismus* akzeptieren.

Das dispositionale Expressivisty denkt, dass moralische Überzeugungen Dispositionen sind, sich in bestimmten mentalen Zuständen zu befinden, welche Wünschen in gewisser Weise ähneln. Genauer gesagt ist die zentrale These dieser Arbeit, dass eine moralische Überzeugung zu haben dasselbe ist, wie disponiert zu sein, in einem bestimmten wunschartigen, nonkognitiven Zustand zu sein, wobei dieser eine ganz bestimmte Struktur haben muss.

Diese Dissertation begründet diese These in einer Reihe von Schritten. Im ersten Teil werden zunächst die Grundsteine gelegt indem eine allgemeine Theorie darüber entwickelt wird, was es bedeutet, dass ein Satz einen mentalen Zustand *ausdrückt* ('express'). Diese Theorie besagt grob, dass ein Satz

---

[2]Ich entgendere in dieser deutschen Zusammenfassung gemäß der eleganten Methode des österreichischen Aktionskünstlers Hermes Phettberg, d. h. ich verwende das Genus neutrum, schreibe statt bspw. '*innen' ein 'y', und für den Plural hänge ich ein 's' an.

einen mentalen Zustand ausdrückt dann und nur dann, wenn (d. h. genau dann wenn; kurz: 'gdw') dieser Satz konventioneller Weise dazu verwendet wird, um einen Sprechakt auszuführen, welcher nur dann als aufrichtig gilt, wenn sein Verwendy oder Sprechy sich in genanntem mentalen Zustand befindet. Diese Theorie des Ausdrückens erfüllt alle Anforderungen, um einen Begriff des Ausdrückens zu liefern, der für die Zwecke des metaethischen Expressivismus geeignet ist.

Der nächste wichtige Schritt ist die Erarbeitung einer Lösung des hartnäckigsten semantischen Problems des metaethischen Expressivismus, das Problem nämlich, zu erklären, wie Sätze, die moralische Ausdrücke enthalten eine kompositionale Bedeutung haben und in logischen Beziehungen zueinander stehen können. Das ist das so genannte *Frege-Geach-* oder auch Einbettungs-Problem. Aufbauend auf den Arbeiten von Mark Schroeder (2008a), wird eine neue expressivistische Semantik, genannt 'Attitude Semantics', für eine Sprache erster Stufe entwickelt. Die grundlegende Idee hierbei ist, dass Expressivistys, die als problematisch erachtete so genannte 'Biforcation' vermeiden können, wenn sie die Struktur jener mentalen Zustände, die durch moralische Sätze ausgedrückt werden und derer, die durch deskriptive Sätze ausgedrückt werden, als isomorph behandeln. Um diesen Strukturen eine philosophisch interessante Interpretation zu geben, wird vorgeschlagen, die ausgedrückten Zustände als dispositionale Zustände aufzufassen. Als eine erste, grobe Annäherung wird der durch den Satz 'Morden ist falsch' ausgedrückte Zustand verstanden als die Disposition es zu befürworten, dass Morden geächtet wird ('disposition to be for blaming for murder').

Im Anschluss an das Frege-Geach-Problem, wende ich mich dem expressivistischen Projekt des so genannten Quasi-Realismus zu. Zentrale Ziele dieses Projektes sind es zum einen zu erklären, wie moralische Sätze wahrheitsfähig ('truth-apt') sein können, und zum anderen, warum die durch solche Sätze ausgedrückten Zustände als moralische Überzeugungen angesehen werden können. Um diese beiden Ziele zu verwirklichen setze ich mich kritisch mit einer Reihe von Theorien der Wahrheitfähigkeit auseinander und liefere einige Gegenbeispiele gegen diese. Darauf aufbauend entwickle ich eine eigene Theorie, welche im Wesentlichen besagt, dass ein Satz wahrheitsfähig ist gdw er konventioneller Weise dazu verwendet wird um eine Behauptung zu machen, d. h. um eine Überzeugung auszudrücken. Diese Theorie erachtet das Ausdrücken einer Überzeugung für explanatorisch elementarer an als das wahrheitsfähigsein. Gemäß dieser Theorie erklären Überzeugungen Wahrheitsfähigkeit und nicht umgekehrt, wie häufig von so genannten Überzeugungsminimalistys propagiert. Eine wichtige Konsequenz daraus ist, dass sich, was ich die 'orthodoxe Strategie' des Quasi-Realismus nenne, als Irrweg entpuppt und umgekehrt werden muss: Expressivistys müssen sich das Recht von moralischen Überzeugungen sprechen zu können verdienen *bevor* sie behaupten können, dass moralische Sätze wahrheitsfähig sind, und nicht, wie allgemein angenommen, anders herum.

Der zweite Teil der Dissertation beschäftigt sich mit der Frage, was es bedeutet moralische Überzeugungen zu haben und wie diese in rationalem Schlussfolgern eine Rolle spielen können. Kurz gesagt: es geht darum wie Expressivistys sich das Recht von moralischen Überzeugungen sprechen zu dürfen verdienen können. Nachdem ich die Hauptprobleme nonkognitivistischer Theorien der Überzeugungen diskutiert habe, allen voran was ich das 'tightrope problem' nenne, argumentiere ich für eine dispositionale Form des Nonkognitivismus über moralische Überzeugungen, welche Teile von Eric Schwitzgebels Überzeugungsdispositionalismus mit Teilen von Sebastian Köhlers 'conceptual role expressivism' verbindet.[3] Dieser dispositionale Nonkognitivismus ist das perfekte Gegenstück zu der im ersten Teil dieser Arbeit vorgeschlagenen dispositionalen Interpretation der expressivistischen Semantik und rechtfertigt damit den gemachten Vorschlag.

Um sich jedoch völlig die Berechtigung zu moralischen Überzeugungen zu verdienen, müssen metaethische Expressivistys darüber hinaus zudem Erwiderungen auf einige epistemologische Einwände liefern, welche die Rationalität moralischer Überzeugungen betreffen. Der erste Einwand, den ich behandle, stammt von Derek Baker (2018) und richtet sich gegen Mark Schroeders spezifische Ausarbeitung des Expressivismus. Ich behandle diesen Einwand, da Schroeders Theorie starke Ähnlichkeiten mit dem in dieser Disseration entwickelten dispositionalen Expressivismus aufweist und man deshalb befürchten könnte, dieser Einwand, sofern er sich als stichhaltig erweisen würde, fände auch gegen letztere Position Anwendung. Glücklicherweise werden wir jedoch sehen, dass Bakers Einwand auf einer Reihe falscher Annahmen beruht und deshalb weder gegen Schroeders Theorie noch gegen den dispositionalen Expressivismus vorgebracht werden kann.

Ein anderer epistemologischer Einwand gegen den metaethischen Expressivismus mit welchem ich mich beschäftige werde, ist Cian Dorrs bemerkenswerter Wunschdenken-Einwand aus (Dorr, 2002). Anders als frühere Erwiderungen greife ich diesen Einwand an, indem ich die zentrale Annahme in Frage stelle, auf welcher dieser Einwand beruht. Diese Annahme besagt, dass es niemals gerechtfertigt sei, eine faktische Überzeugung auf einen wunschartigen Zustand zu gründen. Etwas auf Basis von Wünschen zu glauben sei immer Wunschdenken. Ich argumentiere, dass diese zunächst äußerst plausibel klingende Annahme falsch ist, und zeige, dass unter bestimmten Bedingungen das Schlussfolgern von einem wunschartigen Zustand zu faktischen Überzeugungen epistemisch gerechtfertigt ist. Mein Argument zeigt, dass selbst wenn moralische Überzeugungen wunschartige nonkognitive Zustände sind, wie der Nonkognitivismus behauptet, es dennoch möglich ist, dass das Schlussfolgern von wunschartigen Zuständen zu faktischen Überzeugungen uns Gründe für faktische Überzeugungen liefert. Daraus folgt, dass selbst wenn moralische Überzeugungen wunschartige Zustände sind, Schlussfolgern

<hr>

[3]Siehe (Schwitzgebel, 2002), (Schwitzgebel, 2013), und (Köhler, 2017).

mit moralischen Überzeugungen kein Wunschdenken ist.

Diese Argumente zeigen, dass das dispositionale Expressivisty berechtigt ist von moralischen Überzeugungen zu sprechen und realisiert damit das erste der beiden zentralen Ziele des Quasi-Realismus. Mit Hilfe der zuvor verteidigten Theorie der Wahrheitsfähigkeit, welche Wahrheitsfähigkeit durch die Ausdrückbarkeit von Überzeugungen definiert, kann das dispositionale Expressivisty schlussfolgern, dass moralische Sätze wahrheitswertfähig sind. Damit ist auch das zweite zentrale Ziel realisiert und somit ist das quasi-realistische Projekt dieser Dissertation abgeschlossen.

Im letzten Teil der Dissertation ändert sich die Stoßrichtung. Statt weiterhin den metaethischen Expressivismus stark zu machen und Lösungen für dessen Probleme zu erarbeiten, stelle ich zwei bisher unbemerkte Probleme des metaethischen Expressivismus vor. Beide Probleme betreffen das Thema der moralischen Motivation, das heißt konkreter, die Frage, wie moralische Überzeugungen uns dazu bewegen bestimmte Dinge zu tun. Beide Probleme lassen sich überraschend einfach formulieren. Das erste Problem ist folgendes: Wenn Motivation das Vorhandensein sowohl einer Überzeugung als auch eines Wunsches erfordert (Humeanismus) und darüber hinaus moralische Urteile von sich aus zum Handeln motivieren (Internalismus), dann können moralische Urteile nicht ausschließlich wunschartige Zustände sein. Wenn dieses Argument korrekt ist, dann zeigt das nichts weniger, als dass das berühmte 'Motivationsargument', welches oft als ein Hauptgrund für den Nonkognitivismus angeführt wird, nicht nur die Falschheit des Kognitivismus impliziert, wie allgemein angenommen, sondern ebenso auch die Falschheit des Nonkognitivismus. Das zweite Problem ist folgendes: wenn moralische Urteile mit geeigneten Wünschen kombiniert werden, motiviert uns das plausiblerweise dazu, etwas zu tun. Wenn jedoch Motivation sowohl Überzeugungen als auch Wünsche erfordert (Humeanismus), dann können moralische Urteile nicht ausschließlich wunschartige Zustände sein. Ich diskutiere zwar eine Reihe von Möglichkeiten wie metaethische Expressivistys, und insbesondere dispositionale Expressivistys, versuchen könnten, diese beiden Probleme zu lösen. Letztendlich lasse ich jedoch die Frage offen, wie erfolgsversprechend diese Lösungsversuche tatsächlich sind.

Im Großen und Ganzen liefert diese Dissertation somit eine Reihe wichtiger Beiträge zum Forschungsprogramm des metaethischen Expressivismus. So macht sie beispielsweise Vorschläge zur Lösung einiger äußerst hartnäckiger Probleme, wie des Frege-Geach-Problems in der Semantik oder des Wunschdenken-Problems in der Erkenntnistheorie. Darüber hinaus dürften die Theorien des Ausdrückens und der Wahrheitsfähigkeit auch außerhalb metaethischer Diskussionen von philosophischem Interesse sein. Neben diesen positiven Beiträgen, weist diese Arbeit jedoch auch auf zwei bislang unbemerkte Probleme des metaethischen Expressivismus in der Handlungstheorie hin. Aus all diesen Gründen dürfte die vorgelegte Dissertation für all jene von Interesse sein, die den Expressivismus in der Metaethik entweder verteidigen

oder angreifen möchten, und darüber hinaus für alle, die unsere erstaunliche Fähigkeit moralische Überzeugungen zu haben, gemäß ihnen zu handeln und sie auszudrücken, besser verstehen möchten.

Contents

Contents (detailed)

Preface

I must have been in the fifth or sixth grade, so roughly at age of eleven or twelve, when we did a school excursion to the concentration camp in Dachau. There we were shown, quite unsuspectedly, a documentary movie in which I saw for the first time what the Nazis had done: I saw huge piles of naked emaciated dead bodies, the bodies of men and women of all ages and even little children. I saw close-ups of their bulging eyes and their open mouths. Then we were led into a room with clinkers on the floor and walls and what looked like sprinklers at the ceiling. It was explained to us that this room, though the letters above its door said "Brausebad", was a gas chamber and what this room was actually used for.

This experience had a long lasting, almost traumatic, effect on my young soul and caused me many nightmares even years later. After this excursion I immediately knew and felt one thing for sure: that what the Nazis had done there was abhorrently wrong, and that we must do everything to prevent that something like this can happen again.

A couple of years later, around the age of 16, my best friend Michael handed me an old copy of Friedrich Nietzsche's *Also sprach Zarathustra.* I never had a great interest in books, but at this teen age I was of course fascinated by how radical this 'free spirit' questioned that anything is ever good or evil—at least I took it that he questioned this. At the same time Nietzsche's nihilistic thoughts really shattered me: How could someone think that nothing is ever wrong and so that what had happened in this chamber in Dachau was not wrong? In the following months and years I read through the complete works of Nietzsche, some of his correspondence, and also some biographies about him. I even visited his home in Naumburg and earthened my own Nietzsche bust which I still have on my desk. I also read the works of philosophers who had a great influence on him such as Arthur Schopenhauer, Friedrich Albert Lange, and Paul Rée amongst others. Though I was not officially enrolled at the university, I visited a seminar on Nietzsche, and even gave a presentation on his moral philosophy. I also visited a Nietzsche seminar at the Volkshochschule, and I regularly went to talks at the Nietzsche-Forum in Munich. All that led to an ever growing interest in the foundations of morality.

It took me several years to realize that this interest was so strong that I had to study philosophy at the university. So I quit my job as an optician, and told my wife that I want to go back to school for two years to get the degree necessary to finally enroll to philosophy in 2008.

Now it is 2021 and I finally write this preface to my dissertation on expressing moral belief. Neither the Nazis nor Nietzsche play a visible role in this dissertation, and I have not read a single line of his works since I started doing academic philosophy. Nevertheless this dissertation can be seen as an attempt to reconcile Nietzsche's anti-metaphysical thoughts about the true nature and origins of morality with our everyday experience that we have sometimes very strong beliefs about the rightness or wrongness of things, such as about what happened in the 1940s in Dachau and all the other good and bad things that happen every day on this planet. This dissertation is my attempt to better understand how we can have beliefs about what is right and wrong, how we can take those beliefs to be true or false, what we do when we speak about moral matters, if there is only the facts of nature and no god telling us what to do.

I would be lying if I said that writing this dissertation was easy and always fun. Making new philosophical arguments and bringing them together into a larger whole was hard work, and I was frustrated more than once. Life also can make philosophizing feel like a completely idle endeavor. Right when I started the work on this dissertation in September 2016, our twin daughters were born three month too early after a very difficult pregnancy. Just a couple of month later in May 2017 my father in law died unexpectedly at the age of 63 and left his wife as well as his mother, both in need of care, 400 kilometer away from us. This was a endurance test for our family with three young girls. Then, after a long struggle against her lung disease, my mother in law died in February 2020 at the age of 60. After that came the global pandemic. Meanwhile the dementia of my wife's beloved grandmother had progressed so far that we had to hand her over to a nursing home. There she got infected with the corona-virus and died around Christmas. On top of all of this, the situation in the last couple of month of writing on the dissertation at home, with school and kindergarten closed, and my wife and me alternating in home-office, was also far from being ideal for doing philosophical work.

Nevertheless, having to deal with those strokes of fate and difficulties made me all the more thankful for having a healthy and happy family. I also feel very privileged to have been able and even get paid for trying to better understand what is involved in and how it is possible that we express moral beliefs. I hope that some of the things I say in this dissertation will be found illuminating by others too, and that the contributions it makes to the research program of metaethical expressivism will stimulate further investigations into the field.

Acknowledgments

Chapter 1

Introduction

1.1 Expressivism in Ethics

When someone publicly utters a meaningful sentence like 'murder is wrong', we take that person to *express a moral belief.* More precisely, if someone utters 'murder is wrong', we take that person to assert that murder is wrong—something that can be true or false. Moreover, if we regard the person's assertion to be sincere, we take her to believe that murder is wrong, and we take her to be moved to avoid murdering.

This dissertation aims to answer the question of how we can explain these familiar observations if we assume metaethical expressivism, that is:

> *How can moral sentences express moral beliefs if ultimately they express mental states that are rather like desires?*

In order to answer this question I develop and defend a form of metaethical expressivism. Metaethical expressivism is the combination of expressivism and noncognitivism about ethics. Expressivism, as I shall understand it in this thesis, is a general explanatory claim about meaning:

Expressivism is the view that the meaning of a sentence is explained by the mental state it is conventionally used to express.[1]

I will have to say much more about what expressivism involves in chapter 2 and 3 below. In metaethics, expressivism is traditionally combined with noncognitivism. Noncognitivism is the following view about the nature of moral judgment, where by 'moral judgment' I mean 'any mental state, whatever it is, that is expressed by the conventional use of a sentence containing a moral word':

Noncognitivism is the view that moral judgments are noncognitive, desire-like mental states.

---

[1]Cf. (Gibbard, 2003, 7).

By combining expressivism with noncognitivism we get metaethical expressivism by which I shall mean the following:

Metaethical expressivism: Moral sentences are conventionally used to express noncognitive, desire-like states.

Metaethical expressivism (or short: expressivism[2]) is traditionally regarded as having a number of attractions such as being ontologically parsimonious, being compatible with a naturalistic view of the world, being capable of explaining the vehemency of ethical disagreement, and explaining how moral judgments motivate us to act, to name just a few of its major attractions.

Unfortunately, metaethical expressivism not only has attractions, but also raises a number of questions. One of the more simple ones is the following:

(1) *Expressing*: what it is to express a mental state?

More difficult questions are the following:

(2) *Moral Meaning*: if metaethical expressivism is true and so moral sentences express desire-like states, how can the meaning of sentences be compositional and how can they stand in logical relations to each other?

Answering this question amounts to solving the famous Frege-Geach problem. Another important question is this:

(3) *Moral Truth-Aptness*: how can moral sentences be true or false, when their function is to express desire-like states which themselves are not evaluable in terms of truth and falsity?

(4) *Moral Belief*: how can we legitimately claim to have moral beliefs, when the states expressed by moral sentences are desire-like, and beliefs and desires are categorically distinct?

This thesis provides detailed answers to all these questions. Starting with the first question, I provide a theory of expressing that is compatible with the purposes of metaethical expressivism. After presenting and criticizing several theories of expressing, I make the suggestion to understand expressing in terms of sincerity conditions for speech acts. Roughly, the suggestion is that for a sentence to express a mental state is for that sentence to be conventionally used to perform a certain speech act where this speech act is

---

[2]In this thesis I shall often use 'expressivism' instead of 'metaethical expressivism', and the context will make clear when I mean the more general semantic claim instead of the combination of expressivism and noncognitivism.

sincere only if one is in that specific mental state. This investigation of the notion of expression lays the ground for my answer to the second question.

In order to answer the second question (i. e. solve the Frege-Geach problem) I develop a compositional expressivist semantics which I call 'Attitude Semantics'. The motivating idea underlying the semantic proposal which I make in this thesis is that we can avoid many of the problems that confront a closely related form of expressivism which Mark Schroeder (M. Schroeder, 2008a) has previously developed, if we get rid of a structural asymmetry that lies at the heart of his so-called 'Biforcated Attitude Semantics'. In order to give my own structural solution to the Frege-Geach problem a plausible interpretation, I suggest that expressivists embrace what I call 'dispositional expressivism'. The rough idea of dispositional expressivism is that sentences express sets of dispositions to be in certain mental states. For instance, I propose that 'murder is wrong' expresses the disposition to be for blaming for murder. The dispositionalist proposal, as we shall see, has many philosophical attractions besides offering a plausible interpretation of attitude semantics. For instance, it allows expressivists to draw a clear distinction between the state expressed by descriptive sentences and moral sentences, it connects with the tradition of semantic dispositionalism, and revives some old ideas which have already been defended by one of the most influential figures of the emotivist tradition in metaethics, Charles L. Stevenson.[3] Most importantly, however, the dispositionalist interpretation of expressivism perfectly dovetails with a dispositionalist theory of belief that is compatible with the metaethical expressivist's noncognitivism.

In order to answer the third question, this thesis criticizes a number of theories of truth-aptness and on that basis proposes a different one. Roughly, the proposed theory says that something is apt for being true or false just in case, and in virtue of the fact that, it is conventionally used to make assertions or equivalently, it is used to express a belief. An important consequence of this theory of truth-aptness is that metaethical expressivists need to invert what I call the 'orthodox strategy' of explaining what it is to have a moral belief. The orthodox strategy is to explain moral belief in terms of the truth-aptness of moral sentences. The new theory of truth-aptness precludes this strategy: we cannot explain belief in terms of truth-aptness, if we already explain truth-aptness in terms of belief. For the metaethical expressivists this means that they now need to establish that moral sentences can be regarded as expressing beliefs before they can conclude that moral sentences are truth-apt.

This directly leads to the fourth question about how the expressivists can explain that the noncognitive states expressed by moral sentences can legitimately be regarded as moral beliefs. I suggest that in order to answer this question metaethical expressivists should borrow ideas from a popular

---

[3]His most influential work on emotivism is (Stevenson, 1945).

position in the philosophy of belief according to which beliefs must be characterized in terms of their specific functional roles. Those functional roles, I argue, should be spelled out in terms of dispositional properties. Ultimately, I propose that the states expressed by moral sentences are certain sets of dispositions of agents to be in certain noncognitive states, and that to have such a set of dispositions is precisely what it is to hold a moral belief. This form of noncognitivism, which I call 'dispositional noncognitivism', nicely complements with the before suggested dispositional form of expressivism.

A further important question which metaethical expressivists have to answer, besides the ones mentioned above, is this one:

(5) *Moral Reasoning*: how can moral beliefs figure in rational reasoning, if they are ultimately noncognitive, desire-like mental states?

In order to answer this question I provide detailed replies to two epistemological objections that have been leveled against metaethical expressivism. The first objection comes from Derek Baker (Baker, 2018). His objection fails, I argue, because it rests on several mistaken assumptions about what are and are not genuine requirements of rationality. The second objection is Cian Dorr's famous wishful-thinking objection (Dorr, 2002). Against Dorr I argue that under certain conditions rational thinkers are justified to believe something on the basis of a desire. Moral reasoning, according to metaethical expressivism, satisfies those conditions, and so is not wishful thinking.

Despite all the positive contributions this study makes to metaethical expressivism, this dissertation also raises two problems that have so far been unnoticed in the literature. These problems concern a further question which is the following:

(6) *Moral Motivation*: how can moral beliefs motivate agents to act?

That we find problems with answering this question is rather surprising, since one of the major attractions for becoming a metaethical expressivist, as I have noted above, is that it is generally believed that its noncognitivism is particularly good at explaining moral motivation. Even though I sketch some possible ways for how noncognitivists might be able to solve these problems, I leave it open how viable these solutions are ultimately.

So, in very broad strokes, the grand argument of this thesis can be sketched as follows. Metaethical expressivists can solve their problem with compositionality and logicality (the Frege-Geach problem) if they assume that sentences express mental states with a certain structure and nature. This structure is best interpreted in terms of dispositions which yields dispositional expressivism. In particular, moral sentences express dispositions to be in certain noncognitive states, whereas non-moral sentences express dispositions to be in certain cognitive states. Moreover, the dispositional states expressed

by moral sentences can be regarded as moral beliefs, because they have the marks that dispositionalists about belief claim beliefs to have. Finally, since every sentence that expresses a belief is truth-apt, and the dispositional states expressed by moral sentences can be regarded as belief, this allows metaethical expressivists to conclude that moral sentences are truth-apt. This solves the metaethical expressivists' problems with meaning, belief, and truth-aptness. On top of this, this dissertation shows how moral beliefs can figure as premises in pieces of rational reasoning. Finally, however, this dissertation also raises two new problems for explaining how moral beliefs motivate us to act.

I conclude that even though this dissertation makes some serious advances for metaethical expressivism, and I am generally sympathetic to this research program, I am also only cautiously optimistic about its prospects. Metaethical expressivism has many things to offer for the naturalistically inclined moral philosopher, yet it is not clear that the offerings are worth the theoretical price and the motivation problems raised demand further investigation. For this reason, the dissertation will be of interest to anyone who wants to defend, but also for those who want to criticize an expressivist view about expressing moral belief.

1.2   The Plan of this Study

To give the reader a slightly more detailed description of what is to come in the following chapters, here is an overview of the structure of this thesis.

The thesis is about expressing moral belief, and also about how it motivates action. Accordingly, it is divided into three major parts: the first is about *expressivism*, the second about *noncognitivism*, and the final one about *internalism and Humeanism*.

Part I is about language, and moral language in particular, and contains three chapters. In the *first chapter* (chapter 2) I lay the groundwork by critically discussing the central semantic relation of expressivism: the relation of a sentence's *expressing* a mental state. Although 'expressing' is the name-giving relation of expressivism, expressivists have said surprisingly little about it in the literature. This chapter fill this lacuna. I start with formulating a number of constraints of adequacy on the expressing relation. After showing that several natural proposals fail to satisfy the constraints, I present an explication of the expression relation which is plausible and compatible with the broader commitments of expressivism. Roughly, the proposal is that a sentence expresses a mental state iff that mental state is the sincerity condition of the speech act this sentence is conventionally used to perform. For instance, 'grass is green' expresses the belief that grass is green, because that sentence is used to assert that grass is green, and this assertion is a sincere speech act only if the speaker believes that grass is green.

Equipped with a firm grasp of what it is for a sentence to express a mental state, I turn to expressivism in *chapter 3*. The heart of this chapter is a new solution to the most notorious problem for expressivist semantics: the Frege-Geach problem. After making some general remarks about the semantic program of expressivism and how it might account for compositionality and logicality in terms of functions from mental states to mental states and the rational relations between them, I discuss why such an attitude-functional semantics faces the Frege-Geach problem. After considering its most elementary instance, namely Nicholas Unwin's 'negation problem' (Unwin, 1999), I present Mark Schroeder's celebrated solution, his so-called 'biforcated attitude semantics', developed in (M. Schroeder, 2008a), according to which all sentences, moral and non-moral alike, express pairs of states of the attitude of *being for*. Building on Schroeder's work, I make a proposal for avoiding some of the problems that plague his theory, which results in a new semantics which I call 'attitude semantics'. The proposal is relatively simple: rather than saying that 'murder is wrong' expresses the state of being for blaming for murder *twice* (in order to be a pair) as in Schroeder's theory, expressivists should say that it expresses a state such as the state of *being for being for* blaming for murder, or something structurally similar. The major task is then to show how this idea can be transformed into a formally adequate expressivist semantics. After this task is fulfilled, I suggest a more plausible interpretation of the structure of the states expressed in terms of dispositions, rather than in terms of states of being for, which results in what I call 'dispositional expressivism'.

In the subsequent chapter (*chapter 4*) I turn to the expressivists' project of quasi-realism which involves showing how metaethical expressivism can be compatible with the claim that moral sentences express beliefs and are truth-apt. I explain that the 'orthodox strategy' of quasi-realism is to argue from a minimalist theory of truth-aptness to a minimalist theory of belief. I discuss and reject a number of different theories of truth-aptness on the grounds that they are subject to counterexamples. Finally, I present my own theory of truth-aptness, which says, roughly, that a sentence is truth-apt if and only if, and because it expresses a belief. As a consequence of this theory of truth-aptness, it turns out that the orthodox strategy must be inverted, because we cannot explain belief in terms of truth-aptness, if we explain truth-aptness in terms of belief. Instead, it emerges that the quasi-realists should make what I call the 'belief-to-truth-aptness argument'. Making this argument requires a new strategy. Quasi-realist expressivists now first have to establish that the states expressed by moral sentences merit to be regarded as beliefs despite the fact that they are ultimately noncognitive states. Only after this is shown can they conclude that moral sentences are truth-apt.

This leads to *Part II* of the thesis, which is about noncognitivism about moral belief. *Chapter 5* starts with making more precise the view of noncognitivism and introduces some problems for it, in particular what I call the

'tightrope problem'. This is the problem of finding the right kind of balance between the noncognitivist quasi-realist claim that moral judgments merit to be regarded as beliefs despite their being desire-like states on the one hand, and the Humean claim that beliefs and desires are 'distinct existences' on the other. After giving a brief overview of different theories of belief, I introduce Schwitzgebel's 'phenomenal, dispositional account of belief' (Schwitzgebel, 2002) and compare it to Köhler's more functionalist 'conceptual role expressivism' (Köhler, 2017). This leads to a view I call 'dispositional noncognitivism' which states that beliefs are set of dispositional properties some of which are essential while some of which are only typically for a mental state to be regarded as a belief. I argue that although there remain some issues for dispositional noncognitivism, it also promises to solve some of the traditional problems, and most importantly the tightrope problem.

The next two chapters deal with objections against metaethical expressivism concerning the epistemology of moral belief. In *chapter 6* I deal with an objection put forward recently by Derek Baker (Baker, 2018). This chapter is a slightly modified version of an article that is which I previously published in Erkenntnis (Hengst, 2020). Baker argues that a noncognitivist theory of moral belief like Schroeder's, upon which mine is closely modeled, is unable to explain why certain beliefs are rationally incoherent, and he concludes that thus the noncognitivist theory must be false. I reply to Baker's objection by showing that Baker's argument for the irrationality of those beliefs rests on mistaken assumptions and is thus unsound. I take this discussion to show that it is also highly unlikely that a proposal like mine will fall prey to something like Baker's objection.

Chapter 7 also deals with another objection coming from epistemology, namely Cian Dorr's famous *wishful thinking objection* (Dorr, 2002). This objection is arguably the most difficult objection to metaethical expressivism besides the Frege-Geach objection. Dorr argues that if moral beliefs are desire-like states (noncognitivism), then what looks like perfectly fine reasoning from moral beliefs to factual beliefs, is actually wishful thinking, because one can never be justified in basing a belief on a desire-like state. I argue that this is false. I argue that, quite surprisingly, under certain conditions one can be epistemically justified in basing a belief on a desire-like state. In paradigmatic cases of wishful thinking those conditions are not satisfied, but, according to noncognitivism, in cases of moral reasoning they are satisfied.

Up to this stage of the thesis I have tried my best to defend expressivism and noncognitivism in ethics. This changes in the *final part*. In *chapter 8* I raise two simple, unnoticed, yet fundamental problems for noncognitivism. Both concern its alleged ability to explain moral motivation. The first problem is that the famous *motivation argument*, which is cited as one of the main reasons against cognitivism and for noncognitivism, in fact also entails the falsity of noncognitivism. The second problem is that, intuitively,

moral beliefs, just like non-moral beliefs, when combined with a suitable desire, motivate to act. If noncognitivism is true, and moral beliefs are noncognitive states, however, then this is incompatible with the Humean claim that motivation requires both a noncognitive (desire) and a cognitive state (belief). I suggest several ways out of these problems, but leave it as an open question whether these solutions are ultimately convincing. This concludes my case for expressivism and noncognitivism in ethics.

The *final chapter* summarizes the main claims and findings of this dissertation, and draws a critical conclusion.

Part I

Moral Talk: Expressivism

Chapter 2

On Expressing

This thesis is about expressing moral belief. More specifically, it is about metaethical expressivism. Metaethical expressivism, I said in the introduction, is a combination of two views: expressivism in the philosophy of language, and noncognitivism in philosophical psychology.[1] This first part of the thesis is about expressivism.

Expressivism, as I understand it here, is the view that the meaning of a sentence has to be explained by the mental state it is conventionally used to express. This is only a rough description of the view, of course, and in this and the next chapter our understanding of it will become much sharper. I start this dissertation on expressivism with explaining what kind of theory expressivism is and then investigate the central and name-giving notion of expressivism—the notion of *expressing*. This will answer the above first question of what it is to express a mental state, and so lay the foundation for what is to come.

2.1 What Kind of Theory is Expressivism?

Before we dive into the investigation of what it is to express something, I want to make some general remarks about what kind of theory expressivism is.

Expressivism, I said in the introduction, is the view that the meaning of a sentence has to be explained by the mental state it is conventionally used to express. So, expressivism tells us what moral sentences are *used* for. This suggests that expressivism is concerned with what we *do* with sentences, that it concerns *pragmatics*. Expressivism, however, also attempts to tell us

---

[1]It is not clear in which philosophical discipline noncognitivism fits best. I think it is best described as a view in philosophical psychology if we think of this field of research as being concerned with the metaphysical nature of mental states. Since noncognitivism is a specific view about the nature of the state of belief it might also be appropriate to think of it as belonging to the field of epistemology.

something about what moral sentences mean, and is thus concerned with *semantics*. How does this fit together?

One idea is to link it to the influential view in the philosophy of language that 'meaning is use'. This view is, of course, most prominently associated with the later Wittgenstein (Wittgenstein, 1953). Versions of this idea have prominently been defended for instance in (Brandom, 1994) and (Horwich, 1998). So, expressivism might be viewed as a variant of a use theory of meaning. Unfortunately, the slogan that meaning is use is not very precise, and it is also not clear whether use theories of meaning, and theories of meaning more generally, are best understood as being semantics. This is because, a semantics, on a more narrow understanding, is a theory that systematically assigns semantic values or meanings to expressions of a language. But use theorists of meaning rarely provide a semantics in this narrow sense, and assigned uses to linguistic expressions. So, even if we think of expressivism as a use theory of meaning it is not clear if it is best understood as a semantics in this narrower sense.

Recently, it has been proposed by various authors that expressivism is not best viewed as a semantics, a systematic theory of which expressions mean what, but rather as a *metasemantics*.[2] A metasemantics, or foundational theory of meaning, as it is also called, is a theory that does not tell us what some expressions mean, rather it tells us in virtue of which properties those expressions mean what they mean or come to have a meaning at all. Such a metasemantic conception of expressivism captures the idea that expressivism is an explanatory theory the *explains* meaning. Moreover, this conception also fits with the idea that expressivism is a use theory of meaning, because it is generally assumed in metasemantics that the properties in virtue of which some expression means what it does, are properties concerning the *use* of those expressions. This captures the pragmatic dimension of expressivism.

So are expressivist theories best thought of as a semantics or a metasemantics? There is not always a clear answer to this question. Consider, for instance, the theory proposed by Mark Schroeder (M. Schroeder, 2008a) upon which my own will be modeled in the next chapter. Schroeder thinks that the form of expressivism which he investigates is what he calls a "semantic program" (see the subtitle of Schroeder's book). Even though he clearly does not hold that the mental states that are expressed by sentences are *themselves* their meanings, he treats the contents of those mental states as the 'semantic values' of sentences. In his specific variant of expressivism he treats those semantic values of sentences as ordered pairs of properties. What Schroeder's theory then does is systematically assign those semantic values to expressions of a language which suggests that Schroeder is concerned with doing

---

[2]For more on metasemantics see for instance (Burgess & Sherman, 2014), and in the context of metaethics and especially expressivism see (Schroeter & Schroeter, 2017), (Chrisman, 2016, ch. 1.3), (Ridge, 2014, 8-9) and (Köhler, 2018).

semantics in the narrow sense. Moreover, in his other work (M. Schroeder, 2013), Schroeder defends the claim that those pairs of properties should be taken to play the role that propositions are normally taken to play in order to account for the meaning of sentences. These points suggest that Schroeder himself conceives of his theory as a first-order semantic theory and not a foundational theory of meaning or metasemantics.

On the other hand, however, the 'semantic values' that Schroeder's semantics assigns to moral sentences can hardly be regarded as what we ordinarily take to be their meanings. If someone asks you what 'murder is wrong' means, it will be unhelpful to tell him which ordered pair of properties is its semantic value in Schroeder's expressivist semantics. If someone wants to know what 'murder is wrong' means, it will be much more helpful to tell him that it means that murder is wrong or that it means the same as the German 'Mord ist moralisch falsch'. This suggests that the semantic values in Schroeder's expressivism are perhaps better thought of as being properties of things (namely the contents of the mental states we use sentences to express) that help to explain why 'murder is wrong' means that murder is wrong. This suggests that Schroeder's theory can also be regarded as a metasemantics. Hence, in the case of Schroeder's expressivist theory it is unclear if it is better viewed as a semantics or rather as a metasemantics.

In this thesis I want to remain neutral about the question of whether expressivism in general, and the expressivist theory which I propose in particular, is best regarded as a semantics or as a metasemantics. In my view most forms of expressivism contain elements of both. On the one hand, when expressivists deal with the Frege-Geach problem they seem clearly engaged in solving first-order semantical problems, such as providing semantical rules for the logical connectives, e. g. the negation sign. On the other hand, expressivists also aim to answer more fundamental questions such as how it is possible for an expression to mean anything at all, which suggests that expressivists also want to answer metasemantical questions. In recent years there has been a certain trend for 'going meta' everywhere in philosophy. It seems that in many areas such as meta-normativity, meta-epistemology, meta-philosophy, and meta-metaphysics, philosophers are simply exploring old questions under new labels and metasemantics does not seem to be an exception.[3]

2.2 Why Expressing?

In this chapter, I want to critically discuss the notion of expressing. The goal is to present an explication of the expressing relation which is plausible and compatible with the broader commitments of expressivism.

---

[3]I do not mean to suggest that these questions are not well worth pursuing. My point is simply that the boundaries between first- and second-order investigations may not be as sharp as philosophical labels suggest.

Expressivism is the view that words and sentences have a meaning not in virtue of facts about what they refer to or are about. Expressivism is the view that words and sentences have a meaning in virtue of facts about which mental states they are conventionally used to express. Much attention has been devoted to the question *which* mental states are expressed by which sentences, and I myself shall offer a detailed answer to this question in the next chapter. But before that we should get clearer about what it is for something to *express* a mental state.

In this chapter, I will discuss a number of proposals and argue that at least some of them should be rejected. However, there are also some proposals which I take to be both plausible and suitable for expressivists' needs. Even though I have a preference for an old proposal made by John Searle in his (Searle, 1969) which is both simple and general, I do not want to reject those other plausible proposals.

'Expressing' is the name-giver of expressivism, but why is expressing important to expressivism? One answer is that expressivism is an attempt to overcome the problems that beset a certain descriptivist view in metaethics which has frequently been called *speaker subjectivism.* Speaker subjectivism is the view that, roughly, a moral sentence like 'murder is wrong' uttered by some speaker S should be equivalent to 'S disapproves of murder' or something similar. This view runs into two problems, which I will call, following Schroeder the *modal problem* and the *disagreement problem* (M. Schroeder, 2010b, 67f.). The modal problem is that if speaker subjectivism is true, then 'If S did not disapprove of murder, then murder would not be wrong' is equivalent to 'If S did not disapprove of murder, then S would not disapprove of murder'. However, these two sentences are not equivalent since the latter is tautological, whereas the former is not.[4]

The other problem is the disagreement problem. The problem is simple. When Alice says that murder is wrong and Bob says that murder is not wrong, we intuitively take them to be disagreeing with each other. However, according to speaker subjectivism, what Alice says is equivalent to 'Alice disapproves of murder', and what Bob says is equivalent to 'Bob does not disapprove of murder'. In this case, however, the intuitive disagreement is lost, or at least it is no longer clear that there is any disagreement at all.

The diagnosis given by expressivists is that these problems arise, because speaker subjectivism has a mistaken view about the relationship between 'murder is wrong' and the mental state of disapproving of murder, namely that it holds that the former *reports* the latter. Expressivism hopes to avoid these problems by noting that there is a crucial distinction between reporting that one is in some mental state and *expressing* a mental state.

---

[4]Another take on the problem is that if speaker subjectivism were true, then whether or not something is wrong depends on whether or not someone disapproves of it or not. This, however, runs contrary to our strong conviction that morality is objective, or at least not subjective in such a strong sense.

According to the expressivist, the modal as well as the disagreement problem for speaker subjectivism are not due to something special about moral vocabulary, but simply arise if we do not distinguish between reporting and expressing a mental state. This can be seen by noting that those problems would arise just as well if we would hold that 'grass is green' is equivalent to 'S believes that grass is green'. If 'grass is green' would report that the speaker believes that grass is green, we would again run into the modal and disagreement problems. Since, however, we think that the modal and disagreement problems do not arise for sentences like 'grass is green', the problem with speaker subjectivism suggests that it has a mistaken conception of the relationship between sentences and mental states, namely taking it to be that of reporting to be in some state. Hence, the "fundamental idea of expressivism", as Schroeder calls it, is "that 'murder is wrong' *stands to* disapproval of murder in the *same way as* 'grass is green' stands to the belief that grass is green" (M. Schroeder, 2008a, 18). Expressivists propose that we call this relation, whatever it is, the *expressing* relation. The remarkable thing about the expressing relation, expressivists think, is that, whatever exactly it is, it avoids the two major problems of speaker subjectivism in metaethics.[5]

What then is the expressing relation? We seem to have some pre-theoretical understanding of the expressing relation. The reporting-expressing distinction is often illustrated by examples like the following. If a speaker uses 'I believe that grass is green' he *reports* or describes the fact that the speaker holds the belief that grass is green. In contrast, a speaker who uses 'grass is green' does not report that he believes that grass is green, but rather *expresses* the belief that grass is green. But it is not clear that this pre-theoretical understanding will actually suite all the purposes of expressivism.

Since the notion of expressing is central to expressivism, it is important to have a clear understanding of what expressing is. Unfortunately, both defenders and critics of expressivism have rarely made precise what they take expressing to be, and this has led to some confusion.[6]

In asking what expressing is, we should note that we can ask either for a *definition* or for an *analysis* of the notion of expressing. If we seek a definition, then we aim to provide a set of jointly necessary and sufficient conditions, but do not require that these conditions capture anything of our pre-theoretical understanding of expressing. Some authors explicitly accept

---

[5]Of course, contemporary subjectivists in metaethics think they can solve the modal and disagreement problems without having to become expressivists. The most prominent proponent of subjectivism or relativism is Jamie Dreier. See his (Dreier, 1990).

[6]For instance, Jackson and Pettit in an influential paper (Jackson & Pettit, 1998) presented an objection to expressivism. Their objection crucially rested on an implicit understanding of the expression relation that expressivists should not and do not accept. For discussion of their objection see (Dreier, 2004a) and (M. Schroeder, 2008b).

that their proposed definition of expressing is a purely theoretical notion and may not have anything to do with our ordinary understanding of expressing a mental state.[7] On the other hand, other authors, such as Wayne Davis, explicitly regard their conception of expressing as a genuine analysis of our ordinary notion of expressing (Davis, 2003, 12ff). Since analyses always want to capture our ordinary understanding of some concept, and can therefore also fail to capture our understanding, analyses of expressing can be and have been subject to counterexamples, whereas definitions of expressing are not.

I think that both approaches to expressing, the definitional as well as the analytical, are in some ways problematic. The definitional approach is problematic, because expressivism, as I note above, is in part motivated by appealing to an intuitive difference between reporting and expressing. Hence, a definition of expressing that has not the least to do with our intuitive understanding of it would rob expressivism some of its initial appeal. The analytic approach is also problematic, because it is not clear that our ordinary notion of expressing is sufficiently sharp to admit of an analysis at all, and it is also not clear that our ordinary notion is really suitable for the purposes of expressivism. Thus, I believe, the best approach lies somewhere in between and will rather take the form of a Carnapian *explication* of the notion of expressing.[8] So, a plausible approach to expressing will have to respect our ordinary understanding of expressing at least to a certain extent, but also allow for certain deviations from or revisions of the ordinary understanding when this is required for providing a notion suitable for purposes of expressivism.

Before we take a closer look at expressing, I want to explicitly mention an important metasemantic commitment of expressivism, namely *mentalism*. Mentalism is the view that mental content is explanatorily more basic than linguistic content. This is a very old idea, and is historically associated with John Locke's view that the meanings of words are ideas or impressions on the soul.[9] Mentalism is thought to assist expressivists with their project of explaining the meaning of sentences in terms of the mental states they express, by allowing them to make use of those states' contents. Roughly the idea is that 'grass is green' means what it means, namely that grass is green, because it expresses the belief that grass is green which has the content that grass is green.

There is, however, a complication with this simple picture of how mental content plays a role in explaining the meaning or semantic content of sentences in the context of metaethical expressivism. The complication is that metaethical expressivists deny that there are such things as genuine

---

[7]See for instance (M. Schroeder, 2008b, 29).

[8]Rudolf Carnap introduced the notion of an 'explication' of a concept in his (Carnap, 1951, 3-15).

[9]See (Locke, 1690) and for discussion of Locke's ideational semantics (Hanna, 1991).

moral mental contents or moral propositions, at least if they are thought of a the ultimate explanans of meanings. For instance, they deny there is a genuine moral proposition that murder is wrong, since if there were such a thing, it seems that this proposition could perfectly well also figure as the linguistic content of the moral sentence 'murder is wrong'. This, however, would make the whole expressivistic project pointless, since the goal of this project is precisely to provide an explanation of how moral sentences can have a meaning if there is no such thing as genuinely moral content, either semantic or mental.

So, expressivists are mentalists in the sense that they think that sentences have the meaning they have in virtue of the fact that they express some mental state with some specific mental content, but they deny that 'murder is wrong' means *that murder is wrong* in virtue of this being the content of the mental state expressed by it. So even though expressivists hold that 'murder is wrong' means that murder is wrong in virtue of the fact that it is conventionally used to express a certain mental state, disapproval of murder, say, their claim is not that the content of a sentence derives straightforwardly from the content of the state expressed.[10]

2.3 The Expressing Relation

Let us approach the matter a little bit more systematically. What is expressing? It is clearly a relation. In the following I shall mostly speak as if it were a two-place relation: something expresses something.[11]

What are the relata of the expressing relation? Regarding the right-hand side of the expressing relation, the answer is fairly uncontroversial among expressivists. In line with the expressivists' commitment to mentalism I will focus on mental states as being the things that *are* expressed, and ignore those 'Fregean' or semantic senses of 'express' as we use it when we say that sentences express propositions or 'thoughts', or when we say that predicates express properties or concepts.

But even if it is clear that the things expressed are mental states, we have to draw an important distinction, namely that between types and tokens of mental states. We might signal this difference by saying 'she expresses *her* belief' (token) or 'she expresses *the* belief' (type).[12] There is an important

---

[10]This suggests that expressivists need to make a distinction between the meaning of a sentence and the mental content of the states expressed. One popular idea among expressivists has been to embrace a minimalist conception of proposition in order to account for the meanings of sentence. One idea would be to treat propositions as abstract entities that somehow represent ways of uses. For a minimalist or deflationary account of propositions suitable for expressivists see Köhler (Köhler, 2013).

[11]In reality, however, it might have much more places. For instance, consider how many places we have in: a sentence F at time t in circumstance C as used by some speaker S in order to do action A expresses mental state M.

[12]I do not want to suggest that this way of speaking corresponds to the distinction of

difference between the relation of expressing a token of a mental state and the relation of expressing a type of state. Whereas the former does not allow for insincerity, the latter does. It is not possible that you express *your* belief that p, without you actually holding that belief. But it is possible that you express *the* belief that p, without you actually holding that belief. The latter relation of expressing is in this sense *intentional* or *non-factive*. The intentionality of the expressing relation is important to the expressivists semantic project, because it allows them to explain how a word's or sentence's meaning can be a relatively stable property, despite its being partly determined by what state a speaker is in language is used to express.[13] Primarily for this reason, most expressivists, at least implicitly, assume that the relevant expressing relation has to be non-factive, and thus treat the states expressed to be types of states.

What about the left-hand-side of the expressing-relation? What kinds of things express mental states? In the literature we find a great variety of candidates: agents, sentences, events, speech acts, utterances, judgments. All seem plausible to a certain extent. It makes sense, for instance, to say that *Mary* expressed her or the belief that grass is green. It also makes sense to say that her *uttering* 'grass is green' or her *asserting* it expresses the belief that grass is green. It is also fine to say that the *sentence* 'grass is green' expresses the belief that grass is green. And in many cases it even makes sense to think of *events* as expressing states, as, for example, when someone's *carrying* an umbrella expresses her belief that it might rain.

Some of these candidates feel more natural than others, and some seem to be merely elliptical ways of speaking rather than speaking properly. For instance, do we really think that utterances or speech acts can express a mental states? Likewise, do we really think that sentences are the kinds of thing that express mental states? This seems somewhat unnatural. But being somewhat unnatural of course need not be a devastating objection against an explication of the expressing relation, since it might simply be fruitful from the perspective of semantic theorizing to have a more technical notion of expressing that directly links sentences to mental states.

I suspect that the hesitation to thinking that sentences, events or speech acts can express mental states is due a natural inclination to think that expressing something is first and foremost an action—that expressing is something someone *does*. I too find it natural to think that expressing a mental state is first and foremost an act performed by someone. However,

---

expressing mental state tokens or types in natural language. In ordinary life the distinction mostly does not matter. Normally when a person expresses a belief, we naturally assume that this belief is also a belief of hers, a belief she not only expresses, but also actually holds herself. This is so, because absent any reason to think otherwise, we assume our interlocutors to be sincere.

[13] The importance of this for explaining the stability or constancy of meaning has already been noted by Stevenson. See (Stevenson, 1945, 43f.).

even if expressing is fundamentally something someone does, that does not mean that it is an action in the same way like riding a bike is something someone does. Rather it is an act someone performs *in virtue of* performing some other action. Expressing a mental state might be something someone does in the same way as polluting the air is something someone does *by* driving a car, or it is similar to turning on the light *by* flipping the switch. In other words, we might think of expressing a mental state as being a side-effect of doing something else. However, that side-effect need not be a *causal* effect of someone's doing something, as the examples suggest: rather it could also be viewed as a 'definitional' effect such as someone's doing math by calculating the sum of two numbers, or violating a traffic rule by ignoring a red light.

A useful classification of different types of expressing that incorporates some of the just mentioned features, is made by Wayne A. Davis, and it will be helpful for the following discussion to have this classification in mind (Davis, 2003). Davis distinguishes evidential, word, and speaker expressing, and provides the following examples:

Evidential Expressing: The look on S's face expresses fear.

Speaker Expressing: By saying 'I'm afraid' S is expressing fear.

Word Expressing: The word 'fear' expresses the idea of fear.[14]

What Davis' calls word expression applies not only to words, as the name suggests, but also to whole sentences. I shall therefore use the more general terms *Sentential* or *Semantical* Expressing to also cover cases like: The sentence 'This is fear' expresses the belief that this is fear. Moreover, it might be helpful to distinguish also a fourth sense of expressing, or a second sense of speaker expressing. The idea is that there is a sense in which a *speaker* can express a mental state without him having any control over which state is expressed, namely when he uses a sentence that semantically expresses some mental state. Let us say that a speaker *semantically speaker expresses* a mental state iff the speaker uses a sentence that semantically expresses that state. The difference between speaker expressing and semantic speaker expressing is that a speaker can, in principle, use 'I'm afraid' to speaker express pleasure, say, instead of fear, but it is impossible for a speaker to normally use 'I'm afraid' and not semantically speaker express fear, since that is, after all what the sentence semantically expresses, and individual speakers cannot, merely by wanting to express something else, change that semantic fact.

---

[14]Cf. (Davis, 2003, 43). Davis, like Schroeder, prefers to speak of 'expression' instead of 'expressing'. I use 'expressing' in order to avoid potential confusion with 'expression' in the sense of being linguistic entity like a word or sentence.

It is an interesting and natural question what the relations between these different types of expressing are, and if one of them, as seems reasonable, is more basic than all the others. I shall return to this question below in my discussion of Davis' 'expression theory of meaning'.

In the following, I want to focus on sentential or semantic expressing in the sense that sentence types as a matter of semantics express mental state types, because the expressivist semantics which I shall develop in the next chapter will assign mental state types to sentence types. In the following I will drop the qualification 'types' and 'expressing' means 'semantically expressing' unless stated otherwise. So whenever I say 'mental state' and 'sentence' I mean types.

The question that interests me in the moment is: how can the relation of semantic expressing be explicated so that it is suitable for the purposes of giving such an expressivist semantics?

I suggest that for an explication of the expressing relation to be adequate for the purposes of expressivism is must satisfy at least the following constraints, some of which I have already mentioned in passing:

Intentionality/Non-factivity: It must be possible for a speaker to express a mental state she is not in.

Non-relativity/Constancy: The mental state that is semantically expressed by a sentence must not vary depending on what an individual speaker intends to do with that sentence.

Metasemantic Compatibility: The expressing relation must not conflict with the expressivists' commitment to mentalism.

Generalizability: The account of semantic expressing must apply not only to indicative sentences, but to imperative and interrogative sentences as well.

It will become clear in the following sections, why an explication that does not satisfy these constraints will not be suitable for the expressivists needs. I will now discuss several accounts of expressing and evaluate how well they fare with respect to these constraints.

2.4   Accounts of Expressing

One of the most detailed discussions of the expressing relation is provided in a paper by (M. Schroeder, 2008b) and his book (M. Schroeder, 2008a). There Schroeder discusses and rejects four different explications of the expressing relation, namely the *same content* account, the *causal* account, the *implicature* account, and the Gricean or *indicatory* account. After that he presents his own explication which he calls *assertability expressivism*. In the following I shall present those accounts and Schroeder's arguments against

them as well as his arguments for assertability expressivism. We will see that not all of Schroeder's objections are as convincing as they appear, and that even though assertability expressivism is a viable option for expressivists, there are other and perhaps more attractive options available. Ultimately, I present an explication which I myself favor. It is build on an old remark from John Searle and explicates expressing in terms of *sincerity conditions* of speech acts.

2.4.1   Same Content Account

The first explication of the expressing relation I want to discuss is the following:

Same Content Expressing: A sentence F expresses a mental state M iff F has the same content as M.

This is arguably the most natural account of semantic expressing: a mental state counts as being expressed by some sentence simply in virtue of the fact that they share the same content. The same content account, however, is inadequate for the purposes of metaethical expressivists for two reasons.[15]

First, the same content account conflicts with the expressivists' commitment to mentalism, since they have different orders of explanation. According to mentalism, sentences acquire a content from the mental states they express. The same content account of expressing, however, requires that sentences have a content first, and thereby determine which state they express. Hence, expressivists cannot explain the meaning of a sentence in terms of the state it expresses, when which state it expresses is explained by the meaning of the sentence. So there is a problem with the order of explanation.

The second reason is that, as I have already remarked above, metaethical expressivists deny that there are such things as genuine moral contents, such as the proposition that murder is wrong that could be the object of a moral belief. It is for this reason that they are noncognitivists. They think that what makes moral judgments special is not that they have a special sort of moral content, but that they are special sorts of noncognitive mental states with non-moral contents. But if to judge that murder is wrong is to be in some noncognitive state with a non-moral content, like disapproval of murder, say, then for 'murder is wrong' to express this noncognitive state, they would have to have the same content. But the content of this disapproval state is *murder*, or the concept of murdering. But this can hardly also be the semantic content or meaning of 'murder is wrong'.

---

[15]As the view is currently formulated it is obviously inadequate for another reason. It must be restricted to *beliefs* having the same content, for it does not make much sense to say that 'grass is green' expresses the intention that grass is green in virtue of their having the same propositional content.

Hence, although the same content account is a natural explication of expressing, it is not adequate for the expressivists' purposes, because it conflicts with mentalism as well as with the metaethical expressivists' noncognitivism.

2.4.2   Causal Account

Another account of expressing is not in terms of content, but in terms of causality. The idea is that the expressing relation is a causal relation. This understanding seems to be implicit in the writings of early emotivists such as Ayer, and also in certain passages of Blackburn.[16] The idea can be formulated as follows:

Causal Expressing: A sentence F expresses mental state M iff a speaker's utterance of F is caused by the speaker's being in M.

Schroeder thinks that this account is hopeless as an adequate explication of expressing, but I am not so sure. Schroeder mentions two "fatal" problems for this account, one having to do with lying, the other with minimalism about truth (M. Schroeder, 2008b, 100). Both of these problems are due to the fact that if expressing is a causal relation, you cannot express a mental state you are not in, because a state you are not in cannot cause you to do something. In other words, it is incompatible with the above constraint that any adequate explication of the expressing relation must allow for insincere expressing, that is, be intentional or non-factive. Schroeder says:

> The problem with a causal account of the expression relation, no matter how modified or qualified, is that causation requires the expressed mental state to actually *exist*. But insincerity requires there to be sentences that express mental states that do *not* actually exist. In order to utter a sentence insincerely but with the same meaning as a sincere utterance, it has to be possible for you to express a mental state that you are not actually in. (M. Schroeder, 2008b, 26, italics original)

Schroeder here reminds us that according to expressivism the meaning of a sentence is given by the mental state it expresses. An explication of expressing suitable for this project must allow for insincerity, because if it did not, then someone who speaks insincerely, such as a liar, would not express the same mental states as those who speak truthfully. As a consequence of the expressivists' explanatory project the sentences uttered by liars and those who tell the truth would then mean different things. However, a sentence has to mean the same thing in the mouth of liars as well as non-liars for them to

---

[16]Compare Ayer's talk of "ejaculations" being caused by emotions, and Blackburn's talk of moral words "voicing" or "venting" our states of mind. Cf. (Ayer, 1952, 105) and (Blackburn, 1998, 50).

be able to genuinely disagree with each other. Since the causal account does not allow to express states you are not in, the causal account is inadequate for expressivism.

I think that Schroeder's argument is somewhat too hasty. If we look carefully at Schroeder's writings on expressing, we see that what makes evaluating his argument difficult is that he carelessly switches from talking about sentences expressing states, to speakers and utterances expressing states. That is he ignores the above-mentioned distinction between speaker expressing and semantical expressing. Equally important in this context is that he is insensitive to the aforementioned distinction between expressing *your* mental state token, and expressing *a* mental state type. This is obviously relevant here, because, as I have remarked above, these two understandings differ precisely in that the latter allows for insincerity, whereas the former does not. Schroeder is surely right in claiming that a liar cannot express, in the speaker expressing sense, *his* token belief that p by uttering P, since, after all, he does not hold the belief that p and so the belief cannot cause his utterance. But that does not rule out the possibility that a speaker in uttering a sentence expresses *the* belief that p, and it also does not rule out the possibility that the sentence he utters expresses the belief that p despite the fact that he is insincere. But there still is the problem that causal expressing require that for a sentence to express a mental state every utterance of it must *actually* be caused by a mental state. This problem, however, is easily solved by modifying the causal account by saying that for a sentence to express a mental state it suffices that the state causes an utterance of some specific sentence at least in the sincere or typical cases. So we may refine the causal account in order to take care of liars:

Causal Expressing (refined): A sentence F expresses mental state M iff a speaker's utterance of F is typically caused, in normal, sincere cases, by the speaker's being in M.

I do not want to investigate further whether this refined causal account of expressing will ultimately be adequate, since I believe that there are more attractive accounts anyway. Moreover, it seems that even the refined version might be subject to other problems. For instance, it is not at all clear whether any kind of causal relation between a mental state and someone's utterance suffices for expressing. Think, for instance, of very long and complex causal chains. It seems clear that for a causal account to be viable it would have to determine an *appropriate* length and complexity of the chains causing speakers to utter sentences. This would, however, introduce an unwelcome vagueness into the account. I just want to emphasize that Schroeder's arguments against it may not be as convincing as they appear.

2.4.3   Implicative Expressivism

Another account of the expression relation is in terms of Gricean conventional implicatures. Such an account is advocated in the literature for instance by Stephen Barker (Barker, 2002) and David Copp (Copp, 2001). It says:

Implicative Expressing: A sentence F expresses mental state M iff F, as a matter of conventional implicature, conveys the information that the speaker is in M.

This account, like the same-content account, also seems to be incompatible with the expressivist's mentalism. Schroeder explains that according to the expressivist:

> the mental state expressed by a sentence is no mere pragmatic overlay on language, but what explains how non-normative sentences come to have the propositional *semantic* contents that they do in the first place. Accepting this view would be like holding that "but" denotes conjunction *because* it is conventionally used contrastively. (M. Schroeder, 2008b, 101, italics original)

It is not clear to me what is Schroeder's point here. The analogy with 'but' is misleading. The above would surely be a bad metasemantic theory about how it is that 'but' comes to denote conjunction. But that at most shows that being conventionally used contrastively is not the 'pragmatic overlay' which is relevant for explaining why 'but' to denotes conjunction. 'But', just like 'and', may denote conjunction due to some other information that is conveyed by means of a conventional implicature, such as for instance the information that the speaker is in a 'conjunctive' mental state or ready to draw certain inferences etc. Of course, we would then have to explain what a 'conjunctive' mental state is or what the relevant inferences are, and we cannot say that it is any state expressed by a conjunctive sentence, since that would be circular, but is seems that in principle such an account could be given.

For instance, even though expressivists are not allowed to appeal to semantic contents in their explanation of what state is expressed by a given sentence, their mentalism allows them to appeal to mental contents. And so it seems possible to claim that a conjunctive state is a state the content of which is a conjunction of two contents. For instance the belief that p and q is a conjunctive state, because its content is a conjunction of p and q. Hence, it seems open for expressivists to claim that a sentence containing 'and' has conjunctive semantic content in virtue of the fact that as a matter of conventional implicature it is used to convey the information that the speaker is in a conjunctive mental state.

I do not see any principled reason, why it should be impossible for an expressivist to claim that such sentences convey this information as a matter of conventional implicature, and so express mental states in this way. Schroeder's mistake, I think, lies in having too narrow a conception of what could and could not be conveyed as a matter of conventional implication.

2.4.4 Indicatory Expressivism

The final account of expressing which Schroeder discusses, before he presents his own favored account, is another Gricean account which Allan Gibbard proposed in his book *Wise Choices, Apt Feelings* (Gibbard, 1990) and which he refined in his book *Thinking How to Live* (Gibbard, 2003). Gibbard's account takes roughly this form:

Indicatory Expressing: A sentence F expresses mental state M iff a speaker utters F with the intention to indicate to his audience that he is in M.[17]

The indicatory account of the expressing relation respects the constraint that expressing has to be intentional or non-factive. On this account it is possible to express a mental state the speaker is not in, because a speaker can intend to indicate to his audience that she is in some state without her having to actually be in this state. So this account avoids the sincerity problem of the causal account. Moreover, what mental state is expressed is also 'no mere pragmatic overlay' as it is on the implicative account. What mental state is expressed is not a matter of conventional implicatures, but a matter of the speaker's *intentions*. Furthermore, this understanding of expressing appears to be compatible with the expressivist's mentalist metasemantics: Gibbard thinks that it is by intending her utterance to indicate to an audience what mental state a speaker is in that her utterance comes to have the meaning that it has.

Yet, Schroeder thinks Gibbard's explication of expressing in terms of speaker intentions is subject to two problems.[18] The first is that on this account which mental state is expressed depends on the individual intentions of speakers. This together with expressivism's semantic agenda, generates an unfortunate dependence of meaning on intention:

> If the mental state expressed is merely a matter of the speaker's intentions—of what is going on in her head at the time—then it is natural to think that it could easily happen that a speaker happens to have these intentions with respect to sentences which do

---

[17]Cf. Schroeder 2008b, 26; Schroeder 2010,102.

[18]Actually, Schroeder raises three problems, but the first two problems have the same core, and so I will not consider them both, especially since I think they both rest on the same misunderstanding.

not, intuitively, have the appropriate semantic contents. (M. Schroeder,
2008a, 26)

The problem should be clear. According to expressivism, sentences have
their meaning in virtue of the mental states they express. For instance,
the sentence 'Moriarty has arrived' means what is does due to the fact
that it expresses the belief that Moriarty has arrived. Since on Gibbard's
understanding of the expressing relation to express a mental state is to intend
to indicate that you are in some particular mental state, your utterance of
this sentence expresses the belief that Moriarty has arrived if you intend to
indicate that you believe that Moriarty has arrived. Hence, according to
Gibbard, 'Moriarty has arrived' means what it does, if and because it is used
with the intention to indicate that the speaker believes that Moriarty has
arrived.

Obviously, it cannot be so simple. It is clearly possible that someone
utters some other sentence, for instance the sentence 'I hereby express my
belief that Moriarty has arrived', with the exact same intention which we
just used in order to explain why 'Moriarty has arrived' means what it does,
namely with the intention to indicate to the audience that you believe that
Moriarty has arrived. The problem is that even though it is possible to utter
'I hereby express my belief that Moriarty has arrived' with this intention,
this sentence does not thereby come to mean that Moriarty has arrived. The
problem arises, because on "Gibbard's interpretation, expressing a mental
state is something that a *speaker* does, and which is subject to the speaker's
intentions" (M. Schroeder, 2008a, 29). But what a sentence means does not
and should not vary with a speaker's intentions, at least not that directly.

However, this problem, as Gibbard as well as Schroeder are aware of, can
be solved by restricting the speaker's intentions to allow only intentions that
"take advantage of *already existing linguistic conventions*" (M. Schroeder,
2008b, 105). I agree that appealing to linguistic conventions is one way out of
the problem of expressing being dependent on speaker intentions.[19] Again, I
think that the problem can be avoided if we are careful enough to distinguish
between speaker expressing and semantical expressing, and if we are able
to define a speaker *in*dependent notion of semantic expression in terms of a
speaker dependent notion of speaker expressing. We will see how this can be
done in section 2.4.6.

But there is a second problem for indicatory expressivism, which Schroeder
takes to be more difficult. The problem is that indicatory expressivism
is "uncomfortably unstable" (M. Schroeder, 2008b, 106). What does this
instability consist in? If expressing were spelled out in terms of intentions to
indicate, then an expressivistic semantics would tell us for each sentence of
some language that it is conventionally used by speakers with the intention

---

[19]I want to mention that the problem might also be solved, if we add a normality or
typicality clause as we did above with the causal account.

to indicate that the speaker is in such-and-such mental state. Hence, each sentence would be conventionally assigned indirectly, that is as part of the indicatory intention, the information *that* the speaker is in some specific mental state. Schroeder calls this information a sentence's 'ur-content'. In Gibbard's theory those ur-contents, i. e. the information that a speaker is in such-and-such state, are the things partly in virtue of which sentences have the semantic contents they have. To make this more concrete: indicatory expressivism, on Schroeder's understanding, tells us that the sentence 'grass is green' is conventionally used by speakers with the intention to indicate that the speaker believes that grass is green, and it is in virtue of the fact that speakers use this sentence with the intention to indicate precisely *that* information (i. e. *that* they believe that grass is green) that the sentence 'grass is green' comes to mean that grass is green. "The problem", says Schroeder, "is that on this view, ur-contents turn out to be so robust, that it makes it hard to see why ur-contents aren't normal contents" (M. Schroeder, 2008b, 106). This raises the question why we need normal contents when ur-contents seem already able to do all the work.

Schroeder thinks that Gibbard does not have a satisfying answer to this question, and concludes that indicatory expressivism therefore does not provide an adequate explication of the expressing relation.[20]

Schroeder thinks that the problems confronting indicatory expressivism are essentially due to a single assumption, namely "that the intentions of the speaker which matter [for expressing] are intentions to provide information [the ur-contents] to her audience that she is in the requisite state of mind" (M. Schroeder, 2008b, 29). Schroeder thinks that if we can have a better take on the relation to ur-contents than letting them be what speakers intend to communicate to an audience, expressivists can develop an account of expressing that takes from Gibbard what helps avoiding potential conflicts with mentalism and intentionality, but does not run into the instability worry. It seems that he is right about this, and in the next three sections I present three further conceptions that all drop this problematic Gricean assumption in Gibbard's account, and so avoid the instability.

---

[20]Yet Schroeder himself explains why Gibbard, as a metaethical expressivist and noncognitivist, needs such a distinction between ur-contents and contents. He writes: "[Expressivists] think that we need this distinction, because it is necessary in order to allow normative and non-normative vocabulary to compose under the logical connectives, by giving non-normative sentences an extra level of content to correspond to the only kind of content that normative sentences have" (M. Schroeder, 2008b, 107). This seems to answer Schroeder's question about the expressivists' need for ur-contents as well as normal contents.

2.4.5   Davis' Expression Theory of Meaning

In this section I want to present Wayne Davis' so-called 'expression theory of meaning' which he develops in his (Davis, 2003). Davis' overall goal is to provide a foundational theory of meaning or a metasemantics. As such it covers a lot more ground than is relevant for the topic I discuss in this chapter. But Davis' theory is interesting for two reasons. First, Davis' expression theory of meaning is the most detailed treatment of the foundations of expressivist semantics, and as such should be interesting for expressivists of all stripes. Second, an integral part of Davis' theory of meaning is an analysis of the expressing relation, in particular of speaker expressing. Davis' analysis of expressing is particularly interesting in the present context, because it is a very close relative of indicatory expressivism, which, I shall argue, avoids the two problems of Gibbard's account.

At the beginning of this chapter, I already introduced Davis' distinction between evidential expressing, speaker expressing and word or semantical expressing. This distinction parallels another distinction he draws with regard to *meaning*, and will be familiar to everyone who is acquainted with the cornerstones of Grice's views on meaning.[21]

Davis, as in the case of expressing, distinguishes three different senses of meaning: evidential meaning (which he also calls 'indication'), speaker meaning, and word or semantical meaning. Ultimately, Davis' analysis of meaning is grounded in his account of indication, that is evidential meaning. He has a very nuanced view of what is involved in providing an indication, but this chapter is not the place to present Davis account in all the details. For our purposes only one feature will be relevant, namely that something can be an indication without it been recognizable by anyone to be an indication. This is crucial for avoiding Schroeder's instability worry against indicatory expressivism. I have more to say about this in a moment. The following examples suffice to get at least a grasp of what Davis means by these different senses of meaning:

Evidential Meaning: Boulders mean glacial activity.

Speaker Meaning: By 'boulder', S means 'kilo of cocaine'.

Word Meaning: 'Boulder' means 'large rounded stone block'.[22]

As I said, Davis' interest is primarily in meaning, but as the name of *expression* theory of meaning suggests, expressing mental states plays a central role in this theory. Like the expressivists, Davis is also committed to a form of mentalism. According to his theory "words are conventional signs of mental states", and "meaning consists in their expression" (Davis,

---

[21]For Grice's views see (Grice, 1989).
[22]Cf. (Davis, 2003, 19).

2003, 1). Davis attempts to carry out his project of analyzing meaning by realizing what he calls the 'Gricean Program'. It is in this program where the different notions of meaning and expression are crucial.

What is the Gricean program? In a nutshell, Grice's famous idea was that the meaning of *words* and sentences is somehow determined by what *speakers* mean by them, and that what speakers mean is determined by their *intentions*. On Grice's original conception, for a speaker to mean something is, roughly, to act with the intention to generate a certain effect in one's audience in virtue of the audience's recognition of that intention. As is well-known, however, Grice's rather complicated analysis of speaker meaning in terms of nested intentions is subject to a number of counterexamples. Finding a way to spell out speaker meaning in terms of speaker intentions that is not subject to those notorious problems, resulted in a flourishing literature in the second half of the 20th century.[23]

According to Davis' diagnosis the problems with Grice's analysis all stem from a single mistake, namely that Grice's analysis focused centrally on *audience-directed* intentions. Davis's central idea is to remedy this defect by replacing the Gricean intention to generate a certain effect in the audience with the intention to *provide an indication* that the speaker is in a certain mental state. This is of course pretty much the same conception as that of Gibbard above, and which Schroeder thinks makes the view unstable. Nevertheless, as Davis interprets the relevant intention, he is able to avoid the instability. I now want to explain how.

One point Davis repeatedly stresses is that Grice and many of his followers did not sufficiently distinguish between meaning something and communicating, and that this has been a major mistake in the debate. It is central to Davis' analysis of speaker meaning that a person can mean something without intending to communicate or even without intending to be understood by others. For Davis, meaning-intentions do not involve intentions to communicate. This is relevant in the present context, because Davis analyses speaker meaning in terms of speaker expressing and speaker expressing in terms of intention to indicate (or to provide an indication) that the speaker is in some mental state.

So the crucial difference to Gibbard's Gricean explication of expressing is that according to Davis one can intend to provide an indication without intending to communicate. In other words, on Davis' view, someone can provide something as an indication (and it can in fact be an indication) without intending anyone to recognize it as an indication. Davis shares with Gibbard the idea that for a speaker to express a mental state, and thus to speaker mean something, is for a speaker to *do* something. But for Davis, to mean something is an activity a person can do all by itself; it does not

---

[23]See for instance (Grice, 1968), (Bennett, 1976), (Sperber & Wilson, 1986), (Neale, 1992), (Schiffer, 1972).

require others or an audience. For Davis to *mean* something is like running, whereas for Gibbard it is more like playing tennis: one can run alone, but one cannot play tennis alone. Davis' account of speaker expressing is therefore radically audience-free.

The key take-away is that according to Davis, for a speaker to mean something or to express a mental state does not imply that speakers intend to communicate, that is convey the information, that they are in some state M.[24] On Davis' view, the information, or what Schroeder calls 'ur-content', that the speaker is in some mental state is not something which speakers intend to convey when they express that mental state. Hence, the question of why we need meanings or normal contents besides ur-contents does not arise. On Davis' theory, like on Schroeder's assertability expressivism, ur-contents could not be used by speakers to talk about the world; they rather play the role of a 'device of the semantic theorist'. I conclude, therefore, that Davis' theory shows that a Neo-Gricean version of indicatory expressivism is able to avoid Schroeder's instability objection.

What about Schroeder's first problem for Gibbard's indicatory expressivism? This was the problem that, recall, if expressing a mental state is something that a speaker intentionally does, and which state a speaker expresses crucially depends on which intentions the speaker has, then this seems to make what sentences mean speaker-relative in an unfortunate way. We have seen that in order to solve that problem, expressivists should appeal to pre-existing linguistic conventions, and also distinguish between speaker meaning and semantic meaning. Davis shows us how to do both things at once.

Davis' overall plan is to define semantic meaning as conventional speaker meaning, and speaker meaning in terms of speaker expressing, and speaker expressing in terms of intentions to indicate that the speaker is in a certain mental state.[25] Here is how Davis develops his 'expression theory of meaning' in slightly more detail:[26]

---

[24]This does of course not mean that speakers never intend to convey the information that they are in some state *in addition* to their intention to provide the indication that they are in some state when they mean something or express a mental state. The point is only that on Davis' view the intention to convey that information is not part of the analysis of speaker expressing.

[25]Another route would have been possible for Davis, namely to define semantic meaning in terms of semantic expressing, semantic expressing as conventional speaker expressing, which in turn he can define in terms of speaker expressing.

[26]This presentation of Davis's theory brushes over many of the fine details on which Davis spends a lot of time in his two books. My intention here, however, is not to provide a comprehensive discussion of Davis. The intention here is merely to illustrate in broad outline what an expressivist metasemantics based on a form of indicatory expressivism could look like. For the details of Davis' theory the reader is asked to consult (Davis, 2003) and (Davis, 2005).

Semantic Meaning: A sentence F means that p in language $\mathcal{L}$ iff F is conventionally used by speaker's of $\mathcal{L}$ to speaker mean that p.

Speaker Meaning: A speaker S speaker means that p iff S speaker expresses the belief that p.

Speaker Expressing: A speaker S speaker expresses the belief that p iff S performs a publicly observable action A with the intention to provide A an indication that S believes that p.

This little sketch shows how Davis analyses semantic meaning in terms of speaker expressing. Of course, our interest in this chapter is not so much on semantic meaning, but on the expressivists' concept of semantic *expressing*. Since, however, on Davis' view, speaker expressing is more fundamental than semantic expressing, it is easy to see how Davis can use his analysis of speaker expressing to also define a relation of semantic expressing as conventional speaker expressing that can be viewed as a refined version of Gibbard's explication:

Indicatory Expressing (refined) A sentence F semantically expresses mental state M iff F is conventionally used by speakers S of $\mathcal{L}$ to perform a publicly observable action A with the intention to provide A as an indication that S is in M.

At least prima facie, this refined version of indicatory expressivism does not suffer from any of the problems we have discussed so far. Unlike the same content account it is compatible with mentalism, since it takes mental content to be conceptually or explanatorily prior to semantic content. Moreover, unlike the causal account it allows for insincerity, since a speaker can intend to indicate that she is in some mental state, without being in that state. This is precisely what liars do. Moreover, according to refined indicatory expressivism what state a sentence expresses is a constant property of the sentence that does not depend on the intentions of individual speakers because it appeals to pre-existing linguistic conventions. So if expressivists want to explain meaning in terms of expressed mental states, meaning will not vary with speaker intentions. Finally, this theory is not unstable in the way Gibbard's account is. Hence, refined indicatory expressivism is not obviously less adequate for the purposes of expressivism than Schroeder's assertability expressivism.

On the contrary, refined indicatory expressivism might even be the more attractive account, because unlike Schroeder's and Ridge's conceptions it does not require normative concepts like semantic correctness, assertability, or accountability. Moreover, the explication of expressing that is provided by refined indicatory expressivism is more general than Schroeder's in that it is not restricted to indicative or declarative sentences.[27]

---

[27]Actually, Davis offers an even more general explication that not only applies to

2.4.6 Schroeder's Assertability Expressivism

Schroeder thinks that Gibbard's way of securing the intentionality of expressing is basically right: "a sentence expresses a mental state by being associated with the proposition that the speaker is in that mental state" (M. Schroeder, 2008a, 29). But Schroeder thinks that Gibbard's attempt to spell out how they are associated in terms of Gricean speaker intention is what causes the trouble. In his view, the relation should not hold due to something that the speaker does or intends. Rather being associated with this proposition should be a merely semantic property of the sentence.

According to Schroeder, that the speaker is in such-and-such state is not something that gets communicated, but can be viewed as naming the *assertability condition* of the sentence. This leads to a view which Schroeder calls 'assertability expressivism'. According to assertability expressivism the role of an expressivistic semantics is to assign what he calls "*semantic correctness conditions*" to the sentences of a language (M. Schroeder, 2008a, 30). Those semantic correctness conditions require that a language user is in a particular state of mind:

> [According to assertability semantics each sentence] is associated with a condition under which it is permissible (as far as semantics goes) to assert it. For example, the sentence, 'grass is green' is semantically associated with the condition that the speaker believes that grass is green. As a shorthand, we can say that the sentence 'expresses' this belief. (M. Schroeder, 2008a, 31)

Schroeder argues that assertability expressivism has all the benefits of Gibbard's indicatory expressivism, but none of its problems. First, it is compatible with mentalism. Sentences get their content from the mental states they express: "it is permissible to assert 'grass is green' only if you believe that grass is green, but not because the sentence has the content that grass is green—rather, it has that content because those are the conditions under which it is permissible to assert it" (M. Schroeder, 2008a, 31). Moreover, the expressing relation is intentional, because "insincere utterances are cases in which a speaker breaks the semantic rules governing the language, and asserts something that she is not allowed to assert" (M. Schroeder, 2008a, 31). Furthermore, it avoids the first problem that Gibbard's indicatory expressivism has, because it does not treat expressing as an action someone does, but as having an assertability condition. Assertability expressivism can therefore explain why a sentence such as 'I hereby express the belief that

---

sentences of different grammatical moods, but even applies to the subsentential level. According to him sentences get their meaning from the mental states they express, but since mental states, or 'thoughts', as he prefers to call them, on his view, have ideas as their 'parts', the meaning of a word consists in the idea it expresses. For the details see Davis ideational semantics see (Davis, 2005).

Moriarty has arrived' cannot be used to express the belief that Moriarty has arrived. This is so because the assertability condition of that sentence is that the speaker believes that he believes that Moriarty has arrived, and not that the speaker believes that Moriarty has arrived, which is the assertability condition of 'Moriarty has arrived'.

Assertability expressivism still appeals to ur-contents. But it avoids the instability worry which confronts Gibbard's indicatory expressivism, because ur-contents are not something that speakers intend to communicate to others. Schroeder explains:

> Assertability conditions [...] are a device of the semantic theorist. They are not a kind of *information* that speakers intend to *convey*. So there is no sense in which a community of speakers could get by, managing to communicate information to each other about the world, by means of assertability conditions alone. It is only because some assertability conditions mention beliefs, and beliefs have contents about the world, that speakers can manage to convey information about the world. (M. Schroeder, 2008b, 108)

I agree with Schroeder that assertability expressivism provides an explication of the expressing relation that avoids all of the problems we have discussed so far, and thus might be adequate for the expressivist's purposes. But it is not clear that it does not have problems of its own.

Schroeder himself discusses two potential problems for his account of expressing. The first one he mentions is that 'assertability' and 'semantic correctness' are themselves normative notions, and as such assertability expressivism as a semantics for moral, and more generally, normative terms might fall under its own scope. Schroeder himself confesses that he is not sure whether this is really problematic. But be that as it may, the problem can perhaps be avoided by replacing the normative notion of assertability with a more conventional notion, such as accountability. Such a view about expressing is advocated for instance by Michael Ridge. Ridge's conception of expressing is, roughly, this:

Accountability Expressing: A sentence F expresses mental state M iff a speaker who utters F is thereby liable or accountable for being in M.[28]

Perhaps this solves the problem with the normativity of expressing. The other problem Schroeder mentions is that he thinks it might be possible to raise new modal and disagreement problems for his assertability expressivism.[29] But since Schroeder is again unsure that this is possible, I will ignore this problem.

---

[28]Cf. (Ridge, 2015, 109).

[29]I said something about the modal and disagreement problem in section 2.2 above.

I see another problem with assertability expressivism which has to do with generalizability. Schroeder's explication of expressing does not provide a general explication of what it is for a sentence (declarative, imperative, or interrogative) to express a mental state. It only defines what it is for a declarative sentence to express a mental state, namely to have an assertability condition. But questions do not have assertability conditions.

I believe that it is a constraint on any adequate explication of the expressing relation that it not only must allow declarative sentences, but also imperative and interrogative sentences, to express mental states. The generalizability constraint is reasonable, since it is possible in natural languages to combine declarative sentences and imperative or interrogative sentences with logical connectives. For instance, we can say things like 'If murder is wrong, what shall I do?' or 'If it starts raining, go home!'. It is reasonable to expect that a fully worked-out expressivist semantics for natural languages will have to explain the meaning of sentences like these in terms of the expressing of mental states as well, and not only the meaning of declarative sentences, since questions can contain moral vocabulary as well. If expressivists want to explain the meanings of such sentences also in terms of the mental states they express, they need a single account of expressing that applies to all sentences, and not one for declaratives, and another one for imperatives, and yet another one for interrogatives. That is why I think accounts of expressing must be generalizable.

I do not claim that it is impossible to transform assertability expressivism into a more general 'correctness conditional' semantics. Perhaps Ridge's accountability expressivism does just this.[30] The natural thing to expect from a such a semantics for questions, for instance, would be that it states for each interrogative sentence a condition under which it is semantically permissible to *ask* it, which perhaps is the case when the speaker wants or desires some information. In any case, for Schoeder's theory to be suitable for this broader expressivist program, he would have to show how to generalize his assertability account.

Schroeder thinks that assertability expressivism is superior to Gibbard's indicatory expressivism because it drops the Gricean assumption that to express a mental state is *inter alia* to intend to provide the information that the speaker is in some mental state. In the previous section I we have seen that there is a Neo-Gricean version of indicatory expressivism available that does not make that assumption, and also avoids the first problem to which Gibbard's account is subject, namely that the state expressed varies relative to speaker intentions. If this is correct, then expressivists have a choice if they want to be assertability expressivists or indicatory expressivists.

---

[30] Also compare Daniel Boisvert's 'success-conditional semantics' (Boisvert, 2014).

2.5 Sincerity Expressivism

In this final section on the expressing relation I want to suggest that there is a further explication of expressing which has been ignored so far in the debate. This conception is particularly attractive because it is simpler than the others, non-normative and, due to its relation to speech act theory, also closer to the historic origins of expressivism.

The early expressivists have held that to know what a sentence means we have to know which speech act it is used to perform.[31] How is this related to the contemporary view of expressivism that what a sentence means is explained by the mental state is expresses?

I think the following is a plausible answer. Many speech acts have conditions under which their performance counts as sincere. Let us call those conditions a speech act's *sincerity conditions*.[32] Sincerity conditions typically or always mention mental states. For instance, assertions are sincere only if the asserter holds a specific belief. Likewise, promises are sincere only if the promisor holds a specific intention (e.g. the intention to do what she promises). And requests are sincere, only if one wants what one requests and so on. This leads to the following simple proposal.

Sincerity Expressing: A sentence F expresses M iff F is conventionally used to perform speech act A where A is sincere only if the performer is in M.

The idea is that every sentence of a given language, as a matter of conventions governing the use of that language, is associated with a speech act having a certain mental state as its sincerity conditions. This mental state is the mental state expressed by this sentence.

The idea to define expressing in terms of sincerity is not completely new, but goes back to a proposal which John Searle already made in his 1969 book 'Speech Acts'. There Searle suggests the following:

> Wherever there is a psychological state specified in the sincerity condition [of an act], the performance of the act counts as an *expression* of that psychological state. This law holds whether the act is sincere or insincere, that is whether the speaker actually has the specified psychological state or not. Thus to assert, affirm, state (that p) counts as an *expression of belief* (that p). To request, ask, order, entreat, enjoin, pray, or command (that A be done) counts as an *expression of a wish or desire* (that A be done). To promise, vow, threaten or pledge (that A) counts

---

[31]See for instance (Hare, 1952, ch. 2).

[32]The notion of a speech act's sincerity condition was introduced by Searle in (Searle, 1969, 60ff).

> as an *expression of intention* (to do *A*). To thank, welcome or
> congratulate counts as an *expression of gratitude, pleasure* (at
> *H*'s arrival), or *pleasure* (at *H*'s good fortune).[33] (Searle, 1969,
> 65)

It will be helpful to do a brief excursion into Searle's classic of speech
act theory, in order to see what it has to tell us about speaking a language,
linguistic conventions, constitutive rules and performing acts by obeying
rules. This will help us to better understand the expressing relation in terms
of sincerity conditions and what consequences this has for the overall shape
of an expressivist semantics about which I will have to say more in the next
section.

Searle's hypothesis in his book on speech acts is this:

> that the semantic structure of a language may be regarded as
> a conventional realization of a series of sets of underlying con-
> stitutive rules, and that speech acts are acts characteristically
> performed by uttering expressions in accordance with these sets
> of constitutive rules. One of the aims [of Searle's book] is to
> formulate sets of constitutive rules for the performances of certain
> kinds of speech acts. (Searle, 1969, 37)

Two distinctions are important in order to understand the proposal. One
is the distinction between conventions and rules. The other is the distinction
between constitutive and regulative rules. Conventions, for Searle, are
language-specific ways of realizing underlying general rules. He elaborates:

> The fact that in French one can make a promise by saying 'je
> promets' and in English one can make it by saying 'I promise'
> is a matter of convention. But the fact that an utterance of a
> promising device [e. g. a suitable sentence] (under appropriate
> conditions) counts as the undertaking of an obligation is a matter
> of rules and not a matter of the conventions of French or English.
> (Searle, 1969, 40)

Conventions are arbitrary in the sense that a completely different ex-
pression could have been used in French to make a promise. However, what
could not be different, according to Searle, is that one counts as making
a promise only if one undertakes an obligation. This, according to Searle,
is not a matter of conventions, but rather a matter of a constitutive rule
governing the speech act of promising.

Constitutive rules, other than regulative rules, it is well known, make
certain acts possible that would not be possible if the rules did not exist.

---

[33]Strictly speaking, this definition tells us what it is for a *speaker* or for a *speech act* to
express a mental state. The above definition of semantic expression ('Sincerity expressing')
is received by simply building on what we already know from Davis.

Searle's goal is to uncover the general constitutive rules for various speech acts, since he thinks that knowing what they are will help us with understanding how languages come to have the 'semantic structure' that they have.

So Searle's conception is roughly the following. Each expression or sentence of some particular language like English is associated with a specific convention telling us which speech act one typically, or in general, makes if one utters (or more generally: if one makes intended use of) that expression.[34] These conventions realize underlying language-independent constitutive rules for speech acts. What are those relevant constitutive rules?

In order to find out what the relevant constitutive rules are that underlie the conventions of particular languages, Searle offers analyses for different types of illocutionary acts such as promising, ordering, asserting etc., by providing for each type of speech act a set of individually necessary and jointly sufficient conditions. One important necessary condition which, according to Searle, most illocutionary acts have, is precisely that they have a *sincerity condition*. Moreover, Searle thinks that whether a speech act is sincere depends on the mental states of the speaker.

Moving on from his analyses of speech acts, he attempts to "extract from our set of conditions a set of rules for the use" of the linguistic expressions by which we perform those speech acts (Searle, 1969, 62). Searle offers a number of such rules. In the current context those rules are of particular interest which he claims to be "derived" (Searle, 1969, 63) from the sincerity condition.

In the case of assertion, Searle takes the semantical rule to be that for the use of any 'illocutionary force indicating device' F of assertion: F is to be uttered (used as intended) only if S holds some specific belief.[35] Searle thinks of this rule as a constitutive rule for assertion because it partially defines what it is for a speech act to count as an assertion: a speech act that can sincerely be performed without the speaker holding the relevant belief cannot be an assertion.

Searle thinks that rules such as the one just given for assertion are in fact instances of even more "general underlying rules for illocutionary acts" (Searle, 1969, 63). What are the even more general rules underlying rules? Unfortunately, Searle does not tell us, but in fact the answer seems pretty straightforward at least for the rule 'derived' from the idea that most speech acts have sincerity conditions. A general constitutive rule underlying the semantic conventions of a language seems to be what one could call the

---

[34] Searle is, of course, concerned with natural languages that are spoken languages, that is languages their intended use is to be uttered. But it is easy to imagine languages the intended use of the relevant linguistic items is not uttering, but some other form of usage. I return to this topic in chapter 4.

[35] Cf. (Searle, 1969, 63) modified to assertion.

Maxim of Sincerity: One must utter (more generally: make intended use of)
an expression only if one satisfies the sincerity condition of the speech
act that is conventionally performed by someone who utters (uses) this
expression. In short: be sincere![36]

The maxim of sincerity might be regarded as formulating one of the most
general constitutive rules underlying the semantic conventions of natural
languages. Explicating the expressing relation in terms of sincerity conditions
therefore places the expression of mental states at the heart of Searle's
project to explain the semantic structure of languages. This makes this
notion suitable for the expressivists project of explaining the meaning of
sentences in terms of the mental states they express, and connects it back to
its historic routes in ordinary language philosophy and speech act theory.

Moreover, the explication of expressing in terms of sincerity has all
the properties expressivists need it to have: it is compatible with their
commitment to mentalism since mental states play a central role in it. It
also allows for insincerity because sentences that are conventionally used
to perform speech acts that have a sincerity condition, can easily be used
without satisfying the condition. Moreover, the fact that ultimately what
sentences express is determined by what they are conventionally used to
express by speakers does not make semantic expressing speaker relative
because, again, what a sentence expresses is a matter of conventions and not
of the intentions of individual speakers or their use on specific occasions. Just
because Humpty Dumpty wants 'This is glory' to express the belief that this
is a nice knock-down argument, this will not let this sentence semantically
express this belief.[37] Furthermore, this account is not unstable because it does
not assume that sincerity conditions are some kind of information that gets
communicated; they are a useful a device of the semantic theorist. Finally,
the sincerity account of expressing is a very general account of expressing
and applies not only to declarative sentences, but also to sentences that are
used to perform speech acts other than assertions such as interrogatives that
are used to ask questions.

---

[36]I call this a 'maxim' in allusion to Grice's famous conversational maxims. This maxim
or norm of sincerity might perhaps be regarded as the more general version of Grice's
so-called 'maxim of quality' which includes the prescription to not say what you believe
to be false. Levinson suggested but did not spell out a generalization of the maxim of
quality into a maxim of sincerity (Levinson, 1983, 103ff). See (Braun, 2011) for an actual
formulation of such a maxim. Martinich is skeptical about such moves. He believes that to
'be sincere' is no conversational "supermaxim" because not all speech acts have sincerity
conditions (Martinich, 1980, 219f.). However, if some speech act lack sincerity conditions,
which is plausible, that does not rule out that one must be sincere, in all cases where it is
possible to be sincere.

[37]Compare the dialog between Alice and Humpty Dumpty in Lewis Carroll's classic
'Alice' Adventures in Wonderland' (Carroll, 1936/1871, 213-214). For more on the Humpty
Dumpty theory of meaning see the exchange between (MacKay, 1968) and (Donnellan,
1968).

2.5.1   The Shape of Expressivist Semantics

In this chapter, I have discussed a number of different explications of the expressing relation. This helps us understand better what an expressivist semantics looks like. The task of an expressivist semantics is to systematically assign mental states to sentences, namely the states those sentences *express*. For simplicity let us ignore for the moment that we can form complex sentences and consider a language containing only atomic formulas $F_1,...F_n$.[38] We can think of an expressivist semantics for such a language as being a list like the following:

(1) F_1 expresses M_1,

(2) F_2 expresses M_2,

(i) F_i expresses M_i,...

   Each line of such a list for a given language can be regarded as what Searle referred to as a 'convention' of that language. Different expressivist languages differ in what they write on the left-hand side of each line. For instance, as a matter of linguistic convention, in English the belief that grass is green is expressed by the sentence 'grass is green', whereas as a matter of convention, in German it is expressed by the sentence 'Gras ist grün', and in French by 'L'herbe est verte'.

   Each line of such a list will have a different metasemantic interpretation, depending on what we mean by 'express'. According to assertability expressivism, for instance, the first line says that, as a matter of the conventions of some language $\mathcal{L}$, it is semantically permissible to assert F_1, only if you are in M_1. According to indicatory expressivism, it says that, as a matter of convention in $\mathcal{L}$, F_1 is used by speakers with the intention to indicate that they are in M_1. And sincerity expressivism will say that, as a matter of convention of $\mathcal{L}$, F_1 has sincerity condition M_1, or more precisely, is used by speakers to perform a speech act which is sincere only if they are in M_1.

   Spelled out in slightly more detail, in the case of sincerity expressivism each line (convention) in the expressivists' semantics can be understood as saying something like the following:

- If a speaker S makes intended use U (e. g. utter, write...) of F_i, then S conventionally and typically counts as performing speech act A_i (which is of the assertion-, command-, question-type...), and S's speech act A_i is sincere, only if S is in M_i.

---

[38]I will turn to the task of providing a compositional semantics for a language that contains logically complex sentences in the next chapter. Here I just want to give a general understanding of what an expressivist semantics is.

In addition, as the general constitutive rule underlying all this, the sincerity expressivist's semantics will simply state the above maxim of sincerity, which demands that speech acts be sincere.[39]

In this chapter, I critically discussed several ways to explicate the expressing relation. We have found that some of the most natural understandings of it, in particular the same content and the causal account, are most likely inadequate for the purposes of the expressivist. Other accounts such as Gibbard's indicatory expressivism look more promising, though they are also subject to problems. I have argued, however, that by borrowing key ideas from Davis, a refined version of indicatory expressivism might avoid the problems with speaker-relativity and instability. I also discussed Schroeder's proposal to explicate expressing in terms of assertability. Finally, I presented an ignored but simple explication of expressing in terms of sincerity conditions of speech acts.

In my view, refined indicatory expressivism, assertability expressivism, as well as what I have called sincerity expressivism all provide explications of the expressing relation that are adequate for the overall purposes of expressivism. Even though all three views have a different take on expressing, they satisfy our key constraints: the state expressed by a sentence does not vary with subjective speaker intention (constancy), allows for insincerity (intentionality), and is compatible with mentalism (metasemantic compatibility). With respect to the generalizability constraint sincerity expressivism seems to be the best option.

All the aforementioned properties are crucial for the expressivists' project of explaining the meaning of sentences in terms of the mental states they express. This is so because these properties are required to allow that the mental state that is expressed by a complex sentence is a function of the mental state expressed by its parts. The indicatory expressivist thinks that which mental state speakers intend to indicate they are in by using complex sentences is a function from the mental states speakers intend to indicate they are in by using its parts. Likewise, the assertability expressivist thinks of these as functions from assertability conditions to assertability conditions, and the sincerity expressivist thinks of them as functions from sincerity conditions to sincerity conditions. All three explications allow that what mental state or attitude is expressed by a logically complex sentence is a function of the mental states or attitudes expressed by its parts, where the function is determined by the meaning of the logical connectives that are used to form the sentence. This will be important in the next chapter because

---

[39]Schroeder notes that in the case of assertability expressivism it might not be possible to state a single unified norm of assertion as a general rule underlying the semantics because this conflicts with the expressivists mentalism and noncognitivism. He thinks that each sentence is semantically associated with an *individual* norm of assertion stating when it is permissible to assert that sentence (M. Schroeder, 2008a, 31). If this is in fact true, this is a further reason to prefer sincerity expressivism over Schroeder's assertibility expressivism.

there I will present a novel solution to the expressivist's most notorious problem in semantics: the Frege-Geach problem. This problem will be approached in terms of such functions from expressed attitudes to expressed attitudes, or as I shall call them 'attitude functions'.

Chapter 3

Attitude Semantics and Logic

This chapter offers a version of what Nicholas Unwin calls a "unified and spectacularly simple solution to the Frege-Geach problem" (Unwin, 1999, 350). To this end I develop *attitude semantics*: a formal expressivist semantics and logic for a first-order language. It is shown that attitude semantics has all the merits, and none of the demerits of Mark Schroeder's much noticed 'biforcated attitude semantics' (M. Schroeder, 2008a). Finally, a new form of expressivism is suggested, *dispositional expressivism*, as a philosophically attractive interpretation of the mental states attitude semantics assigns to sentences.

3.1 Introduction

The most notorious problem for expressivism is the Frege-Geach problem. In his influential paper (Unwin, 1999) Nicholas Unwin considers the possibility of a "unified and spectacularly simple solution to the Frege-Geach problem".[1] The "very attractive, even obvious" solution, he says, consists in a particular assumption about the *structure* of the mental states expressed by moral sentences and their negations. But, after raising some problems with generalizing his proposal, Unwin draws the conclusion that "Alas, it does not work!"

In his (M. Schroeder, 2008c) and his book *Being For* (M. Schroeder, 2008a) Mark Schroeder takes up Unwin's proposal, and shows that expressivists can after all work it into a formally adequate and constructive solution[2] to the Frege-Geach problem if, and *only* if, in addition to (i) Unwin's structural proposal (see E' below), they assume (ii) global expressivism, (iii)

---

[1]The citations in this paragraph are all from Unwin (Unwin, 1999, 349f.).

[2]For the notions of *formal adequacy* and *constructiveness* see (M. Schroeder, 2010b, 227-229).

global noncognitivism, (iv) biforcation, and (v) A-type inconsistency. *Global expressivism* is the view that the meaning of *all* sentences is explained by the mental states they express. *Global noncognitivism* is the view that the states expressed by all sentences are ultimately noncognitive states of mind. *Biforcation*[3] is the idea that all sentences express 'biforcated' attitudes which are pairs of noncognitive states. On Schroeder's proposal these biforcated attitudes are pairs of the noncognitive state of *being for* with the restriction that the content of at least one state entails the content of the other. *A-type inconsistency* is a particular kind of rational incoherence, namely the one that holds between mental states iff those states are rationally incoherent, if they are states of the same type, and their contents are inconsistent. Schroeder stipulates the state of being for is capable of A-type inconsistency, and so it is rationally incoherent to be for some content and also to be for its negation. On the basis of these assumptions Schroeder develops one of the most sophisticated and comprehensive expressivist theories to date: *biforcated attitude semantics*.

In this chapter, I will not debate Schroeder's claim that (i)-(v) are sufficient for a solution to the Frege-Geach problem. But I question his claim that they are all *necessary*. The solution presented in this chapter does not require (iii), (iv), and (v), that is it does not require global noncognitivism, A-type inconsistency, and, most importantly, biforcation.[4]

The solution presented in this chapter consists in a formal expressivist semantics for a first-order language with a logic the inconsistencies, truths, and validities of which are coextensive with the inconsistencies, truths, and validities of classical first-order logic. I call this semantics *attitude semantics*. One of the major benefits of attitude semantics, as the name suggests, is that it does not require the somewhat *ad hoc* postulation of biforcated attitudes.[5] Attitude semantics avoids all the problems which biforcation causes, such as the *new new negation problem*, the need for a three-valued 'acceptance logic', the problems with explaining truth-conditions, and perhaps some of the further semantic problems Schroeder lists, such as the problem with binary quantifiers like 'most'. Some of these problems will be treated below.[6]

---

[3]'Biforcation' is not a typo, but supposed to be a pun on 'being for' and 'bifurcation'.

[4]It is not of central importance that the new theory does not require A-type inconsistency (v) because it still requires some notion of rational incoherence (see Sect. 3.5.2 below). It does also not use what Schroeder calls 'B-type inconsistency': a pair of states is B-type inconsistent iff they are rationally incoherent, if they are attitudes of different kinds held towards the same content, such as perhaps disapproval of something and tolerance of the same thing. For more on B-type inconsistency consult (M. Schroeder, 2008a, 48).

[5]In reply to Mark van Roojen, Schroeder tries to mitigate the *ad hoc*-ness worry, by suggesting that biforcation is "motivated by its fruits" (M. Schroeder, 2008a, 98, FN. 5). But if the 'fruits' (i. e. a solution to the Frege-Geach problem, a way to explain the truth-conditions of descriptive sentences, and an account of disbelief...) can be had without biforcation, this reinforces the worry.

[6]For a more detailed discussion of these problems consult Schroeder (M. Schroeder, 2008a, Ch. 8, 9, 10, 12).

Moreover, attitude semantics is also able to capture what Schroeder takes to be an attractive feature of biforcation, namely that it allows expressivists to define a state of *disbelief*.

The central idea underlying Schroeder's biforcated attitude semantics is that a moral sentence like 'murder is not wrong' expresses the noncognitive state of being for not blaming for murder.[7] The simple idea underlying the attitude semantics presented in this chapter is that this sentence instead expresses something like the attitude of *being for being for* not blaming for murder. I say 'something like' because I am interested here primarily in a *structural* solution to the Frege-Geach problem, and only secondarily in some particular interpretation of this structure. I will show how this simple idea can be developed into a formal expressivist semantics and logic that avoids the problems that confront Schroeder's expressivism.

The chapter is structured as follows. In section 3.2, *Compositionality and Logicality*, I make some general remarks on expressivism and the Frege-Geach problem. In the next section, I discuss Nicholas Unwin's *Negation Problem*. In the subsequent section I consider the *Unwinian Proposal*, and Schroeder's development of it which culminates in his 'biforcated attitude semantics'. Section 3.5 presents the heart of the new theory: *Attitude Semantics and Logic*. In the subsequent section, *Comparison*, it will be shown that in attitude semantics it is easier to explain the truth-conditions of descriptive sentences and to define a notion of disbelief than it is in biforcated attitude semantics. In the final section, I suggest that the state expressed by moral sentences rather than being an iterated state of being for being for is more plausibly interpreted as a mental *disposition* to be in some noncognitive state. This *dispositional expressivism* avoids global noncognitivism, and so the radical consequence of Schroeder's theory that all states expressed by sentences are, at bottom, noncognitive mental states.

3.2 Compositionality and Logicality

The main aim of this chapter is to provide a formally adequate and constructive solution to the expressivist's Frege-Geach problem. Expressivism, recall, is the view that explains the meaning of sentences by the mental states they are conventionally used to express.[8] Expressivism, I said, is a *mentalistic* or *psychologistic* semantics: its primary task is to assign mental states to sentences, namely the mental states they express.[9]

---

[7]Actually, in order to qualify as a biforcated attitude, Schroeder suggests that it either expresses this state twice, or it expresses being for not blaming for murder and being for not disliking murder. See (M. Schroeder, 2008a, 113ff.).

[8]In what follows I will often drop the qualifier "conventionally used to". In the following I will presuppose the understanding of expressing in terms of sincerity conditions which I have advocated in the previous chapter, though, as I have argued, expressivism should be compatible with various other explications of expressing as well.

[9]Compare (Rosen, 1998, 387).

What is the Frege-Geach problem? In a nutshell, it is the problem of explaining how meaning can be compositional if metaethical expressivism is true. I will understand the Frege-Geach problem as a combination of two problems: what I call the 'compositionality problem' and the 'logicality problem'. I will start by describing those two problems in turn.

It is widely accepted that any adequate semantics has to account for the fact that meaning is compositional. In other words, any adequate semantics has to give the meaning of a complex expression in terms of a recursive function of the meaning of its parts together with its mode of composition (i. e. its logico-syntactic structure).[10] Since expressivists do not explain meaning, as it is orthodox in semantic theorizing, in terms of truth-conditions, the meaning of a complex sentence cannot be modeled as a function from truth-conditions to truth-conditions either. Since expressivists instead explain the meaning of a sentence by the mental state it expresses, they have to explain the meaning of a complex sentence in terms of recursive functions from the mental states expressed by the parts of the sentence (together with its mode of composition) to the mental states expressed by the complex sentence. I will call the task of providing such attitude-functions for some language the *compositionality problem*.

Since one way of compositional embedding is embedding under logical connectives, we have to ask what the meaning of a logical connective is. In truth-conditional semantics, it is modeled as a recursive function from truth-values to truth-values. Standardly, the meaning of the logical connective 'not', for instance, is a recursive function that 'reverses' the truth-value of the sentence to which 'not' is applied to its opposite value, i. e. it is a function that maps the value *true* to *false*, and *vice versa*. For the expressivist, in contrast, the meaning of a logical connective has to be given by an attitude-function: a function from the mental states expressed by sentences to the mental states expressed by sentences that are formed by means of the relevant connective. In the case of 'not' this has to be a recursive function that 'reverses' the mental state expressed by the sentence to which 'not' is applied to its 'opposite' or 'complementary' mental state. Only if expressivism can provide such attitude-functions for all logical connectives as well as all the non-logical expressions of a language, they can claim to have a provided a compositional theory of the meaning of the sentences of the language.

Besides the semantic properties of sentences, expressivists also have to account for the *logical properties* of sentences. Since the logical properties of

---

[10]The characterization of expressivism given above focuses primarily on the meaning of whole sentences, but the meanings of subsentential expressions can and will as standardly be understood in terms of the semantic contribution they make to the meaning of whole sentences. In the expressivist context this means we will understand their semantic contribution as the contribution they make to (the content of) the mental states they express.

sentences (e. g. that every sentence is logically inconsistent with its negation) depend on the semantic properties of sentences (i. e. on the meaning of the logical vocabulary), and expressivism explains the latter in terms of the mental states they express, expressivism attempts to explain the logical properties of sentences in terms of some properties of the mental states they express. How can expressivists do that?

Expressivists' general idea about logic starts from the observation that one is in a rationally incoherent state of mind if one believes what inconsistent sentences say (i. e. one is in the states expressed by inconsistent sentences). The hope is that this conditional relationship is robust, and holds in the other direction as well. If expressivists can establish the equivalence of the inconsistency between sentences and the rational incoherence between the states they express, then they can treat inconsistency as being definable in terms of rational incoherence. So expressivists aim to establish something like the following as true:

Logical Inconsistency A set of sentences is logically inconsistent iff the syntactic form of the sentences together with the meaning of the logical vocabulary appearing in them guarantees that the set of mental states they express is rationally incoherent.

Once the expressivist has a notion of logical inconsistency, she can use it to define logical truth, entailment and validity in terms of it. So expressivism attempts to explain the logical properties of sentences in terms of rational properties of the states they express. I will call the task of providing such an explanation the *logicality problem*. On my understanding, solving the *Frege-Geach* or embedding problem amounts to solving both the compositionality problem and the logicality problem. I start with the compositionality problem, especially with what appears to be its simplest instance: the compositionality of negation.

3.3 The Negation Problem

Expressivists attempt to provide the meanings of the logical connectives in terms of attitude-functions, that is functions from the states expressed by component sentences to the states expressed by complex sentences. Concerning the unary logical connective 'not' the question is this:

Compositionality of Negation For any sentence A: If a sentence A expresses the mental state M, which state N is expressed by $\neg A$?

In order to answer this question, expressivists need an account of which attitudes are expressed by sentences of their language, and they must provide an answer to what the attitudinal analogue of the truth-functional 'reversing'

is. Since A can be a formula of arbitrary logical complexity, we will not be able to say which attitude is expressed by it, unless we know the attitude-functions for all the other connectives which may appear in the formula. Thus, in order to get a foothold, expressivists need to begin by giving an account of *atomic* sentences and *their* negations. But as Nicholas Unwin has pointed out, even this turns out to be trickier for expressivists than one would expect.

According to Nicholas Unwin (1999), something like the following equivalence schema is implicit in most expressivist theories:

(E)   A speaker S accepts 't is wrong' iff S is ϕ-ing t.[11]

As he understands this schema, the mental state ('ϕ-ing') attributed to S on the right-hand side is the state *expressed* by the sentence accepted by S on the left-hand side. In other words: the mental state expressed by 't is wrong' is the ϕ-ing of t. So (E) can also be written as follows:

(E)   't is wrong'   expresses $\phi(t)$

Note that in this form it is clear that what Unwin calls an 'equivalence' is what we in the previous chapter, following Searle, called a 'convention' about a particular language. Unwin observes that the equivalence schema suffers from the "fundamental syntactic defect" (Unwin, 1999, 341) that there are three ways to insert a negation sign to the left-hand side, but only two ways to insert a negation sign to the right-hand side of the equivalence. Take as an example the sentence 'murder is wrong', and for illustrative purposes replace 'ϕ-ing' with 'blaming for', and 't' with 'murder'.[12] Then we get the following options:

---

[11] I use 't' here as a variable for singular terms. Unwin's original schema is formulated differently. He says that a speaker S accepts '$H!p$' iff S horrays that p (Unwin, 1999, 339). This formulation differs from the one given here in several respects. First, the sentence on the left-hand side ('$H!p$') is a formula of Blackburn's emotivist language E_{ex} (see (Blackburn, 1984b, 193ff)). In Blackburn's language '$H!$' is treated as a sentential operator akin to the deontic operator for 'ought' added to a propositional language. The formulation given here focuses on sentences of first-order languages instead. Another difference, though inessential, is that Unwin focuses on positive evaluations ('ought'/'good') whereas I use 'wrong' as the main example. Finally, Unwin's formulation is less abstract than the one given here, since he does not use a variable ('ϕ-ing') on the right-hand side, but the specific intentional verb 'hooraying'. This suggests that we should treat 'ϕ-ing' as a placeholder for intentional verbs denoting *mental* states, or acts, or relations, and this is how I will primarily understand it here. However, below I will allow that 'ϕ-ing' can be replaced even with non-mental properties such as, for instance, the property of being green.

[12] It is not essential to the account given here that 'ϕ-ing' is replaced with the relation of 'blaming for' in the case of 'wrong'; I only do so to achieve terminological continuity with Schroeder's theory to which I come in a moment. It could equally well be replaced by 'booing', 'disapproving', or something similar.

(E) S accepts 'murder is wrong' iff S is blaming for murder.

(N1) S does not accept 'murder is wrong' iff S is not blaming for murder.

(N2) S accepts 'murder is not wrong' iff *???*

(N3) S accepts 'not murder is wrong' iff S is blaming for not murder.

The negation problem is that the equivalence schema (E) does not allow us to give a plausible interpretation of (N2) by placing a 'not' somewhere into the right-hand side of the equivalence. Here is how Unwin describes the problem:

> (N1) follows from (E) by contraposition, and (N3) is a substitution-instance of (E). However, there can be no similar analysis of (N2), since the sentence ['S is blaming for murder'] does not admit of an intermediate form between external and internal negation. [...] A mere absence of attitude is evidently not enough, since we shall otherwise fail to distinguish (N1) from (N2). (Unwin, 1999, 342)

This situation is unfortunate, since the case we are interested in, the negation case (N2), is precisely the problematic one. Since the only thing we have done with (N2) is that we added in a negation sign on the left-hand side, it is natural to expect a negation sign to appear on the other side of the equivalence as well.[13] The problem is that the (E) does not allow the expressivist to fill out the right-hand side by means of negation, in such a way that we receive a result which is plausible and also different from the right-hand sides of (N1) and (N3). So (E) does not allow the expressivist to answer the question of the compositionality of negation.

3.4   The Unwinian Proposal

Unwin's diagnosis of the negation problem is that it results from a lack of structure in the right-hand side of the expressivist's equivalence schema (E). He asks "where else can the structure come from?" (Unwin, 1999, 344). A "very attractive, even obvious" proposal Unwin makes, is to correct the 'defect' in the equivalence schema by adding the missing structure to the right-hand side. He suggests the following:

---

[13]This expectation might be a natural one, but it is not the only option. The theory of Blackburn allows for readings according to which negated sentences are not used to perform assertions. Instead, on one occasion he says that "negation is expressive of denial", where to deny something is a speech act in its own right, indefinable in terms of assertion and negation (Blackburn, 1988, 511). For accounts along similar lines see Wedgwood (2010), and the bilateral treatments of Cantwell (2013), and Cantwell (2015).

(E') S accepts 't is wrong' iff S is ψ-ing ϕ-ing t.[14]

Or again in terms expressing mental states:

(E')   't is wrong'   expresses $\psi(\phi(t))$

Mark Schroeder agrees with Unwin's diagnosis that if the "problem arises from a lack of structure, there can be only one solution: to add structure" (M. Schroeder, 2008a, 61). Here is Schroeder's particular filling-out of Unwin's proposal:

> The solution is to say, just as all descriptive predicates correspond to belief plus some property that is contributed by the predicate, that all normative predicates correspond to being for plus some relation that is contributed by the predicate. For each predicate, F, there is some relation, R_F, so that '$F(t)$' expresses FOR(bearing R_F to t). So, for example, [...] we might say that 'wrong' corresponds to *being for blaming for*, so that 'murder is wrong' expresses FOR(blaming for murder).[15] (M. Schroeder, 2008a, 58, italics original, variables adapted)

In terms of Unwin's refined equivalence schema (E'), Schroeder suggests for atomic 'wrong'-sentences to replace 'ψ-ing' with the noncognitive attitude of 'being for' and 'ϕ-ing' with the relation of 'blaming for', which yields:

    S accepts 't is wrong' iff S is being for blaming for t.

In terms of expressing and with Schroeder's semi-formal way of denoting mental states:

    't is wrong'   expresses   FOR(blaming(t))

Generalizing from this, he suggests the following semantic clause for atomic sentences:

---

[14]Again, this formulation differs from Unwin's original one which is: A speaker S accepts '$H!p$' iff S desires that one aims at p. See Unwin (Unwin, 1999, 349).

[15]Schroeder uses small capital letters to denote mental states. I will follow him in using this convention. 'FOR(α)' is thus a term denoting the mental state of being for α, where α is a schematic letter for metalanguage formulas describing the content of a state of being for. I will also follow Schroeder in treating the αs as the *contents* of the states of being for, and also as the *semantic values* of sentences (see the definition of atoms below).

atoms If F is an n-place predicate, then $F(t_1, \ldots t_n)$ is a well-formed formula, and corresponds to an $n+1$-place relation, $R^F(z, t_1, \ldots t_n)$, so that if $t_1, \ldots t_n$ are singular terms denoting $o_1, \ldots o_n$, then $F(t_1, \ldots t_n)$ expresses $\text{FOR}\big(\lambda z.(R^F(z, o_1, \ldots o_n))\big)$. We say that $\lambda z.\big(R^F(z, t_1, \ldots t_n)\big)$ is the *semantic value* of $F(t_1, \ldots t_n)$.[16]

How does the structural refinement (E') help with solving the negation problem? Unwin suggests that now that we have enough structure on the right-hand side, the negation sign should be placed in front of the 'ϕ-ing' so that we get:

(N2') S accepts 't is not wrong' iff S is ψ-ing *not* ϕ-ing t.

On Schroeder's interpretation this reads:

(N2') S accepts 't is not wrong' iff S is being for *not* blaming for t.

Or alternatively:

(N2')   't is not wrong'   expresses $\text{FOR}(not \text{ blaming}(t))$

According to Unwin and Schroeder, such an interpretation of the refined equivalence scheme (E') is "not too implausible" (Unwin, 1999, 349), and it does not conflate (N1)-(N3).

Hence, it seems that the expressivist can use the new structural proposal as a blueprint for a rule for negation that avoids the negation problem. Such a rule will say that if A expresses ψ-ing ϕ-ing t, then $\neg A$ expresses ψ-ing $\neg\phi$-ing t. More formally:

$\neg$ If $[\![A]\!] = \lambda z.\alpha$, then $[\![\neg A]\!] = \lambda z.\neg\alpha$.[17]

If this semantic clause is correct, we have an answer to the above question of the compositionality of negation: what the attitude-functional connective 'not' does is that it negates the semantic values of sentences (i. e. the contents of the states expressed). This is its 'reversing' role according to this proposal.

Furthermore, Unwin notes, the proposal has the benefit of fitting the expressivist's conception of logicality, since it allows a straightforward explanation of logical inconsistency in terms of rational incoherence of the states expressed. He explains:

---

[16]Compare (M. Schroeder, 2008a, 78, variables adapted). Schroeder introduces the device of lambda-abstractions in order to describe the properties which are the contents of states of being for in a more precise way than by using gerundival phrases such as 'blaming for'. So '$\lambda z.(z$ is blaming for $o)$' is a term for the property someone has when he is blaming for o.

[17]This is Schroeder's non-biforcated clause for negation. Cf. (M. Schroeder, 2008a, 78). I use '$[\![A]\!]$' to denote the semantic value of A, that is the content of the state of being for that is expressed by some sentence.

> The source of the incompatibility of ['*t* is wrong' and '*t* is not wrong'] is now very clear; someone who accepts both sentences thereby [is for blaming *t* and for not blaming for *t*], and this is clearly irrational. True, it is not strictly speaking *illogical*, since it is a case of inconsistent desires [or states of being for] rather than of inconsistent beliefs, but it is nevertheless an excellent candidate for 'quasi-inconsistency' of the sort that Blackburn requires. (Unwin, 1999, 349)

This all looks as if Unwin has made some major progress with the Frege-Geach problem. If this proposal could successfully be generalized to other logical contexts, and thereby solve the problems of compositionality and logicality, then, Unwin suggests, it would constitute "a uniform and spectacularly simple solution to the Frege-Geach problem" (Unwin, 1999, 350).

But: "Alas", says Unwin, "it does not work!" (Unwin, 1999, 350) because in his view a generalization would require that the first intentional verb on the right-hand side of the refined equivalence schema ('ψ-ing') is replaced either by something like 'believes' or by something like 'desires' (or in Schroeder's interpretation: 'being for'), since otherwise it is hard to see what states we should take to be expressed by *mixed* sentences—sentences that contain both moral and descriptive vocabulary. This means that either all sentences would have to express belief-like states, or that all sentences would have to express desire-like states (e. g. being for).

This leads to a dilemma for the expressivist: either moral sentences express cognitive, belief-like states, or descriptive sentences express noncognitive, desire-like states. Even though Unwin explicitly rejects only the belief-like horn of the dilemma because "then the theory will cease to be expressivist [i. e. noncognitivist]", he seems to think that the desire-like horn is equally untenable. So he concludes that he has "not succeeded in integrating the most naturally revised proposal for negation" (Unwin, 1999, 352).

3.4.1 The New Negation Problem

Schroeder thinks that Unwin is too quick with the rejection of his simple solution. He suggests that the desire-like-option might actually be a bullet worth biting for the expressivist. His proposal is that expressivists should seriously consider the assumption that *all* indicative sentences express the *same* type of noncognitive, desire-like mental state. This implies that if all moral sentences express the desire-like state of, for instance, being for, then all *descriptive* sentences have to be treated as expressing being for as well. In other words, Schroeder proposes that expressivists adopt *global* noncognitivism, i. e. noncognitivism about moral as well as descriptive sentences.

This is a very radical idea as Schroeder himself concedes.[18] Is there
any way to defend it? Schroeder suggests there is. He proposes that in the
descriptive case 'ψ-ing' has to be replaced with 'being for' just as before,
and that 'ϕ-ing' has to be replaced with the relation of 'proceeding-as-
if' (henceforth: 'pai'). Schroeder does not tell us much about the nature
of proceeding-as-if. He only says that it could also be described as the
relation of 'treating-as-settled-in-deciding-what-to-do'.[19] This brings it pretty
close to 'holding-true' which other philosophers would perhaps identify with
'believing'. To illustrate his proposal, he offers the following equivalence for
a descriptive sentence like 'grass is green':

> S accepts 'grass is green' iff S is being for pai grass is green.

The move to global noncognitivism is the crucial step towards solving the
problem of mixed sentences, since now we do not have to ponder about the
question of which state could possibly be expressed by a mixed conjunction
like 'grass is green and murder is wrong' if the first conjunct expressed a
belief-like state, and the second one expressed a desire-like state. Now the
answer is straightforward: conjunctions express states of being for, simply
because all sentences express states of being for.

But there is a problem with this proposal as well. Even if we would
be willing to accept the radical suggestion that 'grass is green' expresses
the noncognitive state of being for proceeding as if grass is green, in order
to solve the problem of mixed sentences, Schroeder's expressivist runs into
what Schroeder calls the 'new negation problem'. The new problem is that
if the expressivist wants her just found semantic clause for negation ($\neg$
above) to apply uniformly (i. e. to moral and descriptive sentences alike),
she will receive false predictions. Recall that in order to solve the negation
problem for moral sentences, Unwin proposed that we place a negation sign
in front of the ϕ-ing, which under the present interpretation for a descriptive
predicate such as 'green' is the relation of proceeding as if. So, the rule
which expressivists need in order to solve the negation problem for moral
sentences, yields the following prediction for a negated descriptive sentence
such as 'grass is not green':

> S accepts 'grass is not green' iff S is being for *not* pai grass is green.

But this seems false. Intuitively, 'grass is not green' should not express
being for not proceeding as if grass is green (this state appears to be too
weak, in some sense).[20] If descriptive sentences express anything like this at
all, it should rather express being for proceeding as if grass is *not* green.[21]

---

[18](M. Schroeder, 2008a, 142)

[19](M. Schroeder, 2008a, 93)

[20]I return to this weakness in the discussion of disbelief in section 3.6.2 below.

[21]Confronted with the new negation problem expressivists might wonder how much of a

Is there a way to modify the semantic clauses for atomic sentences or for negation to yield correct predictions in the moral as well as in the descriptive case? Schroeder thinks that there is such a way, but it requires making a rather *ad hoc* assumption, namely that sentences express *pairs* of states of being for. I explain his 'biforcation'-move in more detail in the next section.

3.4.2 Schroeder's Biforcation Solution

Schroeder thinks that expressivists need not despair in the face of the new negation problem. In order to solve it, he devises a "trick" (M. Schroeder, 2008a, 97). The trick is to treat the states expressed by all indicative sentences as what he calls 'biforcated attitudes'. A biforcated attitude is an attitude consisting of a pair of two states of being for, such that the content of at least one of the states entails the content of the other. For instance, a descriptive sentence P, on the biforcated analysis, expresses not just being for one relation to p, but being for two relations to it. More precisely, Schroeder suggests that

$$P \qquad \text{expresses} \quad \text{FOR}(\text{pai}(p)) \text{ and } \text{FOR}(\neg\text{pai}\neg(p)).$$

Biforcation is required only in order to solve the new negation problem that arises for descriptive sentences. But since the new semantics requires that all sentences expresses biforcated attitudes, moral sentences have to express pairs of states of being for as well, e. g. 'murder is wrong' expresses being for blaming for murder *and* being for blaming for murder. Simply listing the same attitude twice in the moral case may seem philosophically strange—but unproblematic.[22]

Furthermore, he suggests that in order to receive the semantic value of a negated formula, one has to negate both contents of the biforcated attitude expressed by the unnegated formula. This implies that

$$\neg P \qquad \text{expresses} \quad \text{FOR}(\neg\text{pai}(p)) \text{ and } \text{FOR}(\neg\neg\text{pai}\neg(p)).$$

_____

cost it would be to drop the assumption that semantics clauses have to apply *uniformly*. But note that even if expressivists would give two different semantic clauses for negation, one applying to moral sentences, and another one for descriptive sentences, they would again face a problem with mixed sentences because then which clause should we apply in order to determine the semantic value of, for instance, the negation of a mixed disjunction? (Perhaps this problem could be solved by requiring that all formulas of the language must be in conjunctive (or disjunctive) normal form. See (Unwin, 1999, 350f.) for such a proposal.) Expressivists would then have to give semantic clauses for negation that applies to mixed disjunctions, where one disjunct is moral, and the other is descriptive. It is easy to see that this problem would quickly get out of hands. Hence this is not a promising strategy.

[22]However, it leads to yet another problem: the *new new negation problem*, which Schroeder discusses and proposes a solution for in (M. Schroeder, 2008a, 113-115) and (M. Schroeder, 2010a). I ignore this problem here, since it is a problem only for Schroeder's idiosyncratic development of expressivism.

Now it turns out that at least one of the states that are expressed by a sentence and its negation are the intuitively correct ones: P expresses *inter alia* FOR(pai(p)), and $\neg P$ expresses *inter alia* FOR(pai$\neg$(p)). Moreover some states expressed by sentences and their negations are A-type inconsistent with each other, such as, for instance FOR($\neg$pai$\neg$(p)) and FOR($\neg\neg$pai$\neg$(p)).

Schroeder's biforcation-trick makes his whole theory more complicated, and in order to really solve the Frege-Geach problem Schroeder needs to make even more assumptions. For instance, Schroeder needs to draw the further distinction between 'minor' and 'major' attitudes, he needs a notion of 'commitment' between mental states, and he has to assume a 'logic of proceeding-as-if'. Additionally, he has to provide new semantic rules for the other connectives for them to apply to biforcated attitudes, and he needs a three-valued logic.[23]

I shall not go further into the details of Schroeder's elaborate *biforcated attitude semantics*, discuss all the various problems it raises, and illustrate how Schroeder attempts to solve them. Instead, I want to show that a theory can be had that is simpler, more plausible, and less problematic than Schroeder's, while being at least as explanatory.

3.5 Attitude Semantics and Logic

3.5.1 Attitude Semantics

The aim of this section is to show that expressivists can have a formally adequate and constructive solution to the Frege-Geach problem without having to assume biforcation and A-type inconsistency, if they elaborate on the very simple idea which I mentioned in the introduction to this chapter. Just to remind ourselves, the idea is that a moral sentence like 'murder is not wrong' expresses something like being for being for not blaming for murder.

Starting from this simple idea, I develop a formal expressivist semantics for a first-order language $\mathcal{L}$ with the following vocabulary: as non-logical symbols we have countably many constants $a_1, a_2, a_3, \ldots$ and predicates $P_1^n, P_2^n, P_3^n, \ldots$ (for any arity $n \geq 1$). As logical symbols we have the connectives $\neg, \wedge, \vee, \rightarrow$, the quantifiers $\exists, \forall$, and countably infinite variables $x_1, x_2, x_3, \ldots$ The recursive definitions for which strings of symbols count as well-formed formulas in $\mathcal{L}$ are as usual.[24]

Here is a structural proposal for atomic sentences:

(E") S accepts 't is F' iff S is χ-ing ψ-ing ϕ-ing t.

---

[23]Consult chapters 8, 9, 10 of (M. Schroeder, 2008a) for more details.

[24]Notation: I will write 'F' for object language predicates and relations, and 'R' for metalanguage relations. I use 't' as a metavariable for all singular terms (constants and variables) in $\mathcal{L}$.

How can this be interpreted? The idea is that 'χ-ing' always has to be replaced by one and the same kind of desire-like mental state, for instance a state like 'being for'; 'ψ-ing' is replaced with some desire-like state (e. g. again 'being for'), if we are dealing with a moral sentence, and with a belief-like state (e. g. 'pai'), if we are dealing with a descriptive sentence; and 'φ-ing' is replaced with some specific relation or property depending on the predicate in question. Here are two possible interpretations for moral and descriptive sentences:

> S accepts 'murder is wrong' iff S is being for being for blaming for murder.

> S accepts 'grass is green' iff S is being for proceeding as if grass is green.

In terms of the expression of mental states, the suggestion is the following:

'murder is wrong'	expr.	FOR(for(blaming(murder)))
'grass is green'	expr.	FOR(pai(green(grass)))

The attentive reader will notice that I do not change Schroeder's original non-biforcated proposal for descriptive sentences. I only enrich the proposal for moral sentences by adding in a further 'being for'. This dissolves the structural disanalogy in Schroeder's original proposal that descriptive predicates always contribute one and the same relation (pai) to the content of the states expressed, whereas the contribution differs with each moral predicate (blaming, praising, preferring...). The motivating thought underlying attitude semantics is that most of the problems biforcated attitude semantics faces are due to this disanalogy, and that the problems can be avoided if we get rid of this disanalogy.

I am not the first to note this disanalogy in Schroeder's expressivist semantics. This has also been noted by Kölbel in a footnote where he writes: "I believe that many of the problems [Schroeder] subsequently detects for the expressivist are the result of this unfortunate and unmotivated [asymmetry at the level of attitudes expressed]" (Kölbel, 2018, Fn.18). But though Kölbel notices this asymmetry, he makes no suggestion as to how expressivists might get rid of it.

In the spirit of Unwin and Schroeder I suggest that it is 'not too implausible' that the mental state expressed by 'murder is wrong' is being for being for blaming for murder. If we allow that this sentence expresses a state of being for blaming for murder, why not also this slightly more complicated state? I want to stress that I am at this stage of the argumentation only interested in providing a specific structure of the mental states expressed by sentences that allows to solve the Frege-Geach problem. I do not mean to defend this particular interpretation of (E"), especially not the somewhat *ad hoc* iteration of being for being for in the moral case. I will suggest a more

plausible interpretation of the schema at the end of this chapter. For now the question is: How can we formalize this structural idea?

For my proposal to be directly comparable to Mark Schroeder's, I will closely follow his way of formalizing the semantics, though this might not be the only, let alone the most elegant way do formalize it. So I will also use lambda-terms to describe the contents of the mental states expressed by our language. For instance, let us say that to be for blaming for murder is a property, albeit a more complex one than Schroeder's more simple property of being blaming for murder. It is the property Mark has, if he is being for blaming for murder. That is, if $\lambda z.\big(z$ is for (blaming for murder)$\big)$(Mark). Like Schroeder, I shall treat '$\lambda z.\big(z$ is for (blaming for murder)$\big)$' as a term denoting the property of being for blaming for murder. And I treat '$\lambda z.\big(z$ is for (blaming for $t)\big)$' as denoting a function from values of t to properties, namely the properties of being for blaming for that value of t. Our semantics will express the semantic contribution predicates make to the meaning of whole sentences by employing such open lambda-abstractions. So, generalizing, I suggest the following account of atomic sentences:

atom   If F is an n-place predicate, then $F(t_1, \ldots t_n)$ is a well-formed formula, and corresponds to a 2-place relation $R_1^F(z, y)$, and an n-place relation $R_2^F(t_1, \ldots t_n)$, so that if $t_1, \ldots t_n$ are singular terms denoting $o_1, \ldots o_n$, then $F(t_1, \ldots t_n)$ expresses FOR$\big(\lambda z.R_1^F(z, R_2^F(o_1, \ldots o_n))\big)$. Treat $\lambda z.R_1^F\big(z, R_2^F(t_1, \ldots t_n)\big)$ as the *semantic value* of $F(t_1, \ldots t_n)$.[25]

Following the above informal idea, I suggest the lexical entry for 'wrong' is:

wrong   If t denotes o, then $Wrong(t)$ expresses FOR$\big(\lambda z.z$ is for (blaming for $o)\big)$. So, $[\![Wrong(t)]\!] = \lambda z.z$ is being for (blaming for t).

Now let us turn to negation. Informally, what we want from our semantics is a clause for negation that makes the following predictions:

'murder is not wrong'	expr.	FOR$\big($for(not(blaming(murder)))$\big)$
'grass is not green'	expr.	FOR$\big($pai(not(green(grass)))$\big)$

---

[25]The open lambda-term $\lambda z.R_1^F\big(z, R_2^F(t_1, \ldots t_n)\big)$ denotes a function which maps values of t to a particular kind of complex properties—the properties of z standing in relation R_1^F to relation/property R_2^F of t, for that value of t. In other words, in this framework, the semantic contribution a predicate F makes to the meaning of a sentence consists not in one relation R^F, but in two relations, namely R_1^F, and R_2^F.

How can we get that? Here is a suggestion:

$\neg_{\textbf{new}}$ If $[\![A]\!] = \lambda z.\alpha$, then $[\![\neg A]\!] = \lambda z.\neg\alpha^-$.

('α^-' is short for '$\alpha[\neg R_1^F/R_1^F; \neg R_2^F/R_2^F]$'.)

Again, 'α' here is a schematic letter standing for arbitrary λ-formulae of our descriptive metalanguage. What is inside the cornered brackets is saying that each inner relation 'R_2^F' appearing inside 'α' is to be substituted with its negation, and the same must be done with each of the outer relations 'R_1^F'. This may look complicated, but it is just a formal move to place the negation-sign where we want it to be, namely in front of the 'ϕ-ing' in the schema (E").

This new clause, together with Schroeder's non-biforcated clauses for conjunction and disjunction (M. Schroeder, 2008a, 78), the above lexical entry for 'wrong', and the following lexical entry for 'green', allows the expressivist to compositionally assign semantic values to a variety of mixed formulas:

$\wedge$ If $[\![A]\!] = \lambda z.\alpha$, and $[\![B]\!] = \lambda z.\beta$, then $[\![A \wedge B]\!] = \lambda z.(\alpha \wedge \beta)$.

$\vee$ If $[\![A]\!] = \lambda z.\alpha$, and $[\![B]\!] = \lambda z.\beta$, then $[\![A \vee B]\!] = \lambda z.(\alpha \vee \beta)$.

green   If t denotes o, then $Green(t)$ expresses $\text{FOR}\big(\lambda z.z$ is pai $(o$ is green$)\big)$. So, $[\![Green(t)]\!] = \lambda z.z$ is pai $(t$ is green$)$.

In order to see how all of this is supposed to work, consider an example. Let A be the mixed conjunction 'murder is wrong and grass is green'. According to our semantics this expresses the state of being for (being for blaming for murder and pai grass is green). More formally, it expresses $\text{FOR}\big(\lambda z.(z$ is for (blaming for murder)$)\wedge(z$ pai (grass is green)$)$. Now in order to determine the semantic value of $\neg A$, we have to (i) negate the whole term following the '$\lambda z.$' (i. e. α), as well as (ii) all the R_1^Fs and (iii) all the R_2^Fs appearing in it. This gives us: $\lambda z.\neg\big(\neg(z$ is being for $\neg$(blaming for murder)$)\wedge\neg(z$ pai $\neg$(grass is green)$)\big)$. Performing some logical manipulations (de Morgan, cancellation of double negation etc.) we receive the following simplified formula: $\lambda z.\big((z$ is being for $\neg$(blaming for murder)$)\vee(z$ pai $\neg$(grass is green)$)\big)$. In other words, this says that $\neg A$ expresses FOR(being for not blaming for murder or pai grass is not green), which, I suggest, is the intuitive result given the analysis we started with.

Why think this is an intuitive result? The idea is that since '$\neg$(murder is wrong $\wedge$ grass is green)' is equivalent to '$\neg$(murder is wrong)$\vee\neg$(grass is green)', and '$\neg$(murder is wrong)' expresses $\text{FOR}\big($being for $\neg$(blaming for murder)$\big)$ and '$\neg$(grass is green)' expresses $\text{FOR}\big($pai$\neg$(grass is green)$\big)$, it is also plausible that '$\neg$(murder is wrong $\wedge$ grass is green)' expresses a state

the content of which is equivalent to the content expressed by the equivalent sentence '¬(murder is wrong)∨¬(grass is green)'.

The account presented here works for negation, conjunction, and disjunction, and can also be generalized to the quantifiers by following the suggestions of (M. Schroeder, 2008a, 79ff). However, we cannot adopt Schroeder's (non-biforcated) semantic clause for conditional formulas which is:

→ If $[\![A]\!] = \lambda z.\alpha$, and $[\![B]\!] = \lambda z.\beta$, then $[\![A \to B]\!] = \lambda z.(\alpha \to \beta)$.

If we would use this rule, we would receive false results when negating a conditional formula. In order to circumvent this problem we can dismiss $\to$ (and $\leftrightarrow$) from the vocabulary, or define it in terms of $\neg$ and $\vee$ (or $\wedge$), and give no extra semantic clause for the conditional. Alternatively, we can put the interdefinition of the conditional directly into the semantic clause, which is what I do here.

To summarize, here are all the semantic clauses of *attitudes semantics* stated together:

predicates If F is an n-place predicate, then $F(t_1, \dots t_n)$ is a well-formed formula, and corresponds to a 2-place relation $R_1^F(z, y)$, and an n-place relation $R_2^F(t_1, \dots t_n)$, so that if $t_1, \dots t_n$ are singular terms denoting $o_1, \dots o_n$, then $F(t_1, \dots t_n)$ expresses FOR$\big(\lambda z.R_1^F(z, R_2^F(o_1, \dots o_n))\big)$. $[\![F(t_1, \dots t_n)]\!] = \lambda z.R_1^F\big(z, R_2^F(t_1, \dots t_n)\big)$.

connectives/quantifiers If $[\![A]\!] = \lambda z.\alpha$, and $[\![B]\!] = \lambda z.\beta$, then $[\![\neg A]\!] = \lambda z.\neg\alpha^-$, $[\![A \wedge B]\!] = \lambda z.(\alpha \wedge \beta)$, $[\![A \vee B]\!] = \lambda z.(\alpha \vee \beta)$, $[\![A \to B]\!] = \lambda z.(\neg\alpha^- \vee \beta)$, $[\![\exists x(A)]\!] = \lambda z.(\exists x(\alpha))$, and $[\![\forall x(A)]\!] = \lambda z.(\forall x(\alpha))$.

The previous discussion shows that Unwin's simple proposal can in fact be generalized. I think that these rules will assign to all formulas of $\mathcal{L}$ the intuitively correct semantic values or mental states. One reason to think that the rules will always deliver the correct results is the same as the one Schroeder cites for the correctness of his non-biforcated attitude semantics for a purely moral language. The reason is that his semantic rules induce an isomorphism between the sentence structure and the structure of their semantic values.[26] In attitude semantics there is something very close to this isomorphism because for instance $[\![P \wedge \neg P]\!] = \lambda z.(\pi \wedge \neg\pi^-)$, $[\![P \vee \neg P]\!] = \lambda z.(\pi \vee \neg\pi^-)$, and $[\![\neg(P \vee \neg P)]\!] = \lambda z.\neg(\pi \vee \neg\pi^-)^-$, etc. These structural similarities will hold irrespective of the logical complexity of any formula of $\mathcal{L}$.

I conclude that the semantics given in this section provides a simple solution to the compositionality problem that is compatible with expressivism's traditional commitments, yet without requiring biforcation.

---

[26]See (M. Schroeder, 2008a, 80f.).

3.5.2   Attitude Logic

In this section I show how logical inconsistency, truth, entailment, and validity can be defined expressivistically for our first-order language with the attitudinal semantics given in the previous section. The underlying idea here is that a formula A is logically inconsistent with $\neg A$ iff one cannot simultaneously be in the mental states expressed by those formulas without being rationally incoherent. For the atomic case this means that $F(t)$ is logically inconsistent with $\neg F(t)$, for any F and t, because one cannot simultaneously be in the mental states expressed by them without being rationally incoherent. If, for instance, F is 'wrong', then the states will be FOR$\big($for(blaming for t)$\big)$, and FOR$\big($for(not blaming for t)$\big)$.

One might wonder if these two being for states are in fact rationally incoherent. I will, however, not argue for their incoherence here but simply assume it because whether or not the states expressed by inconsistent sentences are plausibly rationally incoherent will strongly depend on the *specific interpretation* we give of the structural proposal.[27] Since I do not want to defend the specific interpretation of the states expressed by moral sentences as being states of iterated states of being for, I am not committed to justifying their incoherence.[28] I suggest the following definition of logical inconsistency:

Logical Inconsistency A set of n $\mathcal{L}$-sentences $\{A_1, \ldots, A_n\}$ is logically inconsistent ($A_1, \ldots, A_n \models$) iff for all assignments of mental states to atomic sentences: the states expressed by $A_1, \ldots, A_n$ are jointly rationally incoherent.[29]

---

[27]In chapter 6 below I will have to say more on the rationality of states of being for, and argue that this is a difficult issue because we would have to know more about the nature of being for in order to be able to determine which requirements of rationality it is plausibly subject to.

[28]Below I suggest that it is a better interpretation to replace the outer 'being for' with 'being disposed to'. Then sentences do not express states of being for, but mental dispositions. It might be much more plausible to claim that being disposed to be for something is rationally incoherent with being disposed to be for its negation, or that one would be rationally incoherent if one manifested these dispositions. Moreover, it might ultimately not be important that the feature that we identify as being coextensive with sentential inconsistency is rational incoherence. It might well suffice that we identify some other systematic, perhaps merely formal feature of the (content of the) mental states expressed by inconsistent sentences to define a notion of inconsistency.

[29]Definitions along these lines can be found in (Blackburn, 1984b), (Gibbard, 2003), (M. Schroeder, 2008a), (Ridge, 2014), and (Baker & Woods, 2015). See also the online discussion at PEA Soup: (Paakkunainen, 2015). One benefit of the definition given here is that there seems to be a natural way to extend it to *semantic* inconsistency, by simply dropping the requirement that the states are rationally incoherent for *all* assignments of mental states to atomic sentences. 'Nicholas is married' and 'Nicholas is a bachelor' will then turn out to be semantically inconsistent sentences because they express the rationally incoherent states of being for pai Nicholas is married and being for pai Nicholas is a bachelor. Their inconsistency turns on the non-logical, but analytical implication: if

As a special case it follows from this that A is logically inconsistent, that is a logical falsehood ($A \models$) iff for any assignment the state expressed by A is rationally incoherent. Since the negation of a logical falsehood is a logical truth, it also follows that A is a logical truth iff for any assignment the state expressed by $\neg A$ is rationally incoherent:

Logical Truth A is a logical truth ($\models A$) iff $\neg A$ is logically inconsistent ($\neg A \models$).

Expressivists will want a general way to demonstrate that the rational relations hold in all and only the right cases. In order to prove that *Logical Inconsistency* will only classify those sets of sentences as logically inconsistent that are classically logically inconsistent, expressivists need to provide some way to formally capture the relevant relation of rational incoherence. Here is one way to do so:

Rational Incoherence A set of n mental states $\{M_1, \ldots, M_n\}$ (of the type of state expressed by sentences of $\mathcal{L}$; e.g. states of being for) with contents $\alpha_1, \ldots, \alpha_n$ is rationally incoherent iff $\alpha_1^+, \ldots, \alpha_n^+$ are inconsistent in the metalanguage logic ($\alpha_1^+, \ldots, \alpha_n^+ \models_{ML}$), where α_i^+ is short for $\alpha_i[\neg\beta^-/\neg\beta]$, for all i.[30]

What this definition says, in effect, is that the mental states that are expressed by sentences of $\mathcal{L}$ are rationally incoherent iff their contents turn out to be inconsistent in the metalanguage logic, after replacing every subformula of the form $\neg\beta^-$ appearing in α_i with $\neg\beta$.[31] In short: Cancel all minus signs in α_i, and then check if the resulting formulas are inconsistent. If they are, then the sentences having theses semantic values must be logically inconsistent.[32]

---

married, then not bachelor.

[30] Please note that according to this definition A-type inconsistent states of being for will also turn out to be rationally incoherent.

[31] If one closes all variables in the lambda-abstractions that describe the contents, one receives first-order language sentences. In this chapter, I take the metalanguage to be governed by a classical logic.

[32] One might wonder if this account is explanatory at all. It explains logical inconsistency of sentences in terms of the rational incoherence of mental states, and the rational incoherence of mental states in turn in terms of the logical inconsistency between the contents of those states in the metalanguage. Haven't we then not just explained the logical inconsistency of object-language sentences in terms of the logical inconsistency of meta-language sentences? And don't we then have to explain the latter in terms of rational incoherence again? The answer is no. Metaethical expressivists are first and foremost interested in how sentences can stand in logical relations if moral sentences do not express cognitive states with moral contents, but noncognitive states with non-moral contents. Accordingly, the contents of all states expressed will therefore be non-moral contents. Since expressivists accept that non-moral contents can stand in logical relations with each other, the explanation of inconsistency in terms of incoherence does not have to be repeated on the meta-language level. I thank Martin Fischer for urging me to clarify this point.

To see how this is supposed to work, consider an example. We want it to turn out that $P \land \neg P$ is logically inconsistent. Generally, if the semantic value of P is π,[33] then the semantic value of $\neg P$ is $\neg \pi^-$. So the semantic value of $P \land \neg P$ is $\pi \land \neg \pi^-$. Accordingly, $P \land \neg P$ expresses $\text{FOR}(\pi \land \neg \pi^-)$. Now, in order to find out if $P \land \neg P$ is logically inconsistent, we have to determine whether $\text{FOR}(\pi \land \neg \pi^-)$ is a rationally incoherent state of mind. This will be so just in case its content is logically inconsistent when all subformulas of its content which are of the form $\neg \beta^-$, are replaced by $\neg \beta$. But the only subformula of this form appearing in the content is $\neg \pi^-$. So replacing it with $\neg \pi$ yields $\pi \land \neg \pi$. This is inconsistent. Hence, the state expressed by $P \land \neg P$ is rationally incoherent, so $P \land \neg P$ is logically inconsistent.

The same reasoning will apply to all other logical inconsistencies (and *mutatis mutandis* to the truths), due to the quasi-isomorphism, mentioned at the end of the previous section, which is induced by our semantic rules between the structure of a sentence and the structure of its semantic value, and by the definition of rational incoherence in terms of inconsistency of semantic value after cancellation of all minus-signs.

With logical inconsistency on board, we can define entailment and validity in a straightforward way:

Logical Entailment $A_1, \ldots, A_n$ logically entail B (i. e. $A_1, \ldots, A_n \models B$) iff $\{A_1, \ldots, A_n, \neg B\}$ is logically inconsistent.

Logical Validity An argument with premises $A_1, \ldots, A_n$ and conclusion B ($A_1, \ldots, A_n \therefore B$) is logically valid iff the premises logically entail the conclusion ($A_1, \ldots, A_n \models B$).

The attitude semantics and logic given in this and the previous section provide a solution to both the expressivist's problem of compositionality and the problem of logicality. Thereby, attitude semantics provides a solution to the Frege-Geach problem of the simple kind Unwin envisaged.

3.6   Comparison

Schroeder claims that his theory is superior to various contemporary expressivist theories, most prominently the ones of Blackburn (1984b), Gibbard (2003), and Horgan and Timmons (2006), because only biforcated attitude semantics really provides a formally adequate and constructive solution to the Frege-Geach problem. But the major drawback of his theory is that the *ad hoc* move to biforcation is necessary "in order to prove all the interesting results about biforcated attitude semantics" (M. Schroeder, 2008a, 175). Attitude semantics shows that one can achieve these results without having

---

[33]I suppress the $\lambda z.$ here for better readability.

to assume biforcation. This is a major point in its favor, but does this suffice to show that attitude semantics is generally the better choice for expressivists?

The aim of this section is to provide some evidence for the claim that attitude semantics is generally superior to biforcated attitude semantics with regard to its various merits and demerits. Since most of the merits of biforcated attitude semantics require biforcation and many of its problems are due to biforcation, it is an interesting question how (non-biforcated) attitudes semantics fares in these respects. I think that attitude semantics has all of the merits of biforcated attitude semantics and is able to avoid most, if not all, of its problems. I will not attempt to provide a full comparison here, but in order to illustrate the potential of attitude semantics for expressivists, I will in the following two sections sketch how it can explain the intuitive truth-conditions of purely descriptive sentences, and provide a notion of disbelief that is not definable in terms of belief and negation.

3.6.1 Truth-Conditions and Mistake-Conditions

Expressivists want to be able to say that purely descriptive sentences have *robust* truth-conditions.[34] Schroeder devotes two whole chapters to showing how biforcated attitude semantics can explain the intuitive robust truth-conditions (henceforth: truth-conditions) of descriptive sentences. Both of his accounts require biforcation though.

In this section I show that attitude semantics is able to explain truth-conditions too. I show this by following Schroeder's second approach, namely his "attempt to generate truth-conditions for [descriptive] sentences directly from the mental states that they express" (M. Schroeder, 2008a, 131).

Schroeder starts with the idea that mental states have mistake-conditions— conditions under which it is a mistake to be in them.[35] The hope is that the conditions under which it is a mistake to be in the states expressed by a descriptive sentence explain the intuitive truth-conditions of the sentence.

Schroeder illustrates this with the following example: it is a mistake to believe that Kerry is president, if Kerry is not the president, i.e. if it

---

[34]Expressivists will ultimately want to say that even moral sentences have truth-conditions. But they normally understand those in some minimalist fashion. I have more to say about minimal truth and minimal truth-conditions in the next chapter (4.1.). In this section I am only concerned with how expressivists might claim purely descriptive sentences to have robust truth-conditions.

[35]Unfortunately, Schroeder is pretty silent about the notion of 'mistake'. One might think that there must be some connection between what he previously said about semantic correctness or assertability conditions, that is condition under which it is semantically permissible to assert something on the one hand, and the mistake conditions of mental state. But the connection does not seem to be so straight forward. Likewise, one might think that there must be some connection between what I have called sincerity conditions of speech acts on the one hand and Schroeder's mistake-conditions of mental states. But again it is not clear what the connection is.

is false that Kerry is president. This suggests, Schroeder proposes, that 'Kerry is president' "has the truth-condition that Kerry is president, because this is the condition that must be satisfied, if the speaker is not to make a mistake either in what to believe or in what to say, given what she believes" (M. Schroeder, 2008a, 132).

Now since both biforcated attitude semantics and attitude semantics assign states of being for to descriptive sentences, expressivists have to try to explain the truth-conditions of those sentences by explaining why the corresponding states of being for have the relevant mistake conditions. Here Schroeder makes two assumptions. First, it is a mistake to be in the state FOR(α), if it is a mistake to be α. Second, it is a mistake to proceed as if p, if p is false, i. e. if $\neg p$ is true.[36] It follows from these assumptions that it is a mistake to be in the state expressed by some atomic descriptive sentence P, if p is false. Why? Well, according to the semantics given here a descriptive sentence P expresses FOR(pai p). By the second assumption, it is a mistake to be pai p, if p is false. But if it is a mistake to be pai p, then, by the first assumption, it is also a mistake to be FOR(pai p). So it follows that it is a mistake to be FOR(pai p), if p is false. So, p is the condition that must be satisfied if a speaker is not to make a mistake in being in the state expressed by P. This, Schroeder suggests, explains why P is true iff p.[37]

However, Schroeder's explanation of truth-conditions does not carry over to negated sentences as construed on his *non*-biforcated view. On this view $\neg P$ expresses FOR($\neg$pai p), which, as we have already seen, is both implausible and problematic. The above assumptions do not entail that it is a mistake to be FOR($\neg$pai p), if $\neg p$ is false, i. e. p is true, which, however, is the condition that would have to be satisfied in order to explain that $\neg P$ is true iff $\neg p$. *Biforcated* attitude semantics does not have this problem because in it $\neg P$ expresses, inter alia, the state of FOR(pai $\neg p$). So Schroeder concludes that "the case of negation again shows that the biforcated attitude approach is necessary" (M. Schroeder, 2008a, 134).

But the possibility of attitude semantics shows that this is false. One of the points of this chapter is that one can avoid biforcation, even within the constraints that Schroeder places on the semantic program of expressivism. In attitude semantics, $\neg P$ expresses FOR(pai $\neg p$), which is both plausible and unproblematic, and this state does have the right mistake conditions to explain the truth conditions of the negated sentence. The explanation is exactly parallel to the one above and runs as follows. By the second

---

[36]Please note, in connection with the previous footnote, that both these assumptions about mistakes cannot be captured in terms of a norm of assertion, assertability conditions or sincerity conditions more generally. Instead, especially the second assumption, bears an obvious resemblance to what has been called the 'truth norm of belief'. For more on the truth norm see for instance (McHugh, 2012), (Whiting, 2010), (Whiting, 2013).

[37]All assumptions and the proof that this holds for descriptive sentences of arbitrary complexity are in an appendix to chapter 10 in (M. Schroeder, 2008a).

assumption, if $\neg p$ is false, then it is a mistake to proceed as if $\neg p$. And if it is a mistake to proceed as if $\neg p$, then, by the first assumption, it is also a mistake to be FOR(pai $\neg p$). So it is a mistake to be FOR(pai $\neg p$), if $\neg p$ is false. This explains why $\neg P$ is true iff $\neg p$.

This shows that expressivists can explain how descriptive sentences get their truth-conditions: they get them from the conditions under which it is a mistake to be in the states expressed by those sentences. However, unlike biforcated attitudes semantics, attitude semantics does not require biforcation to establish this result.[38]

3.6.2 Disbelief

Schroeder argues that even though biforcation might be viewed as a theoretical cost of his expressivism, biforcation "is actually an *advantage* of this account" (M. Schroeder, 2008a, 102, emphasis in original). This is so, he claims, because due to the richer structure of biforcated attitudes it allows him to define a state of disbelief, doubt, or disacceptance. The state of disbelief, Schroeder suggests, is philosophically interesting, because it can help us in dealing with paradoxical sentences such as the Liar-sentence, "surprising" scenarios, or cases of "unreliable testimony" (M. Schroeder, 2008a, 103.). So the possibility to define a state of disbelief decreases the costs of biforcated attitude semantics at least to some degree.

I agree with Schroeder that it would be beneficial to have a state of disbelief, and that biforcated attitude semantics can define such a state. But this would be an advantage over attitude semantics only if it were impossible or at least more difficult to define a state of disbelief in attitude semantics. It is, however, possible and actually easier to define a state of disbelief in attitude semantics than it is in biforcated attitude semantics.

In order to understand Schroeder's account of disbelief, recall that in biforcated attitude semantics every sentence expresses a pair of mental states. According to Schroeder one of these states is what he calls 'major', the other 'minor', where the major state is the one that 'commits' to the other. Schroeder understands commitment in the following way: a state of being for commits to another state of being for iff the content of the first state entails the content of the latter.[39]

Since Schroeder frames his account in terms of acceptance, rather than belief, I will follow his usage here and speak of disacceptance instead of disbelief. According to Schroeder you accept a sentence (or believe what it says) iff you are in the mental state it expresses. Moreover, you deny a

---

[38]I explained only how atomic sentences and their negations get their truth-conditions. Arguably, we will need more assumptions of the kind presented in (M. Schroeder, 2008a, Ch. 10) in order to establish that complex sentences also get the truth-conditions from the mental states they express.

[39]See (M. Schroeder, 2008a, 98).

sentence iff you accept its negation. Finally, you *disaccept* a sentence (i. e. disbelief what is says) iff you are in the minor attitude expressed by its negation.

Now consider an example. In biforcated attitude semantics, if P is a descriptive sentence expressing FOR(pai p), $\neg P$ expresses the pair consisting of FOR(pai$\neg p$) and FOR($\neg$pai p). Given Schroeder's assumption about the 'logic of proceeding as if' that pai p entails $\neg$pai$\neg p$ (M. Schroeder, 2008a, 122), it turns out that FOR($\neg$pai p) is the minor attitude expressed by $\neg P$ (because pai $\neg p$ entails $\neg$pai p by contraposition). So, FOR($\neg$pai p) is the state one has to be in order to disaccept P. Schroeder argues that this is a welcome result, since now it turns out that, in general, disaccepting A (like denying A) is A-type inconsistent with accepting A, though weaker in commitment than denying A (i. e. accepting $\neg A$).[40]

For this proposal of defining a state of disacceptance to work, Schroeder needs biforcated attitudes because without them we can have no major and minor attitudes. In *attitude* semantics there are no major and minor attitudes because there are no biforcated attitudes. Nevertheless, we too can define a notion of disbelief by saying that to disaccept A is to be for the negation of *its* content. E. g., if A has semantic value α, then you disaccept A just in case you are FOR($\neg\alpha$).[41] For instance, if accepting 'grass is green' is FOR$\big($pai(green(grass))$\big)$, then disaccepting 'grass is green' is FOR$\big(\neg$pai(green(grass))$\big)$. Note that this is the same state Schroeder identifies with disaccepting P, but without the detour over biforcated attitudes and major and minor attitudes. The same goes for moral sentences. For instance, you disaccept 'murder is wrong' iff you are FOR$\big(\neg$for(blaming(murder))$\big)$.

Moreover, it follows from the definition of rational incoherence above that disaccepting A (FOR($\neg\alpha$)), like denying it (FOR($\neg\alpha^-$)), is rationally incoherent with accepting it (FOR(α)).[42] So, attitude semantics is able to deliver the same result, but again without requiring biforcation.

There is a further difference worth emphasizing between Schroeder's account of disacceptance and the one I have just sketched. By accepting any sentence of biforcated attitude semantics one *ipso facto* disaccepts its negation. Why? This is so due to the fact that if every sentence A expresses a biforcated attitude, the minor attitude of it always corresponds to disaccepting $\neg A$. So, the rational requirement, if it is one, between accepting a sentence and

---

[40]Cf. (M. Schroeder, 2008a, 102).

[41]Observe the difference to denying A, which is accepting $\neg A$, and so is FOR($\neg\alpha^-$).

[42]If we allow that α entails $\neg\alpha^-$ (an instance of this would be Schroeder's 'logic of proceeding-as-if' which actually is an analogue of the Axiom D in doxastic logic), then disaccepting A will in general be weaker than denying A. If this is allowed, then I suspect that attitude semantics is also able to capture the same (strong Kleene) three-valued "commitment-tables" that Schroeder proposes, and so should be able to account for truth-theoretic paradoxes like the Liar in ways similar, but simpler, to Schroeder's treatment in (M. Schroeder, 2008a, 102-104) and M. Schroeder (2010a). I guess that then attitude semantics also predicts that the only coherent epistemic attitude toward the Liar is to disbelieve both it and its negation.

disaccepting its negation is hard-wired into biforcated attitude semantics.[43] That one accepts A, only if one disaccepts $\neg A$ is a norm that a speaker of Schroeder's biforcated language cannot decide to obey or disobey—it will always hold a matter of semantic fact. This is not a good result, because how rational or irrational one is should not depend in such a strong way on the language one speaks.

The present account also differs from Schroeder's in the following respect: whereas in biforcated attitude semantics every sentence expresses *inter alia* a state of disacceptance, in attitude semantics, as it currently stands, it is not even possible to express disacceptance.[44] There is no sentence in the language that expresses FOR($\neg\alpha$); there are only sentences expressing FOR(α) or FOR($\neg\alpha^-$). But even if there is no semantic means for speakers to express their disacceptance in attitude semantics, it seems possible for them to *be* in a state of disacceptance, e. g. to be FOR($\neg\alpha$). Whether or not a speaker will disaccept $\neg P$, if he accepts P, is then not be a matter of semantic fact, but a rational norm a speaker *can* decide to obey or disobey. I think that this is a further benefit of the present account compared to Schroeder's.

One might worry that it is problematic that one can disaccept a sentence of a language, but be unable to express one's disacceptance.[45] I am not sure how problematic the inexpressibility of disbelief really is, but Schroeder does not seem to find the idea problematic that there might be states that are not expressible using (a particular) language. But one thing that should be noted is that even though, some states might be inexpressible (in the technical sense of 'expressing' that expressivists are employing as their key semantic notion), this does not mean that it must be impossible to *communicate* in some other, possibly pragmatic, way that we are in them. For instance, utterances of first-person attitude ascriptions like 'I do not believe that p', 'I suspend judgment about p', or 'I disbelieve that p', or 'I have not idea whether p' might be taken as implicating, presupposing, or in some other pragmatical way conveying that the speaker disbelieves that p.[46] Moreover, even if in

---

[43]It is controversial if there are rational requirements on the state of disbelief at all, and if there are, what they are. For a treatment of rational requirements on the state of disbelief see Rosa (2016).

[44]In biforcated attitude semantics there is also no sentence that *only* expresses a state of disacceptance because all the attitudes expressed in biforcated attitude semantics are pairs of states where one is major and the other is minor. No state expressed there consists of two minor attitudes.

[45]The, I think misguided, worry that some attitudes are inexpressible in (biforcated) attitude semantics has been discussed by Köhler (2012). More recently the inexpressibility of some states has been used in a objection against Schroeder in (Baker, 2018).

[46]Some authors such a Wolfgang Freitag defend the rather unpopular view, which he calls 'psychological expressivism', that first-personal attitude ascriptions, what he calls 'avowals', such as 'I am in pain', do not express the belief that the speaker is in the attitude he self-ascribes to him, but rather directly expresses the attitude itself (Freitag, 2018). In my view, given the technical understanding of expressing presupposed in this chapter which is in terms of sincerity conditions, this seems false. But if we understand 'express'

the language we are currently considering there is no means of expressing a state of disacceptance, of course nothing prevents us from introducing a new linguistic means to it in order to allow speakers of this language to express disacceptance in this language. For instance, we might simply stipulate that placing a 'mhmm' in front of some sentence results in a well-formed sentence expressing disbelief or doubt about what the sentence says. 'mhmm, grass is green', for instance, we might then stipulate, expresses the state of being for not proceeding as if grass is green.

3.7   Dispositional Expressivism

In this final section I want to sketch a way to avoid *global noncognitivism*, which is the view that to be in the state expressed by a sentence, is to be in a noncognitive state, in our example the state of *for* something. One might have several worries about this view. It is quite a radical and revisionist theory about the nature of the states expressed by declarative sentence, which most philosophers would normally regard as being beliefs. Moreover, since the states expressed by moral and descriptive sentences alike are treated as noncognitive states of being for, this also threatens to blur the traditional distinction between noncognitivism and cognitivism.[47]

A further worry concerns the specific interpretation I have given of the structure of the states expressed by moral sentences. According to attitude semantics accepting 'murder is wrong' is being for being for blaming for murder. I think it is natural to object that the iteration of 'being for being for' in this account is too complicated, too artificial, too higher-order or just another *ad hoc* expressivist assumption which lacks any independent philosophical justification.

I think all of these worries are legitimate. Fortunately, they have an attractive answer. The answer is to replace the outer being for attitude with a mental disposition, and so replace global noncognitivism with *global dispositionalism*. This leads to a promising form of expressivism, *dispositional expressivism*, which is the view that all indicative sentences express not noncognitive states of being for, but mental dispositions. The idea is that instead of saying that accepting 'grass is green' is the state of being for proceeding as if grass is green, and instead of saying accepting 'murder is wrong' is the state of being for being for blaming for murder, we say that the former is the *disposition* to proceed as if grass is green, and the latter is the *disposition* to be for blaming murder. In more schematic form, I propose to replace our previous picture:

---

in the sense of carrying the conventional or conversational implicature that a speaker is in some state, this might be more plausible.

[47]I return to this distinction in my discussion of the 'tightrope problem' in chapter 5.2.4 below.

'grass is green'	expresses	FOR(pai(green(grass)))
'murder is wrong'	expresses	FOR(for(blaming(murder)))

with something like that picture:

'grass is green'	expresses	DISP(bel(grass is green))
'murder is wrong'	expresses	DISP(for(murder is blamed)).

Note that I here not only replaced the outer being-for relation with a disposition, but that I also replaced the proceeding-as-if relation ('pai') with an occurrently-believing relation ('bel'), and the being-for relation ('for') with an occurrently-being-for relation ('for'). This is, of course, optional, and there are plenty other possible replacements. For instance, adopting the view of Soames (2013) that representation is first and foremost a mental act, one could replace 'proceeding-as-if' with the mental act of 'representing-as-true'. Other common dispositionalist proposals in the literature are that to believe a proposition is the

> disposition to assent to utterances of P in the right sorts of circumstances [...,] the disposition to exhibit surprise should the falsity of P make itself evident, the disposition to assent to Q if one is shown that P implies Q, and the disposition to depend on P's truth in executing one's plans. Perhaps all such dispositions can be brought under a single heading, which is, most generally, being disposed to act as though P is the case. (Schwitzgebel, 2019)

Moreover, the inner 'being for' in the moral case could equally well be replaced by 'intending', 'approving', 'seeing-to-it-that', or something similar.[48] It is also possible to treat the objects of occurrently believing and occurrently being for as being propositions, just I have done above. If we treat being for as a propositional attitude we can say that to accept 'murder is wrong' is being disposed to be for the fact that murder is blamed. This is also optional of course, but this move, like the move to dispositions, helps to further reduce of the impression that this attitude semantical account is too higher-order.

The dispositional interpretation replaces the merely theoretically motivated iteration of 'being for being for' in the moral case with the 'disposition to be for' (or something relevantly similar) which is a philosophically much more credible proposal. Under this particular interpretation it no longer turns out that the states expressed by *all* sentences, and especially those

---

[48]Replacing it with the seeing-to-it-attitude might be attractive for it would then be able to connect it to what is called stit-logic. For more on stit-logic see for instance (Belnap, 2001) and (Horty, 2001).

expressed by purely descriptive sentence such as 'grass is green', need to be understood as noncognitive desire-like states of being for. We can now understand descriptive sentences as expressing the disposition to stand in a genuinely cognitive relation (occurrently believing) to a proposition. This is closest to being the orthodox view according to which the state expressed by descriptive sentences involves a cognitive relation to a proposition.

I think that the move to global dispositionalism is a promising idea for metaethical expressivists, and not only for the reasons just given. We will see later (in the second part of this dissertation) that it also helps with the project of quasi-realism. Ultimately, quasi-realists expressivists want to earn the right to identify the states their semantics assigns to sentence with beliefs, since the ordinary view is that declarative sentences are used to express beliefs. The dispositional interpretation allows expressivists to connect to dispositionalist theories of belief, according to which belief, and perhaps, all mental states, are mental dispositions. This is a view that philosophers throughout the 20th century have found and still find plausible. I will have to say much more on the idea that beliefs are dispositions in chapter 5 below.

3.8 Conclusion

In this chapter, I suggested an expressivist semantics, *attitude semantics*, that provides a formally adequate and constructive solution to the Frege-Geach problem. I argued that attitude semantics is superior in many respects to Mark Schroeder's biforcated attitude semantics. Furthermore, I suggested that a philosophically more plausible interpretation of the nature and structure of the mental states expressed by sentences might be in terms of dispositions. Dispositional expressivism is a promising idea that deserves being pursued further, and I will say more to justify this idea in part two of this dissertation which is about moral belief. But before turning to the question of how expressivists might earn the right to moral belief, I want to focus on another part of the expressivists' quasi-realist project, namely that of earning the right to truth and especially to truth-aptness.

Chapter 4

Truth and Truth-Aptness

4.1 Overview

Contemporary metaethical expressivists not only claim that the meaning of moral sentences can be compositional and that they stand in logical relations. Another important goal also is to show that moral sentences are truth-apt and that the states they express can be regarded as beliefs. How can this be squared with their claim that moral sentences express noncognitive mental states that do not seem to be assessable in terms of truth and falsity? In other words they want to answer questions 3 and 4 mentioned in the introduction. Answering those question is the project of quasi-realism.

In this chapter, I will be concerned primarily with the question of truth-aptness. I argue that what I call the 'orthodox strategy' of quasi-realism to establish moral truth-aptness and moral belief is mistaken. The orthodox strategy is to argue first for the truth-aptness of moral sentences, and then to conclude that they express beliefs. This strategy needs to be inverted, I claim. The reason is that once we understand what it takes for a sentence to be truth-apt, we see that we must establish that moral sentences express beliefs before we can establish that they are truth-apt.

The chapter is structured as follows. I start by explaining the expressivists' accommodation project of quasi-realism. Since most quasi-realist expressivists believe that minimalism about truth is the key to earning the right to several allegedly problematic notions, I will give a rough presentation of minimalism about truth. We will see that just as minimalism about truth does not imply, but rather presupposes a theory of meaning, it also does not imply, but presupposes a theory of truth-aptness. I then discuss and reject several theories of truth-aptness mainly by providing counterexamples to them. In passing I also discuss some arguments that have been made in favor of and against expressivism in which truth-aptness plays a central role. Looking carefully at those arguments will sharpen our understanding of some key claims in the debate. This ultimately leads to my own theory

of truth-aptness which says, roughly, that sentences are truth-apt iff and because they can be used to make assertions, which is tantamount to saying that they can be used to express beliefs. An important consequence of this theory is that it shows the orthodox quasi-realist strategy to be mistaken.

4.2   Truth

4.2.1   From Emotivism to Quasi-Realism

Early emotivists such as, most prominently, Alfred Ayer, held, or at least implicitly accepted some quite radical claims about ethics (Ayer, 1952). For instance, Ayer claimed that no moral sentence is ever true because moral sentences are not even apt for being true or false. Moreover, he claimed that moral sentences lack truth- or verification-conditions, they do not express propositions, are meaningless (lack literal or factual meaning), cannot be contradicted, and cannot be used to make assertions or express beliefs. Adding an ethical word like 'wrong' to a sentence, he famously proclaimed, is like speaking in a "peculiar tone of horror" or like adding some "special exclamation marks" (Ayer, 1952, 107). In less colorful terms he claimed that in uttering moral sentences we are "merely expressing moral sentiments", that is we are expressing non-cognitive, non-truth-evaluable states of mind (Ayer, 1952, 107). Moreover, Ayer and the like in various places held that moral sentences are not about moral reality, that is they do not describe or represent moral facts. They also claimed that moral concepts are mere 'pseudo-concepts'.[1]

What is surprising about early emotivists is that they have been willing to accept that if their claims were true, then ethics looks like a totally misguided project, one we should perhaps stop engaging in further. It is also surprising that Ayer made no serious attempt to explain why it nevertheless seems to make perfect sense to attribute truth or falsity to moral sentences, why they appear to have truth-conditions, seem to express propositions, be meaningful, can be contradicted, figure in logical arguments, can be asserted and express what we consider to be our beliefs about moral matters.

If Ayer's emotivism were true, then the error we would all have to be in, would be much bigger than the error moral error theorists ascribe to us, namely the error of believing that moral sentences, assertions, propositions, and beliefs are sometimes true.[2] The error would be much bigger because the error theorist, unlike the emotivist, does not deny that moral sentences are meaningful, express propositions, have truth-conditions, can be used to make assertions and express beliefs. The error theorist only denies that moral sentences are sometimes true because he thinks that their truth-conditions are never satisfied.

---

[1] For an systematic introduction to emotivism see the classic (Urmson, 1968).

[2] The father of moral error theory is John L. Mackie. See his famous (Mackie, 1977) for more. A contemporary defense of error theory is provided by (Streumer, 2017).

Contemporary expressivists are much less radical and want to do better than their emotivist precursors, but also better than the error theorist. They want to preserve as much common-sense about our moral talk and thought as possible. In order to do so contemporary expressivists pursue the project of quasi-realism. Quasi-realism was introduced to the literature by Simon Blackburn, but is also associated with Allan Gibbard's work on expressivism.[3] Quasi-realism is not so much a view rather than a project. It is the expressivists' attempt to "earn the right" (Blackburn, 1984b, 197) to such realist-sounding notions like 'truth', 'fact', 'reference', 'belief', 'knowledge' and 'assertion' and others. Quasi-realism is also sometimes described as the project of explaining why moral sentences have the surface-structure they have, namely the same as sentences about ordinary descriptive matters. For instance, Blackburn himself described quasi-realism as

> the enterprise of explaining why our discourse has the shape it
> does, in particular by treating evaluative predicates like others ...
> It thus seeks to explain, and justify, the realistic-seeming nature
> of our talk of evaluation. (Blackburn, 1984b, 180)

Since it is an open question what is all 'realistic-seeming' about moral talk and thought, it is also an open question which claims quasi-realists ultimately want to establish. But it is fairly uncontroversial that they want to establish at least the following three claims:

Moral Truth: Some moral sentences are true.

Moral Truth-Aptness: Moral sentences are truth-apt.

Moral Belief: Moral sentences express beliefs.[4]

On the one hand, these three claims are arguably central to our ordinary understanding of morality. On the other hand, these claims seem to be incompatible with the key commitments of metaethical expressivism. Moral Truth seems to be ruled out by the metaethical expressivists' commitment to anti-realism, according to which there are no robust moral facts. If there are no moral facts, how can there be truth about them? Moral Truth-Aptness and Moral Belief seem to be ruled out by the expressivists' claim that moral sentences express noncognitive states: if noncognitive states cannot be assessed in terms of truth and falsity, how then should moral sentences

---

[3]The central works are (Blackburn, 1984b), (Blackburn, 1993), (Blackburn, 1998), and (Gibbard, 1990) and (Gibbard, 2003).

[4]A fully worked-out quasi-realist expressivist might want to establish more than those claims. For instance, she might also want to establish that moral sentences express moral *propositions*, and moral predicates express moral *properties* or *concepts*, that moral sentences represent *moral facts*, and that that there is moral *knowledge* etc. I will, however, only focus on those three claims which I take to the the most central ones.

be truth-apt? And if moral sentences express noncognitive desire-like states, how can they express beliefs, when beliefs and desires are distinct?

The expressivists' preferred way of realizing the quasi-realist project has been to embrace some form of minimalism or deflationism about the above notions. In particular, it is thought that minimalism about truth will be particularly helpful for getting the quasi-realist project up and running. It is not totally clear why truth has been regarded as central, but a natural idea is that many of the other problematic notions, such as 'truth-aptness', 'fact', 'representation', etc. can be defined in terms of or by reference to truth, and that other notions such as 'belief', 'knowledge', and 'assertion' can be defined in terms of those notions. So I start my discussion with minimalism about truth.

4.2.2   Minimal Truth

It has often been thought that adopting some form of minimalism about notions such as truth, fact, meaning, representation, reference, truth-aptness, truth-condition, proposition and belief, will help the expressivist with her quasi-realist project. There have, however, also been serious concerns about whether minimalism and expressivism can successfully be combined or whether minimalism even shows expressivism to be false. We will see below that these incompatibility worries fail to be successful attacks on contemporary quasi-realist forms of expressivism. There is also the problem of so-called 'creeping minimalism', which is the problem of how to distinguish quasi-realism from realism, once the quasi-realist project has been successfully realized. I will not enter this latter debate here because I do think that this problem is not a really deep one, and I take some refined version of Dreier's proposed solution, his 'explanation explanation', to be basically on the right track.[5] Even if minimalism ultimately enables the quasi-realist expressivist to 'earn the right' to speak of moral beliefs, sentences or propositions as representing moral facts, there remain clear differences in the realist's and the quasi-realist's explanations of *why* we have the right to so speak.

In this section I will provide an overview over key minimalist theses about *truth*. This overview is by no means comprehensive. There is a huge literature on minimalism about truth, and I will only scratch the surface here. Digging deeper into the details, let alone dealing with technical issues surrounding the question of whether or not the semantic paradoxes force defenders of truth minimalism to abandon classical logic, for instance, is beyond the scope of this study. For the purpose of this study I will only sketch what is necessary in order to understand what minimalism about truth potentially achieves for quasi-realist expressivists, and, more importantly, what it is unable to achieve, namely a theory of truth-aptness. This will let us see clearer where

---

[5]For more on the 'creeping minimalism'-worry see (Dreier, 2004b) and (Simpson, 2018).

quasi-realist expressivists have to do the real work. To anticipate my later claim: the real work has to be done in establishing a theory of belief that is compatible with noncognitivism.

So what is minimalism about truth? According to minimalism about truth, truth is not a substantive property (such as, for instance, correspondence to facts, coherence, or utility...) whose complex nature requires deep metaphysical inquiry. It should not surprise us that Ayer accepted an early version of minimalism about truth according to which truth or 'true' is redundant. He claimed that:

> we find that in all sentences of the form 'p is true,' the phrase 'is true' is logically superfluous [...] Thus to say that a proposition is true is just to assert it, and to say that it is false is just to assert its contradictory. And this indicates that the terms 'true' and 'false' connote nothing, but function in the sentence simply as marks of assertion and denial. (Ayer, 1952, 88f)[6]

Since such a redundancy theory of truth is nowadays generally regarded as untenable, contemporary minimalists make more modest claims about truth and often base their theory of truth on what has been called the truth or equivalence schema:

(T) 'p' is true iff p.

Defenders of truth minimalism claim that this schema forms the conceptual core of the concept of truth. (T) does not offer an explicit definition of truth, but only an implicit one, by characterizing the concept of truth. To have the concept of truth, minimalists claim, is to be disposed to accept all (perhaps all non-paradoxical or non-pathological) instances of this schema. They also often present their view by saying that predicating truth of some sentence or proposition is equivalent to asserting that sentence or proposition, e. g. to call 'snow is white' true is simply a way to call snow white. However, according to them, the truth predicate is not redundant, because it fulfills important functions which enable us to say things we could not say without it. First, it enables 'semantic ascent': it allows us to freely switch between talk about sentences or propositions to talk about the world. Second, and more importantly, the function of the truth concept, minimalists claim, is to allow us to form generalizations and speak about propositions without directly asserting them, like in 'Everything the pope says is true' or 'What Amanda said is true'. Due to these functions the truth predicate is often said to be purely logical, and the property expressed by it is a logical property.

---

[6]Ayer's claims about truth and falsity parallel his claims about goodness and wrongness. Famously, Ayer claimed that the presense of 'good' ('wrong') in a sentence adds noting to its factual content or meaning (Ayer, 1952, 107). Convinced by Moore's open question argument, Ayer held that alethic as well as ethical concepts are 'unanalysable' or primitive.

It is its usefulness in these contexts that justifies or explains why we have and employ a concept of truth.

Minimalists who take (T) to capture everything significant that can be said about our concept of truth can be divided along two dimensions. The first dimension concerns how to understand the equivalence. The second dimension concerns how to interpret (T)'s left-hand side. With regard to the first dimension, minimalists can decide whether they take the equivalence to be material, analytic or necessary . I take it that minimalists who regard (T) to be the very foundation of the concept of truth will not interpret the 'iff' in play here as material equivalence. Instead, they will either take the equivalence to be analytical or necessary. However, treating it as analytic is tantamount to saying that 'p is true' means the same as 'p' which is just the redundancy theory of truth which for good reasons has fallen largely into philosophical discredit. So it appears that minimalists should conceive of the equivalence in (T) as necessary equivalence. At least I will assume this in what follows.

With regard to the second dimension, minimalists can decide whether they take the things of which truth is predicated to be sentences or to be propositions. However, both options seem problematic. The problem with taking sentences to be the primary truth bearers is that then (T) will come out as false, if interpreted as a conceptually necessary claim. To see this consider a sentential reading of (T):

(ST) (The sentence) 'p' is true iff p.

It appears to be central to our concept of truth that for instances of (T) to be true it is required that the object-language sentence denoted by "p" in the left-hand side *means the same* as the meta-language sentence 'p' on the right-hand side. However, (T) does not mention meaning at all. Since what a sentence means is an empirical fact about a particular language, there will be instances of (T) that are false, and so (T) cannot formulate a necessary truth. Hence, sententialism about (T) is regarded as problematic.[7]

The problem that the sentence mentioned on the left-hand side may mean something other than what is said on the right-hand side is avoided if (T) is interpreted not as being about sentences, but about what they mean or the propositions they express:

(PT) (The proposition that) p is true iff p.

---

[7]A common reply to this objection is that the sentences denoted by "p" must be *interpreted* sentences, that is have a meaning. But even so, (T) only captures truth concepts that are relativized to specific languages. It does not capture our ordinary, general concept of truth.

Obviously, the '*p*'s mean the same on both sides, since they are the same. However, the problem with (PT), as a theory of truth, is that it appears to be trivial. It has been suggested that perhaps the best way for propositional minimalists is to bite the bullet, and accept the triviality because saying that something is trivial is not the same as saying it is false.

Both the (interpreted) sentential minimalist as well as the propositional minimalist presuppose a notion of meaning or propositional content in order to explain truth. This has led many philosophers to think that minimalists about truth cannot explain meaning or propositions in terms of truth because this would result in circularity.

In order to see why minimalism about truth is thought to be incompatible with a theory of meaning that explains meaning in terms of truth, consider how a paradigmatic truth-conditional semanticist, like Donald Davidson, would explain meaning.[8] Davidson's theory of meaning, as developed in (Davidson, 1967), attempts to replace P and p in the following axiom schema in such a way that we receive a list of all the true instances of (M) for some particular natural language $\mathcal{L}$:

(M) P means (in $\mathcal{L}$) that p.[9]

However, even if such a list can be given, it is not achieved in a finite and compositional way, and so it violates two important constraints of adequacy for any theory of meaning. Davidson famously suggested that in order to avoid this problem, 'P (of $\mathcal{L}$) means that p' has to be explicated as 'P is true-in-L iff p'. Roughly put, the idea is that 'snow is white' means that snow is white, *because*, as a matter of the convention, $\mathcal{L}$ is such that 'snow is white' is true-in-$\mathcal{L}$ iff snow is white. In other words, Davidson proposed to replace the meaning-schema (M) with Tarski's truth-schema (T) which would allow him to determine the truth-conditions, and thus meanings, of the sentences of some language in a recursive, and thus compositional, way.

But minimalists about truth cannot pursue this way of spelling out meaning. An adequate theory of meaning for some language has to be such that knowing the theory is sufficient to know the meanings of all sentences of the language. According to Davidson's theory, to know the meaning of a sentence is to know when it would be true. Hence, knowing Davidson's theory of meaning amounts to nothing more than knowing all the instances of (T). This is, however, precisely what minimalists claim is required to have the truth-concept in the first place. Michael Dummett saw this incompatibility of minimalism about truth with truth-conditional semantics clearly and described it as follows:

---

[8]Note that Davidson himself did not the T-schema as his theory of truth. He only employed a compositional theory of truth in order to explain the compositional features of language.

[9]'P' is the name of an $\mathcal{L}$-sentence, and 'p' a sentence that articulates what the sentence says.

> if we accept [minimalism about truth]... we must abandon the
> idea... that the notions of truth and falsity play an essential rôle in
> an account of [meaning]... [Because if] someone should gain from
> the explanation that P is true in such-and-such circumstances
> an understanding of the sense of P, he must already know what
> it means to say of P that it is true. If when he enquires into
> this he is told that the only explanation is that to say that P is
> true is the same as to assert that P, it will follow that in order
> to understand what is meant by saying that P is true, he must
> already know the sense of asserting P, which was precisely what
> was supposed to be explained to him. (Dummett, 1959, 148f.)

In order to avoid the circularity of explaining meaning in terms of truth,
while explaining truth in terms of meaning, prominent defenders of truth
minimalism, such as Paul Horwich and Hartry Field, think their theories of
truth need to be supplemented by theories of meaning that broadly follow
the Wittgensteinian idea that the meaning of sentences has to be explained
by the conditions under which it is appropriate to *use* of them and not in
terms of conditions under which they are true.[10]

Hence, minimalists about truth hold that being disposed to accept all
(non-pathological) instances of (PT) is all there is to having the truth concept.
Furthermore, they agree that truth is a property, albeit only a logical one. So
it does not carry any heavy ontological commitment concerning the nature of
truth. This latter point is what makes minimalism about truth particularly
attractive to the quasi-realist expressivist qua anti-realist about moral facts.
Moreover, expressivism, as one form of use theory of meaning is suited for
combination with minimalism because it decidedly rejects the view that truth
has any significant role to play in the explanation of meaning.

So far we have found that minimalism about truth seems to require a
theory of meaning which does not explain meaning in terms of truth, but
rather in terms of use. It is important to note, however, that this does not
mean that meaning and truth are totally disconnected from each other. On
the contrary, meaning and truth are said to be conceptually or analytically
connected. This connection is usually captured by what John Collins calls
the 'truth from meaning principle' (Collins, 2002, 506):

(TM) If P (of $\mathcal{L}$) means that p, then P is true-in-$\mathcal{L}$ iff p.

According to Collins, this principle "expresses perhaps our central un-
derstanding of the relation between truth and meaning" (Collins, 2002,
506).

---

[10]See (Horwich, 1990) and (Field, 1986).

We are now in a position to see why minimalism about truth might help the expressivist with pushing forward his quasi-realist project. A fully worked-out expressivist theory of meaning specifies for each sentence of some language what it means. Each theorem of this theory will have the form of the antecedent of (TM). This means that for each sentence of the language we can immediately infer the truth-condition of that sentence. What is most important from the perspective of metaethical expressivism, is that this will hold irrespective of whether a sentence is moral or not. For instance, if an expressivist theory of meaning is able to explain how 'murder is wrong' means that murder is wrong, we are justified in saying that 'murder is wrong' is true iff murder is wrong. Since the minimalist expressivists holds that truth is minimal, nothing metaphysically robust is introduced to their metaethical theory by allowing that moral sentences have truth-conditions.

Expressivists who embrace minimalism about truth as the point of departure for their quasi-realist project usually hold that once a minimal notion of truth is established, this allows them not only to freely speak of moral truth, but also of moral *facts*, moral sentences as representing or being *about* those facts, moral predicates denoting moral *properties*. Here is, for instance, Blackburn:

> Because of minimalism we can have for free what look[s] like a ladder of philosophical ascent: 'p', 'it is true that p', 'it is really and truly a fact that p' ..., for none of these terms, in Ramsey's view, marks an addition to the original judgement. You can as easily make the last judgement as the first—Ramsey's ladder is lying on the ground, horizontal. ... From its top there is no different philosophical view than from the bottom, and the view in each case is just, p. (Blackburn, 1998, 295f.)

I do not want to provide detailed accounts of how minimalism about truth is thought to be extended so that it also provides minimalist notions of facts, representation, and others. A more detailed discussion of how such an extension might proceed is provided in (Sinnott-Armstrong, 2006, 27f) and (Tiefensee, 2010, Ch. 2). Perhaps such an extension for some of those notions can be successful, but I will not pursue these matters further, especially since the point of my discussing minimalism is not so much to give an understanding what it does achieve for the expressivist, but what it does not achieve.

On the positive side, the forgoing discussion suggests two things. First, since expressivism is a theory of meaning which does not explain meaning in terms of truth, but in terms of use, it appears in principle to be compatible with minimalism about truth. Second, once expressivists have completed their theory of meaning, they can embrace minimalism about truth and are able to assign minimal truth-conditions to all sentences of their language. In

fact this will be a trivial matter because once we know what a particular sentence means, we can immediately infer, via (TM) or (PTM), its truth condition. Since truth is minimal, saying that sentences have conditions under which they are true places no metaphysical burden on the minimalist quasi-realist expressivist. Thus, minimalism about truth is a step forward in realizing the quasi-realist project.

Nevertheless, the previous discussion also showed that there are important things that minimalism about truth does not deliver, but rather presupposes. Most importantly, it emerged that by embracing minimalism about truth, expressivists are not released from the task of developing a theory of meaning. On the contrary, a theory of meaning has to be developed *before* expressivists can enjoy the fruits of truth minimalism. Minimalism about truth complements expressivism, but it does not help with, let alone, solve their semantic and logical problems about how sentences come to mean what they do and how they can figure in logical arguments as they do. Truth, from the perspective of minimalism, is simply too minimal to do any real work in explaining meaning and logic.

One might think at this point that minimalists have to reject the common view that to know the meaning of a sentence is to know under which conditions it is true. I think what they have to reject about this view is the usual understanding of it. The usual understanding is that the view states an equivalence, namely that one knows that P means that p iff one knows that P is true iff p, where this is supposed to provide an analysis of 'means that' in terms of truth conditions. Minimalists should reject this order of the analysis, and claim instead that the equivalence provides an analysis of 'is true under such-and-such conditions' in terms of meaning. We can signal this difference of analytical focus by saying that to know under what conditions a sentence would be true is to know its meaning, instead of saying that know the meaning of a sentence is to know under what conditions it would be true. Minimalists should think of 'is true under' as the *analysandum* and 'means that' as the *analysans*.

4.3 Truth-Aptness

4.3.1 The Plan

Even though in the past metaethical expressivism has often been defined as a view about the truth-aptness of moral sentences, namely as the view that moral sentences are not truth-apt, truth-aptness has not quite received the attention it should have received in the literature on expressivism. In the following I want to fill this lacuna.

The overarching goal of this chapter and the next is to develop in detail a novel way of how expressivists can establish two central claims of their quasi-realist project, namely:

Moral Truth-Aptness: Moral sentences are truth-apt, and

Moral Belief: Moral sentences express beliefs.

There is a dominant strategy of argumentation for these two claims in the literature. The strategy is to argue first for the truth-aptness of moral sentences, and in a second step argue from the truth-aptness of moral sentences to their expressing beliefs. Call this the 'orthodox strategy'. A typical argument exemplifying the first step in the orthodox strategy for the truth-aptness of moral sentences can be found in (P. Boghossian, 1990) and (Wright, 1992) and has the following schematic form:

(1) A sentence is truth-apt iff, and because, it is such-and-such (theory of truth-aptness)

(2) Moral sentences are such-and-such. (alleged fact)

(3) Hence, moral sentences are truth-apt. (from 1, 2, via modus ponens)

The second step of the strategy is to extend this argument as Neil Sinclair has done in (Sinclair, 2006):

(4) If a sentence is truth-apt, then the state it expresses is a belief. (minimalism about belief/Wright's tie[11])

(5) Hence, moral sentences express beliefs. (from 3, 4, modus ponens)

Originally, instantiations of these arguments, (1)-(3) and (3)-(5), were proposed as *problems* for expressivism because its authors had old-school, non-quasi-realist expressivists in mind which were taken to reject (3) and (5). I am interested, however, in the question of how good these arguments are for the contemporary quasi-realist expressivist who, unlike his old-school predecessors, does not reject (3) and (5), but instead wants to use these arguments in order to *establish* (3) and (5), that is Moral Truth-Aptness and Moral Belief. I will argue that the usual ways of making those arguments do not help expressivists establishing (3) and (5) because they rest on mistaken theories of truth-aptness, that is they rest on false instantiations of (1).

The primary aim of this chapter is to propose a particular analysis and explanation of truth-aptness. In other words, I aim to justify a particular filling-out of the 'such-and-such' in (1). The analysis and explanation I shall defend is, roughly, the following:

> *Belief explains Truth-Aptness*: A sentence is truth-apt iff and because (i. e. in virtue of the fact that) it expresses a belief.

---

[11]I will explain what minimalism about belief and Wright's tie are below.

Given the explication of the expressing relation in terms of sincerity conditions defended in chapter 2, the thesis that belief explains truth-aptness is equivalent to the following:

> *Assertion explains Truth-Aptness*: A sentence is truth-apt iff and because there is a conventional way of using it to make an assertion.

Since commonly a sentence is defined as being assertoric iff it is truth-apt, we can also capture the proposal of this chapter in the catchier slogan that *sentences are assertoric when they can be asserted.*

Ultimately, I shall argue, expressivists should be making roughly the following argument for the truth-aptness of moral sentences, which I call the 'belief-to-truth-aptness argument':

(6) If a sentence can be used to make assertions (used to express a belief), then it is truth-apt.

(5) Moral sentences can be used to make assertions (express beliefs).

(3) Therefore, moral sentences are truth-apt.

The important consequence of my discussion of truth-aptness will be that expressivists need to invert the orthodox strategy of arguing from the truth-aptness of moral sentences to their expressing beliefs. In order to make the belief-to-truth-aptness argument they need to proceed the other way round. They need to argue from the expression of moral belief to the truth-aptness of moral sentences. This puts the usual quasi-realist dialectic upside-down and has important consequences. On the one hand it makes establishing that moral sentences are truth-apt and express beliefs more difficult than expressivists have previously though. On the other hand, if the new strategy can successfully be carried out, it will put the expressivits' quasi-realist claims about moral truth-aptness and moral belief on much more solid ground precisely because of this difficulty.

The chapter is structured as follows. I start by making some remarks about why expressivists need a theory of truth-aptness. After that I do some groundwork by explaining what a theory of truth-aptness is, what different types of theories of truth-aptness one should distinguish, and I explain which kind of theory of truth-aptness I think we should defend. Next I discuss Paul Boghossian's argument that minimalism about truth implies some particular kind of theory of truth-aptness, namely so-called disciplined syntacticism. I then present two arguments that have been based on disciplined syntacticism: the Boghossian-Wright argument for moral truth-aptness and the moral belief argument (also known as the moral belief problem) by Neil Sinclair. Then I present some objections against disciplined syntacticism, as well as

against its successor theory: Allan Gibbard's disagreementism. Then we will have a look at Michael Smith's platitudes-respecting theory of truth-aptness, and his truth-inaptness argument against expressivism. Careful examination of some objections that have been put forward against Smith's argument will be instructive insofar as this will sharpen our understanding of truth-aptness and it will provide the idea for my own theory of truth-aptness. Finally, I will put everything together and present a more precise version of the belief-to-truth-aptness argument.

Two features might make following some of the arguments in this chapter difficult, so I want to point them out. First, most of the arguments discussed here have originally been presented with old-school expressivism in mind. Old-school expressivists held that moral sentences are *not* truth-apt and do *not* express beliefs because they took this to follow rather straight forwardly from their claim that moral sentences express noncognitive states. Contemporary expressivists, however, want to establish exactly the opposite and try to reconcile noncognitivism with truth-aptness and moral belief. Consequently, arguments that have originally been put forward against (or in favor of) old-school expressivism, now must be assessed as arguments in favor of (or against) contemporary expressivism. The other feature that might make it hard to keep track of the overall argumentation in this chapter is that I am primarily arguing for an inversion of the order of explanation between belief and truth-aptness. Consequently, some of the claims I reject will look very similar to claims I accept. For instance, ultimately, I agree with the minimalist about belief that every state expressed by a truth-apt sentence can be regarded as a belief, but I do not think of this as an analysis of belief, as the belief minimalist thinks, rather I think of it as an analysis of truth-aptness.

In this chapter a number of notions take center-stage. These notions will be truth, truth-aptness, proposition, assertion, belief, and sentences. Once we inquire into these notions, we quickly lose our sense for what is fundamental and what is derived. Many of these notions seem interdefinable and even inter-explainable. So it may appear that no definition or explanation is unquestionably better than the other. For instance, can propositions be true or false because they are the objects of propositional states such as belief? Or can propositions be the objects of propositional states because they can be true of false? Can sentences be asserted because they express beliefs? Or do they express beliefs because they can be asserted? My proposal here can be seen as one particular way of drawing a coherent and non-circular picture of the relations between some of these notions, but I do not have an argument that it is the only coherent picture or even the correct picture. Nevertheless I think that the proposal indeed draws a pretty good picture, and at least it is expressivist-friendly in the sense that it treats only pragmatic material as fundamental, namely linguistic signs, language use, conventional norms, and speech acts such as assertions. However, even though this proposal

is expressivist-friendly it attempts to not make it too easy for something
to count as truth-apt because this would be inadvisable from a dialectical
standpoint. If something can be achieved too easily, then it is not clear that
it is really an achievement.[12] I take it to be a point that speaks in favor
of my analysis of truth-aptness, that it requires expressivists to develop a
theory of belief *before* they can conclude that moral sentences are truth-apt.

Overall this chapter makes two new contributions to the debate. It offers
a novel theory of truth-aptness in terms of use, and it shows that expressivists
cannot establish the truth-aptness of moral sentences unless they already
have a theory of belief in general, and of moral belief in particular. This
chapter deals with truth-aptness. Developing a theory of belief will then be
the task for the next chapter.

4.3.2   Expressivists Need a Theory of Truth-Aptness

Do expressivists need a theory of truth-aptness, and if so, why? Not everyone
believes that expressivists need a theory of truth-aptness. For instance, Mark
Richard has argued that expressivists need not care about truth-aptness
once they have a solution to the Frege-Geach problem, that is once they
have explained the semantical and logical properties of their target language
(Richard, 2015). Having such a solution, it is argued, would contain, *inter
alia*, an explanation of the semantic and logical properties of 'true' and
sentences that contain it. Nothing more would be needed.

I am sceptical that this really frees expressivists from the need of having
a theory of truth-aptness. Having a semantics and logics for 'true' does
not tell us to which expressions it can properly be applied to. We need a
theory that tells us to which expressions 'true' can properly be applied to,
before we can even start explaining the semantical and logical properties of
'true'-sentences. For this reason, expressivists need a theory of truth-aptness
even if they are ultimately able to solve the Frege-Geach problem. Moreover,
a theory of truth (or of the concept of truth or the truth predicate) by itself
also does not tell us to which entities truth can properly be applied. Even if
we know that applying 'true' to some expression to which it can be applied
is in some sense equivalent to that expression itself (as minimalists about
truth claim), does not bring us in a position to know to *which* expressions it
can be applied. So, we need to know what sorts of thing we can plug into
the Tarskian equivalence schema, (T) or (PT) from above, before we can
accept any instance of it. This is also noted by John O'Leary-Hawthorne:

---

[12]This echoes a common objection to the expressivists' quasi-realist projects, and also
to minimalism about truth, namely that expressivists set themselves the tasks so easy that
it is not really clear that anything has been achieved.

> When confronted with the disquotational schema ['P is true' iff
> p], one might reasonably inquire as to the proper range of its
> instances. What English sentences can be plugged in? It is
> because this question demands an answer that we can be certain
> that the disquotational schema doesn't tell us all we need to know
> about truth... (O'Leary-Hawthorne, 1994, 215)

Hence, expressivists qua minimalists about truth need a theory of truth-aptness because without a principled way of saying which things can be substituted into the T-schema to yield sentences such as 'it is true that murder is wrong', expressivists cannot make use of minimalism about truth.

Another reason why many expressivists need, or at least think they need (I shall argue against this below), a theory of truth-aptness is that many expressivists want to embrace the following view about belief:[13]

> *Minimalism about Belief*: any state M is a belief iff M is expressed (or expressable) by a truth-apt sentence.

For the minimalist theory of belief to get off the ground, a theory of truth-aptness is obviously needed.[14]

The most important reason why expressivists need a theory of truth-aptness is that moral sentences *appear* to be truth-apt: we commonly ascribe truth and falsity to moral sentences and we do not regard this practice as somehow defective. Quasi-realists want to explain why we are justified in engaging in this practice. The most satisfying explanation would of course be that they appear to be truth-apt because they actually *are* truth-apt. I see two ways in which quasi-realists could try to show that moral sentences actually are truth-apt. One would be by providing an account that explains, on very general grounds, why certain expressions can be regarded as truth-apt that is compatible with expressivism's overall commitments. The other way would be a more formal approach: define a language, stipulate to which expressions of the language a truth-predicate can properly be applied to, and then give a semantics for the language and the truth-predicate. According to the first approach whether or not some expression is truth-apt depends on general features of the expression and is something that can be discovered. According to the formal approach an expression counts as truth-apt simply in virtue of the fact that we as semantic theorists assign some truth-value to it.

---

[13]Minimalism about belief is not always explicitly embraced, but is implicit in the accounts of many writers on expressivism. E. g. it is often assumed that an expressivist will give a semantic clause for 'believes that' along the following lines: 'S believes that p' means the same as 'S is in the mental state expressed by p'. Cf. (M. Schroeder, 2008a).

[14]Since I will below argue that truth-aptness must be explained in terms of belief, rather than vice versa, it should be clear that expressivists who will accept my arguments cannot simultaneously be minimalists about belief, since they would then be caught up in a explanatory circle. So in this case minimalism about belief no longer can be regarded as a reason for the expressivists' need for a theory of truth-aptness.

I think both ways of proceeding might be fruitful, but I shall pursue the first way instead of the second here. The reason for this choice is that metaethicists are primarily interested in the properties of the languages we actually speak, like English. Answering the question whether the moral sentences of English are truth-apt by making a stipulation seems unsatisfactory. Whether the moral sentences we actually use are truth-apt or not is normally regarded by expressivists an important philosophical question that should not be answered by fiat.

4.3.3   Theories of Truth-Aptness

Now that we know that and why expressivists need a theory of truth-aptness, we should be clearer on what a theory of truth-aptness is. The task of a theory of *truth*, roughly, is to tell us what truth is or what it takes to understand the truth predicate or to have the truth concept. But what is the task of a theory of *truth-aptness*? The task is not to provide a definition of truth-aptness, since the definition of truth-aptness is fairly uncontroversial. The following definition is generally accepted:

> *Truth-Aptness*: Something is truth-apt if and only if it *can* be either true or false.[15]

Proponents as well as opponents of the view that moral sentences are truth-apt all agree that this is how truth-aptness is defined.[16] This can be read off from Ayer's understanding of emotivism which includes the claim that moral sentences say nothing which "*can* be either true or false" (Ayer, 1952, 107, italics added). Similarly, Paul Boghossian describes emotivism as the view that moral sentences are such that "nothing about the world *could* render them true or false" (P. Boghossian, 1990, 160, italics added).

---

[15]Perhaps a more general definition is that something is truth-apt iff it can have some truth-value.

[16]Richard Holton is an exception (Holton, 2000). Holton holds that an expression is truth-apt iff it *is* either true or false. I do agree with the right-to-left direction: if something is true or false, then it can be true or false. Hence, it will be truth-apt. But that something is truth-apt does not imply that it is actually either true or false. A sentence containing a vague or gappy predicate may on some occasion be neither true nor false, but that is not sufficient for claiming that the sentence is not apt for being either true or false. If Heinz Erhardt is a borderline case of baldness, 'Erhardt is bald' is neither true nor false. (If you are an epistemicist about vagueness, you will deny this conditional. But in that case you might want to replace the example with a view according to which, for instance, 'The present king of France is bald' lacks a truth-value in the actual world.) Nevertheless, 'Erhardt is bald' could be true or false, namely, if Erhardt were either clearly bald or clearly not bald. Hence, even though 'Erhardt is bald' may lack a truth-value, it could have one. If a sentence could *in principle* bear a truth-value, then it should qualify as truth-apt. I ignore the controversial question whether sentences such as the Liar should be regarded as truth-apt.

But if it is uncontroversial that 'being truth-apt' means 'being able to be true or false', then what is the task of a theory of truth-aptness? In my view the central task of a theory of truth-aptness is *explanatory*.[17] It tells us what kinds of entity can be assessed in terms of truth or falsity, and what about these entities makes them apt or suitable for being assessed in terms of truth or falsity. In other words, it is a theory that tells us of some particular type of entity the properties *in virtue of which* that type of entity is suited for bearing a truth-value. The 'in virtue of' is supposed to highlight that this theory not only states necessary and sufficient conditions for truth-aptness, but also offers an *explanation* of why the things that are truth-apt are truth-apt. If one wishes, one could equally say that a theory of truth-aptness, as I understand it here, tells us what properties *ground* or *constitute* something's aptness for being assessed in terms of truth and falsity. Therefore, I take a theory of truth-apt to have the following form:

Theory of Truth-Aptness: for all X, X is truth-apt *because* X is such-and-such.

I make the following assumption about the relation expressed by 'because'. The relation is asymetrical and satisfies the following: If A because B, then necessarily (if B, then A).[18] This because-relation is the one that also holds in cases where we say things like that the singleton-set {Sokrates} exists because Sokrates exists. Making these assumptions allows me two things. First, due to the asymmetry we cannot explain belief in terms of truth-aptness (minimalism about belief) if we explain truth-aptness in terms of belief. The second property allows me to go from 'something is truth-apt because it is such-and-such' to 'if something is such-and-such, then it is truth-apt'. This latter implication is what I need for my final truth-aptness argument at the end of this chapter.

Theories of truth-aptness can be distinguished along several dimensions. For instance, they can be about different types of truth-bearers, they can be minimal or robust theories of truth-aptness, they can be conservative or radical, monist or pluralist, and they can be internalist or externalist. The theory of truth-aptness I shall propose below is a theory that regards linguistic expressions (well-formed formulas) as truth-bearers, shuns talk of 'minimal' or 'robust', is conservative, monist, and externalist. I shall briefly discuss what I mean by these terms now.

---

[17]A similar idea can be found in the work of Huw Price who distinguishes between explanatory and analytic accounts of truth-aptness, or as he says, 'statementhood'. See (Price, 1988, 2-4).

[18]Cf. (de Rizzo, 2020).

Types of Truth-Bearers

Before inquiring into which property constitutes something's truth-aptness it is sensible to ask what kinds of thing can be truth-apt or be bearers of truth. Several candidates naturally spring to mind: propositions, sentences, speech acts such as assertions, or mental states such as beliefs. It is controversial among philosophers which of these are the primary bearers of truth, and so what kind of thing is to be regarded as the ultimate possessor of truth-aptness and which things are truth-apt only derivatively.[19] Since expressivists are primarily concerned with the question of whether moral sentences, that is linguistic expressions, are truth-apt or not, and not so much with the question of whether moral propositions, beliefs, or other types of potential truth-bearers are truth-apt, my focus will be exclusively on the truth-aptness of linguistic expressions. In particular my focus will be on the truth-aptness of well-formed formulas (short: formulas). I will use the more general term 'formula' and reserve the term '(declarative, indicative, or assertoric) sentence' for those formulas that are truth-apt. This follows common usage in logic textbooks that define a sentence as any formula that can be either true or false, or more generally, that can bear a truth-value. Defined this way, the question 'is this sentence truth-apt?' is trivial, since nothing can be a sentence and not be truth-apt, whereas the question 'is this formula truth-apt?' is non-trivial. Hence, another way of asking what makes a formula truth-apt is asking what makes a formula a sentence.

Internalist vs. Externalist

In saying that a theory of truth-aptness needs to tell us the properties in virtue of which some entity is a truth-bearer, I do not mean to suggest that these properties must be internal or essential to their nature. For instance, it is often claimed that propositions are the kinds of thing that are necessarily truth-bearers: a thing that cannot be true or false cannot be a proposition, since propositions can be true or false by definition. It will become clear that I do not think that whether or not some formula is a truth-bearer is due to features internal to it. Instead, I think that what makes something truth-apt are features inessential or external to them, such as our rule-governed linguistic use of it, a use it has as a matter of convention. The question is what the linguistic conventions must be like so that making some formula subject to them lets it qualify as a truth-bearer.

---

[19]Richard Kirkham provides a more comprehensive list and discussion of what kinds of things can be bearers of truth, and advocates a very 'tolerant' view, namely that "there are no restrictions *in principle* on what kinds of entities can possess truth or falsity" (Kirkham, 1995, 59). According to Kirkham, all things can be bearers of truth—even teddy bears! (Kirkham, 1995, 61).

Conservative vs. Radical

Any theory of truth-aptness must provide a criterion by which we can systematically tell apart the truth-apt from the non-truth-apt formulas. A reasonable additional constraint of adequacy on a theory of truth-aptness is that it classifies all and only the intuitively truth-apt formulas as truth-apt. Following O'Leary-Hawthorne we can call theories of truth-aptness that respect this latter constraint *conservative*, and theories that ignore it *radical* (O'Leary-Hawthorne, 1994). For instance, intuitively 'murder is wrong' is a truth-apt formula of English, whereas 'Ouch' or 'Close the window!' are not. A conservative theory of truth-aptness will yield the intuitively right predictions about such cases, whereas Ayerian emotivism, as a radical theory, may be happy with classifying 'murder is wrong' as not truth-apt, though intuitively it is.[20] I assume that the expressivists' quasi-realism commits them to a conservative theory of truth-aptness.

Minimalist vs. Robust

Theories of truth-aptness can be minimal or robust. I said above that minimalists about truth need a theory of truth-aptness. But what kind of theory? Minimalists about truth tend to classify the theories of truth-aptness they embrace also as 'minimal'. Unfortunately, it is rarely made precise what makes a theory of truth-aptness or the property of truth-aptness 'minimal', and it is not even clear that 'minimal' in 'minimal truth' means the same as in 'minimal truth-aptness'. For instance, sometimes the truth-aptness of some formula is described as minimal because it comes 'for free', for instance with certain surface features of the truth-apt formula, or can be established in an 'easy' (Jackson, Oppy, & Smith, 1994, 291) or 'fairly trivial' (O'Leary-Hawthorne, 1994, 216) way. Other authors classify every formula as minimally truth-apt if it is not 'strictly speaking' truth-apt (Lenman, 2003a, 32). A more natural and more illuminating understanding is to take 'minimally truth-apt' to mean 'apt for minimal truth'. This seems to be the understanding Boghossian has in mind in making his incompatibility argument (s. below). A really general understanding of 'minimal' is provided by Michael Lynch. He defines a property as minimal iff it is not "substantive" or "metaphysically transparent", where a property is metaphysically transparent iff "one can know all the essential facts about the property just by grasping the ordinary folk concept of that property" (M. P. Lynch, 2009, 2). The problem I see with this understanding of being a minimal property, however, is that, unlike with the concepts of truth or belief, I am sceptical that there really is something like an 'ordinary folk concept' of truth-aptness at all.

---

[20]The scope of the conservativeness constraint is limited, however, because it makes sense to speak of intuitively truth-apt formulas only in the context of natural languages, that is languages we can have sufficiently robust intuitions about.

Perhaps what unites minimalists about truth-aptness is not a shared view about what minimal truth-aptness is, but a view about what minimal truth-aptness is not. The common denominator seems to be their view that something can be truth-apt without being apt for robust truth in the sense of correspondence with (robust) facts. In my view, however, the diversity of conceptions of 'minimal' suggests that it is not a very illuminating or useful notion. For this reason I suggest that we avoid talk of minimal truth-aptness.

Another reason why we should avoid talk of 'minimal' truth-aptness is dialectical. Saying that something is minimally truth-apt may easily give rise to the impression that it is not *really* truth-apt after all. This is a problem all minimalist accounts have, be it of minimal truth, minimal truth-aptness, minimal belief, minimal proposition etc. The best way to get rid of this unwanted impression and the confusion different understandings of 'minimal' may cause, is to shun the label 'minimal' in the context of a theory of truth-aptness, and evaluate some particular theory of truth-aptness by what it actually says, and not by the label under which it is marketed.

Monist vs. Pluralist

Another important distinction is that between monist and pluralist theories. This distinction can be drawn with respect to many theories. For instance, monists concerning *truth* hold that there is only one kind of truth, whereas pluralists hold there is more than one kind of truth.[21] Most contemporary minimalists about truth, such as Horwich, are monists, rather than pluralists: they think that minimal truth is the *only* kind of truth, and that there is no 'real' truth besides it, since minimal truth is real truth. However, others, such as (M. P. Lynch, 2009) hold that some areas of discourse such as ethics, aesthetics, or mathematics allow only one kind of truth (perhaps minimal), whereas others such as physics allow for another kind of truth (perhaps robust).

Just as we can be monists or pluralists concerning truth or belief, we can be monists or pluralists concerning *truth-aptness*. Truth-aptness monists hold that there is only one kind of truth-aptness, whereas pluralists hold that there is more than one kind. Most so-called minimalists about truth-aptness are monists, rather than pluralists: they think that minimal truth-aptness is the only kind of truth-aptness. However, a minimalist about truth-aptness may also hold that some areas such as ethics etc. allow only for minimal truth-aptness, whereas others allow for more robust kinds of truth-aptness.[22]

---

[21]The same distinction can be drawn with respect to belief. For instance, some hold that quasi-realists expressivists can make sense of moral beliefs if they distinguish between different types of belief, e. g. minimal and robust beliefs, and thus by being pluralists concerning belief. See (Sinclair, 2006) below.

[22]One must be careful not to lump together the minimal-robust distinction and the monist-pluralist distinction. One can be a pluralist concerning truth-aptness, but hold that all kinds of truth-aptness are minimal or that all kinds of truth-aptness are robust, for example.

Monism and pluralism each have their own advantages and disadvantages. For instance, one may have the intuition that there is truth in ethics as well as in physics, but that the truths of physics are somehow more firm than those of ethics. For someone with this intuition it is advisable to be a pluralist regarding truth. On the other hand being a pluralist might also be considered a cost, because it raises problems the monist does not face. One such problem is that a pluralist has to tell us for different areas of discourse why the relevant formulas in them are truth-apt in one sense, but not in another, and they have to tells us where exactly the boundaries are.[23]

I do not want to suggest that expressivists cannot be pluralists about truth-aptness. But *prima facie* it more advisable to have a monist theory of truth-aptness, than to have a pluralist one. Being a monist has the clear benefit of avoiding the above problem of explaining why there are boundaries for certain notions of truth-aptness and where they are. Moreover, monism generally does not give rise to the impression that there are first-class notions and second-class notions of truth, truth-aptness, beliefs and so on, and that in some areas of discourse, such as ethics, we have a right only to the second-class notions.

Relative vs. Absolute

As with many other properties one may think that the property of truth-aptness is absolute or relative. Even though I think it should be obvious, it is rarely noted that truth-aptness, like truth, is a language-relative, or language-dependent property. It does not make sense to ask if 'Schnee ist weiss' is truth-apt independently of a language. Likewise, it does not make sense to ask if 'Faki si konori' is grammatical, meaningful, or true unless we are told in which language. We could here allude to Frege and say that only in a context of a language can a formula have a truth-value.

4.3.4   Boghossian-Wright Argument for Truth-Aptness

With all these distinctions in place we can now turn to the question of this chapter: what makes something apt for truth? The notion of truth-aptness featured centrally in the debate about whether expressivism is compatible with minimalism about truth. So it is worth reconsidering this debate, and critically assess the proposals that have been made there.

---

[23]A related problem is how to account for logically complex formulas some components of which are truth-apt in one sense while others are truth-apt in another sense. For instance, if 'grass is green' is truth-apt in a different sense than 'murder is wrong' is truth-apt, what do we say about their conjunction? I do not mean to suggest that this is an unsolvable problem, but any solution will require making certain important assumptions about the relevant notions of truth-aptness, like the assumption that at least one notion needs to be a subset of the other. For instance, one has to assume that every robustly truth-apt formula is also minimally truth-apt.

A problem with discussing the arguments from the compatibility debate is that all authors in it operated under the old-school premise that expressivism implies or contains the view that moral sentences are *not* truth-apt. However, contemporary quasi-realist expressivists, as I have already mentioned a couple of times, reject this view and instead want to establish and justify precisely the opposite claim, namely that moral sentences *are* truth-apt. As a consequence the complete debate has to be turned upside-down and all the arguments made in it appear in a new and interesting light.

The compatibility debate originated from an argument made by Paul Boghossian in (P. Boghossian, 1990). Boghossian argues for the claim that minimalism about truth is incompatible with the old-school expressivists' claim that moral sentences are not truth-apt. His argument is roughly as follows. Minimalism about truth, he claims, implies a certain form of minimalism about truth-aptness, namely disciplined syntacticism (a view to which we will turn shortly). This form of minimalism about truth-aptness makes it easy, Boghossian thinks, for moral sentences to be truth-apt. This, however, directly contradicts the old-school expressivists central claim. Hence, expressivism is incompatible with minimalism about truth-aptness, and thus also incompatible with minimalism about truth.[24]

Boghossian's incompatibility argument was intended as a argument against expressivism. However, from the perspective of the contemporary quasi-realist the argument gets a different heading. What originally was an argument against the claim that moral sentences are not truth-apt, can now be adopted by expressivists as an argument for the claim that moral sen-

---

[24]Actually, Boghossian's argument is richer than the one I give in the main text. Since his additional points are not relevant to my argument, I shall only mention his additional points in this footnote. Since Boghossian thinks that expressivists deny that moral sentences are truth-apt, he concludes that they must have a more robust notion of truth-aptness in mind than the minimal notion of truth-aptness implied by minimalism about truth. According to Boghossian, the old-school expressivist must presuppose a conception of truth-aptness, and thus of truth, that is "richer than the deflationary [i. e. minimalist]" (P. Boghossian, 1990, 165). One way of responding to this line of argument would be to defend pluralism about truth or truth-aptness. If one pursues this strategy, then one will argue that there is more than one notion of truth-aptness, for instance truth-aptness$_1$ and truth-aptness$_2$. The pluralist can then claim that some formulas, for instance moral ones, are truth-apt in the sense of truth-apt$_1$, whereas others, non-moral formulas say, are (also) truth-apt$_2$. This pluralist strategy is implicit in early emotivist views such as Ayer's, and recently (Lenman, 2003a) and (Stoljar, 1993) explored to possibilities of pursuing the pluralist truth-aptness strategy explicitly. Boghossian, however, vehemently denies the possibility of being a pluralist regarding truth or truth-aptness (cf. (P. Boghossian, 1990, 165, Fn.17), and so, he thinks, expressivists cannot hold on to their claim that moral sentences are not truth-apt, if they want to be minimalists about truth, and thus minimalists about truth-aptness. As I mentioned above, I, like Boghossian, set aside the pluralist strategy. I do so not because I think it is impossible to defend a pluralist theory, but because I think that in general a monist strategy is preferable. Being a monist avoids a lot of problems the pluralist faces, such as the problem of justifying every single notion of truth-aptness, and the problem of the truth-aptness of compound formulas some parts of which are truth-apt$_1$ whereas others are truth-apt$_2$.

tences are truth-apt. This changes the whole dialectic of the incompatibility debate. Expressivists no longer need to try to block Boghossian's argument, for instance by attacking its premises. Instead, they should try to defend it by defending its premises and answering objections to it.

The aim of the next section is to show that, unfortunately for the expressivist, Boghossian's argument is not a good argument for the claim that moral sentences are truth-apt, since the premises on which it rests are mistaken.

Grounding Truth-Aptness in Truth

Where shall we start to look for criteria for a formula's being truth-apt? A natural and promising idea is that the property or concept of truth itself might tell us something about what things have to be like for them to be apt for truth. Boghossian formulates something like this idea in the form of a constraint. He says that:

> Any proposed requirement on candidacy for truth must be grounded in the preferred account of the nature of truth. (P. Boghossian, 1990, 165)

Walter Sinnott-Armstrong similarly claims that a "minimal theory of truth ... puts constraints on which sentences can be true or express truths" (Sinnott-Armstrong, 2006, 21). One might actually think that truth-aptness not only must be grounded in truth, but that a theory of truth straightforwardly delivers or implies a theory of truth-aptness. Consider, for instance, what a robust theory of truth seems to imply about truth-aptness. According to a simple correspondence theory of truth, truth is the property of corresponding to how the world is. To which things can this property be applied? The natural answer is: only to things that can correspond to how the world is. One way of spelling out what it is for something to correspond to the world is to say that it describes or represents a way the world might be and the world actually is that way. This way of spelling out the correspondence theory implies, or at least strongly suggests, that something cannot be true or false, that is, is not be truth-apt, if it does not describe a way the world might be. So the correspondence theory of truth by itself seems to provide a criterion for something's aptitude for truth.

But what if the preferred account of the nature of truth is not a robust one, but minimalism about truth? There appears to be a problem. If minimalism about truth is taken to be the view that truth is not a robust property or not even a property at all, how could truth itself provide us with any criteria for which things it can properly be applied to? How could minimal truth tell us something about what formulas need to be like for them to be able for being true or false? We cannot ground truth-aptness

in the nature of truth, if, as minimalists about truth tell us, "truth has no nature" (M. P. Lynch, 2009, 5).

However, even if many minimalists about truth accept the negative claim that truth is not a property or does not have a 'nature' in which truth-aptness can be grounded, most of them also accept some positive claim to the effect that truth has a *function*. The minimalist about truth may modify Boghossian's grounding constraint accordingly and claim that truth-aptness must be grounded not in the nature, but in the function of truth.

What could this look like? As I have noted in the section on truth, minimalism about truth is a family of theories of truth that include different positive claims. So perhaps not every version of minimalism about truth is equally well placed to ground truth-aptness in it.[25]

Boghossian, however, thinks that minimalism about truth grounds some specific minimal theory of truth-aptness, and he has a particular version of truth minimalism in mind which one could call 'assertion minimalism about truth'. According to assertion minimalism about truth, the function of truth or 'true' is that it is a device for assertion: 'true' allows us to make assertions about the world by making assertions about sentences (semantic ascent). This 'functional' part of this form of minimalism about truth indeed tells us something about truth-aptness, Boghossian thinks:

> Since, [according to minimalism about truth], the overall effect of asserting that a sentence is true is just to assert the sentence itself, the requirements on truth predication [i. e. truth-aptness] must include whatever requirements attend candidacy for assertion itself (P. Boghossian, 1990, 164, italics added)

I think that this conception is pretty plausible, especially when one accepts that truth-aptness must be grounded in the function of truth, and one thinks that the primary function of truth is assertion.[26]

What are the requirements for candidacy for assertion? Boghossian immediately provides an answer, and it is this answer that, in my view, has led many philosophers in the compatibility debate into a wrong direction.

---

[25]For instance, those minimalists who merely claim that to know what truth is to know that for any substitution of an appropriate formula into (T) that instance of (T) holds. It is not clear that the 'function' truth has according to this form of minimalism can be spelled out in such a way that truth-aptness can be grounded in it. The problem is that this form of minimalism about truth by itself does not tell us what the formulas have to be like to be appropriate for being substituted into (T).

[26]Note that one need not be a minimalist about truth to find Boghossian's proposal plausible. Even a correspondence theorist of truth might think that requirements on truth predication must include requirements for assertion, if she also thinks that in addition to truth being correspondence, say, 'true' is a device for assertion.

According to Boghossian:

> clearly, a sentence must be both meaningful and assertoric in
> form if it is to be a candidate for assertion. (P. Boghossian, 1990,
> 164)

Slightly more detailed Boghossian says that a formula is truth-apt only
if it satisfies two "minimal requirements":

> first, [it] must be significant, and second, it must be declarative
> in form. Unpacking somewhat, the requirements are that the
> [formula] possess a role within the language: its use must be
> appropriately *disciplined* by norms of correct utterance; and that
> it possess an appropriate *syntax*: it must admit of coherent em-
> bedding within negation, the conditional, and other connectives,
> and within contexts of propositional attitude (P. Boghossian,
> 1990, 163, italics added)

Boghossian's conception of truth-aptness has become quite popular and
is now known as *disciplined syntacticism*, a label invented by James Lenman
in (Lenman, 2003a). I will have a closer look at Boghossian's syntax and
discipline requirements in the next section. For now let us just note that, as
they are stated here, these requirements formulate only necessary conditions
on truth-aptness. However, Boghossian thinks that "these requirements would
seem also to be *jointly sufficient* for truth conditionality, on a deflationary
understanding of truth" (P. Boghossian, 1990, 164). In his view they are
jointly sufficient because it is difficult to see "how any ... further requirement
would be motivated [if one accepts minimalism about truth]" (ibid.).

His idea seems to be something like this: if 'true' expresses not real prop-
erty, but is only a device for making possible certain assertions (semantically
ascending and making assertions about the world by making assertions about
sentences), then a formula that has all the properties a formula needs to have
in order to be a candidate for assertion, seems to also have *all* the properties
a formula needs to have so that 'true' can be applied to it and thus serve
the purpose for which we have it, i. e. making certain assertions possible.

I am inclined to agree with Boghossian's view that truth-aptness needs to
be grounded in truth, and I am even more inclined to accept it if we think of
it as the claim that the function of truth justifies our account of truth-aptness.
It seems reasonable that anything that explains how something can serve
the main purpose for which we have the truth-predicate is sufficient for
explaining how this something is apt for being true or false.

Since we are now familiar with Boghossian's view about truth-aptness,
Boghossian's incompatibility argument can be reformulated as this:

(8) Minimalism about truth grounds disciplined syntacticism about truth-aptness. (grounding)

(9) A sentence is truth-apt iff and because it is syntactically sophisticated and disciplined. (disciplined syntacticism)

(10) Moral sentences are syntactically sophisticated and disciplined. (alleged fact)

(3) Hence, moral sentences are truth-apt. (contrary to expressivism)

(11) Therefore, minimalism about truth and expressivism are incompatible.

This argument rests on Boghossian's claim that minimalism about truth implies some particular theory of truth-aptness, namely disciplined syntacticism, which Boghossian thinks of as a minimalist theory of truth-aptness.

Before I turn to my own criticism of Boghossian's argument I want to discuss an objection against it put forward by Richard Holton in (Holton, 2000). Holton thinks that Boghossian's incompatibility argument fails, because it rests on the grounding claim (8). He objects that even if some account of truth grounds some account of truth-aptness, from the mere fact that truth is minimal it does not follow that truth-aptness is also minimal and cannot be a robust or substantive property. Holton devices an analogy between being true and being truth-apt on the one hand, and being a winning lottery ticket and being a winning or a losing lottery ticket on the other.[27]

Holton's analogy shows, I think correctly, that the claim that truth is not a substantive property "doesn't *entail* that no substantial characterization can be made of being either true or false" (Holton, 2000, 9).[28] I agree with Holton that minimalism about truth does not imply a theory of truth-aptness that also must be classified as being *minimal*. However, Holton's argument does not show that minimalism about truth does not ground or imply *any* particular conception of truth-aptness at all. In particular, Holton's arguments do not show that minimalism about truth does not imply disciplined syntacticism. Holton's argument at most shows that the property of being syntactially disciplined need not be thought of as being a *minimal* property. This, however, does not block the central bit of Boghossian's argument for the claim that moral sentences are truth-apt. What would block Boghossian's argument, would be the claim that minimalism about

---

[27]He argues that even if being a winning lottery ticket (being true) would not be a substantive property, this does not show that the features something must have so that it can be either a winning or a losing lottery ticket (be truth-apt) are not substantive. As he observes, something cannot be a winning or a losing lottery ticket (be truth-apt) unless it is a lottery ticket, and whether or not something is a lottery ticket or something completely different (a person, say) is due to substantive properties.

[28]The same conclusion has also been reached by (Jackson et al., 1994) a couple of years earlier. They say: "No argument that what divides the true from the false among the truth apt is not the possession of a substantive property can shake the fact that what divides the truth apt from the non-truth apt is a substantive property" (Jackson et al., 1994, 301).

truth does not imply disciplined syntacticism. But Holton's analogy does not establish that. Hence, Holton's objection as an attack on Boghossian's incompabitibility argument fails.

It remains an interesting question whether minimalism about truth grounds disciplined syntacticism, but I shall not pursue it here. The reason for this is that many philosophers have willingly adopted Boghossian's disciplined syntacticism as a good or even the correct theory of truth-aptness independently of minimalism about truth. Therefore, expressivists do not actually need Boghossian's grounding claim to make a truth-aptness argument. They only need claims (9) and (10) to derive (3). It is this simplified argument and disciplined syntacticism to which I turn now.

The Boghossian-Wright Argument for Truth-Aptness

Many philosophers think of disciplined syntacticism as the correct theory of truth-aptness independently of what they think the correct theory of truth is. They can make the following argument, which in this form has not only been made by Boghossian but also by Crispin Wright in his (Wright, 1992). A similar argument has also been made by Paul Horwich in (Horwich, 1993). The argument can be formulated as follows:

The Boghossian-Wright Argument for Truth-Aptness:

(9) A formula F is truth-apt iff and because F is syntactically sophisticated and disciplined.

(10) Moral F's are syntactically sophisticated and disciplined formulas.

(3) Hence, moral F's are truth-apt.

This argument entails what quasi-realist expressivists want to establish. Is it a good argument for the truth-aptness of moral sentences?

One might try to object to the argument by arguing against (10), that is against the claim that moral sentences are syntactically sophisticated and disciplined (short: syntactically disciplined). One way to do so would be by using a deep reading of 'admits of coherent embedding'[29], and claim that even though moral sentences superficially appear to be embeddable, they are not really embeddable, which, it might be claimed, is precisely what is shown by the Frege-Geach problem. It should be clear from the discussion in chapter 3, however, that expressivists will not accept this objection because they do, of course, believe that the Frege-Geach problem can be solved and so moral sentences *really* admit of coherent embedding. So arguing against expressivism by questioning (10), that is by questioning the embeddability of moral vocabulary, is not a promising objection against the Boghossian-Wright argument.

---

[29]Cf. Boghossian's quote in the previous section which is also requoted below.

Another strategy to block the argument is by arguing against (9), that is against disciplined syntacticism as being the correct theory of truth-aptness. Since I think this objection is successful, and thus shows the argument to be a bad argument for the truth-aptness of moral sentences, let us have a closer look at what this view actually says and what problems it faces.

4.3.5 Disciplined Syntacticism

According to disciplined syntacticism, recall, for a formula to be truth-apt,

> first, [it] must be significant, and second, it must be declarative in form. Unpacking somewhat, the requirements are that the [formula] possess a role within the language: its use must be appropriately disciplined by norms of correct utterance; and that it possess an appropriate syntax: it must admit of coherent embedding within negation, the conditional, and other connectives, and within contexts of propositional attitude. (P. Boghossian, 1990, 163)

With minor variations in the details, other authors followed Boghossian's proposal. For instance, James Lenman defines a formula to be truth-apt iff it satisfies the following two conditions:

> The *Syntax Condition*: That it have the syntactic surface features of a paradigmatic truth-apt sentence—that, in particular, it be apt for significant embedding in negations, conditionals, propositional attitudes and other subsentential constructions.[30]

> The *Discipline Condition*: That its use be subject to a sufficient degree of normative discipline establishing clear standards of appropriate and inappropriate use. (Lenman, 2003a, 34)

So, disciplined syntacticism has two separate parts. Two things must be noted about these parts. First, Boghossian explicates what seems to be a semantic part (that a formula be "significant" or "meaningful") in terms of norms governing "correct utterance" or, more generally, in terms of norms for *use*. This is due to the fact that Boghossian's discussion of truth-aptness is situated in the context of use theories of meaning, what he calls 'non-factualism' (P. Boghossian, 1990, 159). The second thing to note is that Boghossian's other condition, the syntactic condition, does not appear to be purely syntactic, because it is not clear what exactly he means

---

[30]Note that Lenman's Syntax Condition is more general than Boghossian's in that it does not require that sentences are declarative in form, but only that they have the same form as paradigmatic truth-apt formulas of that language.

by saying that a formula must admit of embedding that is "coherent" or "significant".[31]

Boghossian seems to think that the syntax condition and the discipline condition are equally important. Others, such as James Lenman, but also Michael Ridge, think that the two conditions play different roles. For instance, according to Ridge, declarative or assertoric form is what is definitive of truth-aptness, whereas discipline provides the "deeper explanation" of what is "special" about declarative form so that it defines truth-aptness (Ridge, 2014, 198). On Ridge's view, the relevant sort of discipline is that declarative formulas are subject to norms of assertion, namely that "one must have a corresponding ... belief."[32]

Another point I want to make, to which I have already appealed previously, is that two different readings of the syntax condition are implicitly embraced in the literature: let us call them the 'superficial' one, and the 'deep' one. The superficial one holds that whether or not a formula 'admits of coherent embedding' is purely a matter of grammar: it is something which we can observe simply by looking at the grammatical structure of some formula and the grammatical rules of some language. For instance, the predicate 'is wrong' in English when applied to a singular term results in a grammatical sentence which again results in a grammatical English sentence when embedded under 'not' etc. On the other hand, the deep reading holds that whether or not a formula admits of coherent embedding is a much more complicated question which is not purely syntactical. Rather the question is whether a formula is embeddable in the sense of having "embedding-aptitude" (Tiefensee, 2010, 35f.). In the debate about the Frege-Geach or embedding problem no one questioned that a conditional with a moral antecedent is a syntactically well-formed formula of English. Moreover, everyone agrees that such a formula is meaningful. In this context the question is how it can be possible that moral sentences admit of such coherent embedding, *if* expressivism is true. The question is not whether moral sentences intuitively admit of coherent embedding, which is taken for granted. On the deep reading of the syntax condition an answer to the question of whether or not a formula admits of

---

[31]It might be objected that coherent or significant embeddability is a purely syntactic property if we think by understand 'coherent' in the sense that all formulas, non-moral and moral, are subject to the *same* syntactic rules, so that it cannot be that negating a non-moral sentence is grammatical, whereas negating a moral sentences is not.

[32]With his focus on belief Ridge's view resembles my own view which I introduce below. But there is an important difference. Ridge defends a hybrid theory, namely his 'ecumenical expressivism', according to which moral beliefs are hybrid states constituted by robustly cognitive belief states, and non-cognitive states. Since his account of moral thought contains that cognitivist element, his account of truth-aptness can facilitate this belief component for his claim that sentences expressing moral thoughts are truth-apt. Since I am interested in this dissertation in non-ecumenical, or as they are sometimes called 'pure' forms of noncognitivist-expressivism, I cannot rely on a cognitivist element for an account of truth-aptness.

coherent embedding depends on whether or not it gives rise to a Frege-Geach problem or not. In the literature, however, the general tendency has been to think of the property of being syntactically disciplined as a superficial property that can easily be detected just by looking a some formula. This superficiality also explains why disciplined syntacticism has been classified as a minimalist theory of truth-aptness.

Disciplined syntacticism can be criticized in a number of ways. Let's start with the perhaps least damaging points. The syntax condition as well as the discipline condition are underspecified. It is not clear when a formula is 'appropriately' disciplined or is subject to a 'sufficient degree' of discipline. It is also not clear what *kind* of discipline Boghossian has in mind, that is what the relevant norms for 'correct utterance' or 'appropriate use' are. He must have something specific in mind, since not every kind of norm seems relevant. For instance, there clearly are different norms for the correct utterance of declarative English formulas like 'Bob will shut the door'. One such norm might be that one must utter this sentence only if one believes that Bob will shut the door. But this formula also seems to be subject to the norm that, at least in many typical situations, one must use this formula only if one intends to bring it about that Bob shuts the door. However, being subject to this latter norm, seems irrelevant for the truth-aptness of 'Bob will shut the door'. So, the disciplined syntacticist needs to work out his theory in much greater detail, in particular she needs to tell us what the relevant norms of use are, before we can try to assess its plausibility.

Some authors think that we do not have to embrace the syntax as well as the discipline condition, but that truth-aptness is *only* a matter of syntax. Stoljar, for instance, thinks that a formula is truth-apt, if embedding it under 'true' results in a formula as well. Ridge describes this as the view that "truth-aptness comes 'for free' with the declarative form" (Ridge: 198). But pure syntacticism is false. Even if 'gavagai' and 'etihw' were part of the vocabulary of English and functioned syntactially like predicates, so that 'Gaby is gavagai' or 'Snow is etihw', were well-formed formulas, that would not make them truth-apt. Merely having some specific syntactic form by itself is insufficient for truth-aptness.

Moreover, syntactic form is not only insufficient for being truth-apt, it also seems to be unnecessary. Why, for instance, should we think that truth-apt formulas must admit of syntactic embedding? We can easily imagine a language that contains only atomic formulas, such as 'snow is white', 'grass is green' etc., yet think these formulas can be true or false.

There is a further problem for pure syntacticism. There is no syntactical form of the declarative as such. For instance, p, $\exists x F(x)$, $H!p$ are all declarative formulas of propositional logic, first-order logic, and Blackburn's emotivist language respectively, but they do not share a common syntactic form. It might be objected that I confuse syntactical form with logical form. But what is a declarative sentence in the logical sense? In logic textbooks

one usually finds a formula to be defined as being a declarative, indicative, or assertoric sentence iff it is truth-apt. But this is of no help for the (disciplined) syntacticist who tries to do exactly the opposite, namely define truth-aptness in terms of assertoric form.

In general, syntactic form seems irrelevant to truth-aptness. If a formula F_1 in L_1 means the same as another formula F_2 in L_2 (or is subject to the same norms of use) and F_1 is truth-apt in L_1, then F_2 seems truth-apt in L_2, and does so irrespective of its form. In other words: truth-aptness survives modification of form as long as meaning remains unaltered. If 'B!murder' in Blackburn's emotivist language means the same (or has the same use) as 'murder is wrong' means in English (namely that murder is wrong), then 'B!murder' seems to be truth-apt even though it is of a very different form than 'murder is wrong'. For this reason, meaning or use, appears to be a much better guide to truth-aptness than form.

I do not take these points to be full-blown objections. I have just made a couple of critical remarks on disciplined syntacticism as a theory of truth-aptness. But there is also a well-known objection against disciplined syntacticism, namely Jamie Dreier's case of accostivism (Dreier, 1996), or Walter Sinnott-Armstrong's modification of it: greetivism (Sinnott-Armstrong, 2006, 23). I will use greetivism for the purpose of illustration. According to greetivism, uttering 'Hello, Bob' is used to greet Bob. We can introduce the predicate 'is hello' to English and stipulate that 'Bob is hello' is a well-formed formula which is also used to greet Bob. 'Bob is hello' is disciplined, since there is a clear standard for correct use: use it only to greet, where to greet is "to express a kind of respect or pleasure at contacting someone" (Sinnott-Armstrong, 2006, 25). Moreover, Dreier and Sinnott-Armstrong argue, 'Bob is hello' admits of coherent embedding, simply due to the fact that it is a declarative sentence. Despite all this, and this is the point of the objection, it seems we really do not think that 'Bob is hello' could be truth-apt. Hence, Dreier's argument shows that being syntactically disciplined is insufficient for being truth-apt.

Dreier's objection makes old-school expressivists happy who hold that moral sentences are not truth-apt. If being syntactically disciplined is insufficient for being truth-apt, one cannot derive the truth-aptness of moral sentences from the fact that they are syntactically disciplined. But what about quasi-realist expressivists? They are unhappy about Dreier's objection because if disciplined sytacticism is untenable, then the Boghossian-Wright argument is unsound, and so can no longer be used to justify the quasi-realist claim that moral sentences are truth-apt.

Dreier's challenge has consequences not only for the Boghossian-Wright argument for the truth-aptness of moral sentences, but also for all other arguments that rely on disciplined syntacticism. One such argument is made by Neil Sinclair, and this argument is particularly interesting for expressivists since it aims to establish the quasi-realist expressivists claim that moral sentences express beliefs. Even though Sinclair's argument for moral belief

fails, because it rests on the Boghossian-Wright argument, and thus rests on disciplined syntacticism, it fails for other reasons as well. It will be instructive for later, to see what these other reasons are, and so we will now have a look at Sinclair's argument.

4.3.6   Sinclair's Argument for Moral Belief

In the previous section I discussed and rejected disciplined syntacticism as a plausible theory of truth-aptness. In this section, however, I let my discussion of truth-aptness rest for a moment and discuss another argument that is also based on disciplined syntacticism, namely Sinclair's argument for moral belief (Sinclair, 2006). Even though this argument rests on disciplined syntacticism, and for the reasons just disussed is unsound, discussing this argument is nevertheless relevant to my discussion because Sinclair's argument can be seen as an attempt to establish the quasi-realist expressivist's claim that moral sentences express beliefs. One of the overarching aims of this dissertation is to establish exactly this claim, it will be interesting to see what this argument is, and whether it could be rescued if we replace the mistaken disciplined syntacticism with a more plausible theory of truth-aptness. We shall see, however, that combined with what I think is the correct theory of truth-aptness, Sinclair's argument for belief becomes circular, and thus useless for quasi-realist expressivists. A further reason why Sinclair's argument is relevant is because it raises what Sinclair calls the 'moral belief problem'. Since any argument that entails that moral sentences express beliefs gives rise to this problem, it will be interesting to see what this problem is and what Sinclair's proposed solution to it is, even if Sinclair's particular argument for this claim fails. Another reason why Sinclair's argument is relevant to my discussion is that Sinclair suggests it is based on a conceptually necessary connection between truth-aptness and belief. That connection also plays the central role in Smith non-truth-aptness argument with which I deal in the next section. Moreover, that connection also appears in my own truth-aptness argument. So this is a good point for an introduction of the connection.

As I have mentioned in the introduction to this chapter, Sinclair's argument for moral belief takes the Boghossian-Wright argument for the truth-aptness of moral sentences, (9) and (10) to (3), one step further by claiming that there is a conceptual connection between truth-apt sentences and the expressing of beliefs. Unfortunately, Sinclair is not telling us what exactly the alledged connection is. He only says that:

> There is a conceptually necessary connection between a sentence
> being truth-apt and sincere assertoric use of that sentence serving
> to express a belief (whose content is captured by such usage).
> (Sinclair, 2006, 250)

This does not tell us what the connection is, but Sinclair's argumentation suggests that is something like the following conditional principle:

> *Sinclair's principle*: If a sentence is truth-apt, then sincere assertoric use serves to express a belief.

Sinclair suggests that the connection formulated in this principle is the same as one that is advocated by Crispin Wright. Wright, however, said something different, namely that

> assertion has the following analytical tie to belief: if someone makes an assertion, and is supposed to be sincere, it follows that she has a belief whose content can be captured by means of the sentence used. (Wright, 1992, 14)

Call this latter principle *Wright's tie*. I take Sinclair's principle and Wright's tie to be both plausible, but they are different in important respects. So before we take a closer look at Sinclair's argument, I want to point out some important differences between Sinclair's and Wright's formulations.

First, Sinclair mentions a couple of things that are absent in Wright's tie, namely sentences, truth-aptness, and expressing. These additions make differences. For instance, since Sinclair appeals to expressing, the 'sincere' is irrelevant in his principle. I argued above that a plausible constraint on any account of the expressing relation is that it must allow that a person can express a state she is not in. E. g. if someone asserts 'grass is green', she expresses the belief that grass is green irrespective of her being sincere or insincere. Any notion of expressing that we use to explain meaning must be insensitive to speakers truthfulness in this respect if we want meaning to be stable. In contrast to Sinclair's formulation, in Wright's tie the appeal to sincerity is crucial, since it is clearly false that if someone makes an assertion it follows that she has a belief.

Another important difference is that in Wright's formulation, sentences appear not to be important. Wright focuses on someone's making an assertion—he does not require that assertions must be made by doing something with sentences even though doing something with sentences is the usual way of making assertions.

Furthermore, the notions of truth or truth-aptness are completely absent in Wright's tie and so it seems that the tie links together totally different things than Sinclair's links together. Whereas Wright claims that his principle states an analytical tie between assertion and belief, the connection Sinclair seems to have in mind is much more complex and states that if one (sincerely) asserts a truth-apt sentence, one expresses a belief. So Sinclair's connection links truth-aptness to expression of belief.[33]

---

[33]Actually, it is not so clear what the *focus* of Wright's tie actually is. Is it about

With this alleged conceptually necessary connection at hand, Sinclair extends the Boghossian-Wright argument as follows:

> If moral sentences are syntactically disciplined, and if being syntactically disciplined suffices for being truth-apt, which in turn suffices for sincere assertoric use of such sentences serving to express beliefs, then sincere assertoric uses of moral sentences express beliefs. (Sinclair, 2006, 251)

In a slightly more organized form, Sinclair's argument can be reformulated as follows, where (9),(10), and (3) are the same as in the Boghossian-Wright argument:

Sinclair's Argument for Moral Belief:

(9) A formula F is truth-apt iff and because F is syntactically disciplined.

(10) Moral F's are syntactically disciplined.

(3) So, moral F's are truth-apt.

(12) If F is truth-apt, then F in its sincere assertoric use serves to express a belief.

(5) Hence, moral F's in their sincere assertoric uses serve to express beliefs.

Please note that this argument does not work if we replace (12) with Wright's claim that if one makes an assertion and is sincere, one holds a belief. That claim together with (3) does not entail (5). So, the differences between Wright's tie and Sinclair's principle, are in fact important for Sinclair's argument to go through.

Sinclair conceives of this argument as raising a problem, which he calls the 'moral belief problem'. The problem, according to him, is that this argument "entail[s] the falsity of expressivism", which Sinclair takes to be the view that "[m]oral sentences in their sincere assertoric uses do *not* serve to express beliefs" (Sinclair, 2006, 250, emphasis added). As emphasized several times now, this problem is of course only a problem for old-school expressivists, but not for quasi-realist expressivists who do not deny (5), but want to establish it. The quasi-realist expressivist, in contrast, would welcome Sinclair's argument as an attempt to establish precisely what she wants, namely that moral sentences express beliefs. Of course, quasi-realists

---

assertion, or belief, or sincerity? Does it tell us that assertions are sincere only if one holds a *belief*, instead of, say, a desire. Or does it tell us that it is *assertions* instead of commands, say, that are sincere only if one holds a belief? Or does it tell us that assertions are *sincere*, instead of correct, say, if one holds a belief? I shall suggest below that it makes a constitutive claim about assertion.

will not actually welcome Sinclair's argument because, as we already know, it rests on a mistaken theory of truth-aptness, and therefore is unsound. Therefore, it also does not really constitute a moral belief problem.

But please note that even if Sinclair's argument does not constitute his moral belief problem, because it rest on the mistaken view of disciplined syntacticism, there remains a further, more general moral belief problem for the quasi-realist which exists independently of Sinclair's argument for the claim that moral sentences express beliefs. The problem is that however expressivists will be able to establish that moral sentences express beliefs, this claim brushes over the fact that expressivists need to preserve at least some difference between moral beliefs and non-moral beliefs. I will make this more general moral belief problem for expressivists more precise in section 5.2.4 as the 'tightrope problem'.

Let us however, ignore all this for a second, and have a look at how Sinclair proposes to solve the alleged problem which he thinks his argument gives rise to. Considering this solution is interesting because, if it were viable, it would not only help the old-school expressivist, but it would also help the quasi-realist expressivists with their the just mentioned more general moral belief problem.

Sinclair's solution to his moral belief problem is a pluralist one. He proposes that expressivists should become belief pluralists and distinguish between minimal and robust beliefs. It should be clear how making such a distinction might help with the moral belief problem. If the quasi-realist expressivist embraces minimal as well as robust notions of belief, then she can claim that (12) links truth-aptness not to belief simpliciter, but only to minimal belief. In this case the expressivist could accept Sinclair's conclusion that moral sentences express beliefs because if these are only minimal beliefs we thereby make room for a difference between moral and non-moral belief. Non-moral beliefs, the expressivist can then claim, are beliefs in the minimal as well as in a robust sense, whereas moral beliefs are also beliefs, but only in a minimal sense.[34] This would prevent quasi-realist expressivists from becoming cognitivists, and thus solve the moral belief problem.

For this solution to be plausible, we would need to know what a minimal belief is. Sinclair proposes that expressivists become *minimalists about belief.* Minimalists about belief, according to Sinclair, hold the view that "being expressible by a truth-apt sentence is *all there is* to a mental state's being a belief. Thus what we may call a *minimal belief* is simply any state of mind that is expressed by sincere assertoric use of a sentence that is syntactically disciplined" (Sinclair, 2006, 253, emphasis added). In other words, Sinclair suggests that expressivists accept the following view:

---

[34]According to Sinclair robust beliefs are minimal beliefs plus some property R. So, for him robust beliefs are a subset of minimal beliefs.

Minimalism about Belief: The mental state M expressed/expressible by
    some formula F is a (minimal) belief iff F is truth-apt.

However, expressivists should kindly reject Sinclair's proposal to solve the
moral belief problem by adopting minimalism about belief. Minimalism about
belief is problematic for at least the following reasons. First, minimalism
about belief may falsely classify some states as beliefs that clearly are not
beliefs. This is due to the fact that it is unclear what makes a state *expressible*
by a truth-apt sentence. If, for instance, it turns out that some desire state
is expressible by a truth-apt sentence, then desires will be classified as
beliefs, and this is not something quasi-realist expressivists want. Second:
minimalism about belief trivializes Sinclair's principle (12). Recall that
Sinclair now takes the principle to say that if a sentence is truth-apt, it
expresses a minimal belief. But if to be a minimal belief just is to be a
state expressed by a truth-apt sentence, we get: if a sentence is truth-apt
it expresses a state that is expressed by a truth-apt sentence. This sounds
fairly trivial.[35] A final reason against minimalism about belief is that below
I shall argue that we should analyze truth-aptness in terms of belief, which
is clearly incompatible with the belief-minimalist's opposing view that belief
should be analyzed in terms of truth-aptness.

I agree with Sinclair that expressivists must be able to draw a distinction
between moral beliefs and non-moral beliefs. However, I think that the
distinction should not be made by distinguishing between minimal and robust
beliefs. Expressivists should rather seek to make the distinction between
moral and non-moral belief by pointing to differences in their respective
functional roles or dispositional profiles. I will show how this can be done in
the chapter on moral belief. In any case, we should not multiply minimalisms
beyond necessity.

We now know how Sinclair thinks expressivists should deal with the moral
belief problem to which his argument gives rise. But we still need to know
how good Sinclair's argument actually is for moral belief. How plausible
are the premises of Sinclair's argument? Let's start with (10). Sinclair, like
Boghossian and Wright, thinks that there is "undeniable evidence that moral
sentences are syntactically disciplined" (Sinclair, 2006, 252). I think that
almost everyone will accept this, and I already noted above that expressivists
think that moral sentences are embeddable even in the deep sense. So (10)
seems safe.

What about (12), that is Sinclair's principle? In my view, premise (12)
is also quite plausible—at least I think that some precisification of it is true.
(12) is also defended by (Jackson et al., 1994) and plays the central role in
their argument against moral sentences being truth-apt. We will have a look

---

[35]This threat of trivialization as also been noted by (Tiefensee, 2010, 42). Sinclair himself
thinks the vacuity or triviality is to be expected (Sinclair, 2006, 253). But even if it is to
be expected, this does not mean it is harmless.

at their argument and the role (12) plays in it in the next section, where I will also offer a more precise formulation. I want, however, to emphasize, that even though I think that there is a link between truth-aptness and belief, I reject Sinclair's proposal that (12) links truth-aptness to *minimal* belief. It might feel natural to link truth-aptness to minimal belief, if we think, as Sinclair does, of truth-aptness as also being minimal in some sense.[36] But just as I believe that it is preferable to be a monist about truth-aptness, I believe that being a monist about belief is preferable too, and that we should avoid the notoriously blurry notion of 'minimal'.

Of course, the major problem with Sinclair's argument for moral belief is that it rests on (9), that is disciplined syntacticism, and that this is a mistaken view about truth-aptness. Hence, (9) and (10) do not prove (3), which means that even though (10) and (12) are somewhat plausible, Sinclair's argument for moral belief cannot get off the ground. Crucially, for the expressivist this means that she cannot use Sinclair's argument to justify her quasi-realist claim that moral sentences express beliefs.

It may seem, however, that the falsity of (9) need not completely ruin Sinclair's argument because it could still turn out that the *correct* theory of truth-aptness will allow that moral sentences are truth-apt. Whether this project will be successful depends of course on what the correct theory of truth-aptness is. I will ultimately argue that the argument fails because, given the correct theory of truth-aptness, a Sinclair-style argument will become circular.

To summarize: The Boghossian-Wright argument for the truth-aptness of moral sentences as well as Sinclair's argument that such sentences express moral beliefs were originally seen as posing problems for expressivists, though what the authors had in mind was old-school expressivists. My interest here was in how good these arguments are as arguments for the quasi-realist project of establishing the moral sentences are truth-apt and express belief. We found that both arguments are subject to a number of problems. The major problem is that both arguments rest on disciplined syntacticism which is a mistaken view about truth-aptness. Though for the old-school expressivist this would be good news, for the quasi-realist expressivist this is bad news, since she can use neither argument to justify her claim that moral sentences are truth-apt or her claim that they express moral beliefs. I noted that one might think that the second half of Sinclair's argument could be rescued

---

[36]However, even though this might feel natural, it would be nice to have an argument for why we should think that the beliefs expressed by minimally truth-apt sentences are minimal beliefs. Just as I have mentioned before, we should be careful with the label 'minimal', for it is not clear what it means, and in Sinclair's case 'minimal' apparently does not mean the same in 'minimal truth-aptness' as in 'minimal belief'. Moreover, even if 'minimal' meant the same in both, it is not clear why the belief expressed by a minimally truth-apt sentence has to be minimal as well. This clearly echos Holton's objection against Boghossian's grounding claim from above. In principle nothing speaks against minimally truth-apt sentences expressing robust beliefs, or robustly truth-apt sentences expressing minimal beliefs.

if we find a more plausible theory of truth-aptness which also allows to regard moral sentences as being truth-apt. Expressivists could then still pursue the orthodox strategy and use a Sinclair-style argument to argue from truth-aptness to belief.

Allan Gibbard thinks he has a theory of truth-aptness that allows him just this. I will, however, argue that even Gibbard's more plausible theory of truth-aptness is subject to counterexamples, and so it becomes less clear that expressivists should pursue the orthodox strategy exemplified in the Boghossion-Wright and Sinclair's arguments.

4.3.7   Disagreementism

Many have accepted that cases like Dreier's 'Bob is hiyo' (accostivism) and Sinnott-Armstrong's 'Bob is hello' (greetivism) are counterexamples to disciplined syntacticism and looked for ways to supplement the theory with further conditions so that it becomes immune to these kinds of challenge.

It is widely agreed that one of the best supplementations is provided by Allan Gibbard (Gibbard, 2003).[37] Gibbard's starting point is again minimalism about truth. Minimalism about truth, on his understanding, is the view that 'true' is a device of expressing agreement. According to Gibbard, 'true' is used to agree, and 'false' is used to disagree. With this version of minimalism about truth in the background (let's call it 'agreement minimalism'), he suggests that what is wrong with cases like 'Bob is hello' or 'Bob is hiyo' is that one cannot disagree with the speech acts one counts as performing by making conventional use of them, that is one cannot disagree with an accosting or a greeting, respectively. More precisely, one cannot disagree with the state one has to be in if an act of accosting or greeting is to be sincere. But if 'false' is used to disagree, and one cannot disagree with 'Bob is hello', then 'Bob is hello' cannot be truth-apt. This, according to Gibbard, explains why 'Bob is hello' is not truth-apt despite its being syntactically sophisticated and disciplined. In addition, what is needed is that one can dis/agree with it. The idea of Gibbard "is that some sentences do not express states of mind that it is possible to agree or disagree with — and that such sentences are not truth-apt." (M. Schroeder, 2010a, 158). This formulates at least a necessary condition on truth-aptness. Perhaps when combined with syntactic discipline it is also sufficient. So we get the following view of truth-aptness which we could call 'disagreementism':

> *Disagreementism:* A formula F is truth-apt iff, and because, it is syntactically disciplined and one can disagree with the state expressed by it.[38]

---

[37]See (M. Schroeder, 2010a).

[38]I do not claim that this is the view Gibbard actually holds. It is merely one way of spelling out a theory of truth-aptness based on his remarks. For instance, it is not clear if

Is disagreementism the correct view of truth-aptness? There are a couple of things that speak in its favor. First of all, it is a round package: if 'true' and 'false' are used to agree and disagree with something, then a formula can be true or false only if it can be agreed or disagreed with. This broadly follows the above discussed idea that truth-aptness must be grounded in the function of truth. But we have seen that other packages are round too. For instance, if 'true' and 'false' are used to assert or reject some things, then a formula can be true or false only if it can be asserted or rejected. Another round package, the one I shall defend below, is the following: if truth is what beliefs aim at, then a formula can be true or false only if it expresses a belief. So being a round package is hardly decisive.

Another reason that speaks in favor of disagreementism, at least from the perspective of expressivism, is that the notion if disagreement, or rational incoherence, is also central to their account of logic as we have seen in the previous chapter.

The major advantage of disagreementism is of course that it rules out cases such as 'is hiyo' and 'is hello', and that it also rules out paradigmatically non-truth-apt cases such as 'ouch' or 'damn it'. Hence, disagreementism is a conservative theory of truth-aptness.

Despite all these attractions, I think that disagreementism is not the correct view of truth-aptness because it is nevertheless easy to construct counterexamples. For instance, imagine an artificial language that assigns intentions to some declarative sentences. One can disagree with an intention, but does this make a sentence expressing it truth-apt? The answer seems to be: no. Or suppose that 'Horray, Bob' is used to approve of Bob, just like its declarative reformulation 'Bob is horray' is. Suppose furthermore that approval is a state one can disagree with.[39] It disagrees, for instance, with disapproval of Bob which, suppose, is the state expressed by 'Bob is boo'. Should we say that those sentences are truth-apt? The radical truth-aptness theorist will perhaps bite the bullet and say yes. But for the conservative truth-apt theorist, this does not seem to be an option. 'Bob is horray' is not a truth-apt formula despite its expressing a state it is possible to disagree with, and being syntactically sophisticated and disciplined.

The hesitation to call 'Bob is horray' or 'Bob is boo' truth-apt may be due to several reasons. For instance, it may be due to the fact that 'Bob is boo' runs into a negation problem, and more generally into the Frege-Geach problem. If 'Bob is horray' expresses approval of Bob, then which

___

Gibbard really thinks we need to supplement disciplined syntacticism with disagreement or if expressing a state with which it is possible to disagree is already by itself necessary and sufficient for being truth-apt.

[39]In case you think approval is not a state one can disagree with, simply replace my example with any plausible expressivist theory the noncognitive states of which allow for disagreement, such as for instance, the before discussed theory of Schroeder and the attitude of being for.

state is expressed by 'Bob is not horray' and what explains their logical inconsistency?

Another reason for hesitating to call 'Bob is horray' truth-apt, even though it is syntactically disciplined, and can be disagreed with, seems to be that even though we know what it is used for, that is which state it is used to express, we nevertheless do not really know what it means.[40] My conjecture is that we do not have the slightest clue what it means, though we know what it is used for, because we cannot even make sense of what it would be to *assert* 'Bob is horray' or to believe that Bob is horray. So I conclude that disagreementism is not a viable theory of truth-aptness.

4.3.8   Platitude-Respecting Theory of Truth-Aptness

We find another theory of truth-aptness in the context of an argument made by Frank Jackson, Graham Oppy, and Michael Smith (Jackson et al., 1994) and Michael Smith's (Smith, 1994a) and (Smith, 1994c). Smith et al. defend an argument for what they take to be the old-school expressivists' "distinctive claim that moral judgments are not truth-apt." Their non-truth-aptness argument involves two steps: "The first is to advance a plausible necessary condition for a sentence to be truth-apt... The second is to show that the sentences in question do not satisfy the necessary condition" (Jackson et al., 1994, 298). The relevant necessary condition, according to Smith et al. is that sentences are truth-apt only if they express beliefs. The argument is roughly as follows. Sentences are truth-apt only if they express beliefs. Since expressivists, according to Smith et al., deny that moral sentences express beliefs, it follows that moral sentences are not truth-apt.

Originally, Smith et al. offered this argument as a way for old-school expressivists to defend their non-truth-aptness claim against the Boghossian-Wright argument. It aims to show that moral sentences are not truth-apt, by assuming that moral sentences do not express beliefs. It is clear that contemporary quasi-realist expressivists will not accept this kind of argument. They do not want to use the claim that moral sentences do not express beliefs in order to establish that moral sentences are not truth-apt. They want to establish the truth-aptness of moral sentences as well as their expressing beliefs. Hence, the Smithean argument, which was originally presented on behalf of old-school expressivism, poses a threat to the contemporary quasi-realist expressivist. So the latter will try to block the argument.

In the following paragraphs I will first present Smith's theory of truth-aptness, then present his argument for non-truth-aptness of moral sentences, and then discuss several objections that have been raised against the argument. My aim is *not* to defend Smith's argument against these objections, but to use the objections to receive a sharper understanding of Smith's conception of truth-aptness. This will be helpful for the next chapter where

---

[40]For a similar thought see (Sinnott-Armstrong, 2006, 24).

I will use the insights from our discussion to formulate my own conception of truth-aptness.

Smith thinks the Boghossian-Wright argument, and therefore also Sinclair's argument for the truth-aptness of moral sentences are "simply unsound" (Smith, 1994a, 3). He takes issue with the first premise (premise (9) above) of their truth-aptness arguments, that is with disciplined syntacticism. Like Dreier and Sinnott-Armstrong, Smith denies that being syntactically disciplined is sufficient for being truth-apt. According to him, more is needed for being truth-apt. He thinks that disciplined syntacticism is not the complete story about truth-aptness. Instead, Smith and his co-authors defend what they describe as a "platitude-respecting theory of truth-aptness" (Jackson et al., 1994, 295). According to this theory:

Platitude-respecting theory of truth-aptness: "a sentence is truth apt if and only if it has those features that connect platitudinously with having a truth-condition." (Jackson et al., 1994, 295)

This is of course pretty uninformative. What are the features that connect platitudinously with being truth-apt? Interestingly, Smith et al. suggest that Wright's tie states a "central platitude governing truth-aptness" (Jackson et al., 1994, 294), and so needs to be included in any plausible theory of truth-aptness. They believe it is a "mistake" that "Wright, while granting the analytic connection with belief, does not explicitly include it in his account of discipline" (Jackson et al., 1994, 295).

As I have already noted in my discussion of Sinclair's moral belief problem above, Wright's tie is not a platitude connecting to *truth-aptness*, since it does not mention truth-aptness at all. Smith et al., however, unlike Sinclair, are well aware of this fact. Nevertheless, they are willing to accept that in *addition* to Wright's analytical tie between *assertion* and belief

> there [*also*] is an analytical tie between *truth-aptness* and belief. A sentence counts as truth-apt only if it can ... be used to give the content of a belief, specifically, the belief of someone who asserts the truth-apt sentence. Part of the story about rich patterns of usage [i. e. discipline] required to confer truth-conditions [i. e. truth-aptness] must be a story about using the sentence to express belief. (Jackson et al., 1994, 294, italics original)

So, Smith et al. suggest that any theory of truth-aptness must include this tie between truth-aptness and belief. But even so the theory of truth-aptness Smith et al. offer is incomplete. It is incomplete because they do not attempt to list *all* the platitudes that connect with truth-aptness. They write that various conditions must be satisfied if a formula is to be regarded as truth-apt, but leave it open when the list is finished. In their view, for a

formula to be truth-apt "it must have the appropriate syntax, it must be disciplined, it must be possible to use it to give the content of a belief [i.e. express a belief], and so on and so forth." (Jackson et al., 1994, 295). So their theory does not state a sufficient condition for truth-aptness.

But is that a defect? It must be noted that although the platitude-respecting theory of truth-aptness is incomplete, it is not subject to the aforementioned counterexamples. For instance, it easily handles Dreier's hiyo/hello-challenge. Recall that 'Bob is hiyo' and 'Bob is hello' are not used to assert, but to accost and to greet, respectively. Unlike assertions, accostings and greetings do not require beliefs in order to be sincere. So, given that we explicate expressing as I suggested in the second chapter, 'Bob is hiyo' and 'Bob is hello' do not express beliefs. So, it follows from the platitude-respecting theory of truth-aptness, such sentences are not truth-apt.

Smith's theory of truth-aptness is also not subject to my 'Bob is boo'-challenge. As long as we do not assume that the state of approval expressed by 'Bob is horray' can be regarded as a belief, it will turn out as not truth-apt, which, from the perspective of the conservativist, is as it should be.

It is also noteworthy that this theory of truth-aptness is compatible with the key idea of disagreementism. It is part of our concept of belief that no state is a belief, if one cannot disagree with it. Alternatively: if some state is a belief, then one can disagree with it. Hence, if some sentence expresses a state one cannot disagree with, it also cannot express a belief, which in turn means that a sentence expressing such a state is not truth-apt. So, due to the reference to belief, the platitude-respecting theory of truth-aptness captures the key idea of disagreementism.

Even though the platitude-respecting theory of truth-aptness is incomplete in the sense of stating only necessary conditions for truth-aptness, Smith can use his theory to make his argument for the non-truth-aptness which I described at the beginning. In a more organized form we can now state it as follows:

Smith's Argument Against Truth-Aptness

(12) If F is truth-apt, then F expresses a belief.[41]

(13) Moral sentences do not express beliefs.

(14) Hence, moral sentences are not truth-apt.

We have already seen that Smith, like Sinclair, thinks (12) is a conceptually necessary connection, or in Smith preferred terminology, a 'platitude'. The difference between Sinclair and Smith is that Smith makes (12) to be

---

[41]Smith et al say: "If S is truth-apt then S can give the content of a user's belief", in other words, if a sentence is truth-apt, then users are "using the sentence to express belief" (Jackson et al., 1994, 298).

an integral part of his theory of truth-aptness, whereas Sinclair thinks of it only as a kind of bridge-principle linking truth-aptness to belief.

What about (13)? According to Smith et al., expressivists accept (13) because they hold that "there is no such thing as the belief that torture is wrong—no such thing as the *belief* whose content is given by the sentence 'Torture is wrong' " (Jackson et al., 1994, 298). Slightly more detailed Smith explains this as follows:

> what the expressivists are saying is just that, even though the sentences we utter when we make moral judgements have the appropriate syntax, they do *not express belief*. Why? Because, by the expressivist's lights, moral judgements have a connection with the will that they could not have if they were expressions of belief. Thus, according to these expressivists, moral judgements express our desires and not our beliefs; moral sentences are therefore not truth-apt. (Smith, 1994c, 21, italics added)

As already noted, Smith et al. think that their argument does expressivists a favor because they have old-school expressivism in mind. From the perspective of contemporary expressivism, the argument poses a threat. In thinking about how contemporary expressivists could block Smith's argument, they might look at some objections that have been put forward against it.

Divers and Miller's Objection to Smith

Smith's argument has been criticized by several authors. For instance, John Divers and Alexander Miller think that Smith's argument is *invalid*, because it equivocates on 'belief' (Divers & Miller, 1994). Once again the thought is a pluralist one. Like Sinclair, Divers and Miller distinguish between minimal and robust notions of truth-aptness as well as of belief. Importantly, they think that the conceptual connection formulated in (12) holds only if its antecedent as well as its consequent appeal to either only minimal or only robust notions.[42] Since they assume that Smith thinks of his platitude-respecting theory of truth-aptness as a minimalist theory, they think that the right hand side of premise (12) must appeal to minimal beliefs as well.[43]

---

[42]Divers and Miller hold: "It follows from the links between truth-aptitude and belief that if the (declarative) sentences of moral discourse are robustly truth-apt then their sincere uses are expressions of correspondingly robust beliefs" (Divers & Miller, 1994, 16). Since this has been a recurring theme in this chapter, it should by now be clear that this is false: it is logically possible that such principles link minimal notions to robust notion and *vice verca*.

[43]Even though Smith is willing to describe the platitude-respecting theory of truth-aptness as being 'minimalist', he thinks it is not as 'minimally minimalist' as disciplined syntacticism. Cf. (Jackson et al., 1994, 295f.).

According to them "Smith's expressivist shows that since moral commitments (essentially) have motivational implications, and robust beliefs don't, moral sentences are not fit for the expression of robust beliefs." (Divers & Miller, 1994, 16). The problem, according to Divers and Miller, is that Smith's expressivists do not show that moral sentences do not express minimal beliefs.[44] Hence, they claim that Smith actually advances the following fallacious argument:

Smith's Argument (fallacious version):

(15) If a formula F is minimally truth-apt, then it expresses a minimal belief.

(16) Moral sentences do not express robust beliefs.

(17) Hence, moral sentences are not even minimally truth-apt.[45]

Smith gives a comprehensive reply to Divers and Miller charge of fallaciousness. Essentially, he makes two points. First, Smith, like Boghosssian and myself, rejects Divers and Miller's pluralist strategy, in particular their distinction between minimal and robust beliefs as well as between minimal and robust truth-aptness. This, of course, immediately blocks their objection. Second, and more importantly, he shows that their objection rests on a misunderstanding. Divers and Miller think that Smith takes his platitude-respecting theory of truth-aptness to presuppose a Humean theory of the mind according to which 'no belief can be a desire', and that it is due to this Humean theory that expressivists deny moral sentences to express beliefs. But this is only partly correct. Smith agrees that his theory of truth-aptness "in fact *presupposes* a psychological theory" of belief (Divers & Miller, 1994, 13). But he does not hold that his theory of truth-aptness presupposes any *specific* theory, such as, for instance, a Humean theory of psychology. Of course, if one accepts that a sentence is truth-apt only if it expresses a belief, one will not be able to say which sentences lack truth-aptness, unless one adopts *some* particular theory of belief, and the choice of a theory of belief will have consequences for which sentences turn out as lacking truth-aptness. He says:

> minimalism is a common ground theory ... because theorists who embrace quite different philosophies of mind — Humeans and anti-Humeans alike — can all accept it. However, in order to use a minimalist theory of truth-aptitude to figure out which sentences *are* truth-apt, these different theorists need to 'plug

---

[44]Recall from above that Sinclair proposed minimalism about belief as a solution to his moral belief problem.

[45]Compare their argument (Divers & Miller, 1994, 17).

in' their own favourite philosophy of mind — a Humean or an anti-Humean theory, for example — and, depending on which theory they plug in, they will end up thinking that quite different sentences are truth-apt. Without such a philosophy of mind, the minimalist does not have a theory that tells us which sentences are truth-apt at all; he is like a cook who has a recipe but no ingredients. (Smith, 24)[46]

So, it is true that if one holds a Humean theory of belief, as most expressivists have historically done, this does have consequences for which sentences qualify as not truth-apt. But Smith's theory of truth-aptness does not force anyone to be a Humean in the philosophy of mind. Hence, Divers and Miller's objection rest on a misunderstanding and is thus not successful.

Holton's Objection

Another objection to Smith's argument comes from Richard Holton (Holton, 2000). Holton criticises Smith's argument not as invalid, but as *unsound.* He objects that Smith suggests that expressivists (and everyone else) accept premise (12) because it is platitudinous or analytical. Holton, however, thinks it is not analytical. More precisely, Holton disentangles premise (12) into two separate claims (a) and (b). Here is Holton's own reformulation of Smith's argument:

Smith's argument (Holton's version):

(a) Truth-apt sentences, when uttered sincerely, make assertions.

(b) Assertions are expressions of beliefs.

(c) Moral sentences, uttered sincerely, do not express beliefs, since they have the wrong direction of fit.

(d) Moral sentences are not truth-apt.[47]

Holton claims that (a) and (b) are not analytical because "both make substantial philosophical claims which can quite coherently be questioned." (Holton, 2000, 159). I want to note three things about this claim. First, Holton does not back it up by any sort of argument. He simply takes it to be intuitively clear that neither (a) nor (b) are analytical. Second, even if Holton were right, and Smith's premise (12) is not analytical or platitudinous, that

---

[46]This passage suggests that Smith actually takes belief-expression cum disciplined syntacticism not only to be necessary, but also to be sufficient for truth-aptness. If that is correct, Smith's theory must be seen as the precursor to my own theory of truth-aptness below.

[47]This reformulation is taken word for word from (Holton, 2000, 158).

need not mean that it is false. So, even if (a) and (b) are not analytical, this does not prove Smith's argument to be unsound. Third, Holton's evaluation of Smith's premises as being non-analytical crucially depends on Holton's specific reformulation of them. The matter looks different, I shall argue, if we pay more attention to the formulation of the premises. That is what I want to do in the next paragraphs.

Sharpening Smith's premises

Holton disentangles Smith premise (12) into two separate claims. This might be helpful, but (a) and (b) are not the best way to disentangle it. Consider (a) first, that is the claim that "truth-apt sentences, when uttered sincerely, make assertions". The most important problems with it are the following.[48] First, it is unclear why Holton speaks of sincere utterance. Only speech acts can be sincere or insincere, but uttering is not a speech act.[49] Moreover, Wright's tie, on which Smith bases his premise (12) is about assertion, and not about utterance. If we replace 'uttered' with 'asserted' as Holton does, (a) becomes uninformative: clearly, if we (sincerely) assert something, we make an assertion. Second, like above in Sinclair's argument, the 'sincerely' is superfluous in (a). Even notorious liars succeed in making assertions by uttering truth-apt sentences. The problem with liars is not that they pretend to make assertions by uttering sentences, while they do not really make assertions. The problem with liars is that they make assertions by uttering sentences while pretending that their assertions are sincere though in fact they are not. This does not prevent liars from performing assertions by uttering truth-apt sentences. So by ignoring the superfluous 'sincerely' and sticking with 'utterance', we receive something like the following reformulation of Holton's (a):

(a') If F is truth-apt in L, and S utters F, then S makes an assertion.

Is (a') an analytical truth? I do not think so. Even if we are told that L contains truth-apt formulas, and we are also told which ones they are, this by itself will not tell us *how* the truth-apt formulas must be used in L so that the users are actually regarded by the linguistic community as making assertions by using them. *Uttering* truth-apt sentence need not be the way to make an assertion in that community.

Yet, even if (a') is not analytically true, that does not mean it is never true in any language. Surely, in English, one way of making an assertion is by uttering a truth-apt sentence. In fact, this is the way it is typically and conventionally done in English: it is a rule (or regularity) of English

---

[48]A less relevant problem, but still problem is that strictly speaking sentences are not things that are asserted, though they may be used in order to perform assertions.

[49]In Austin's terminology, it is a locutionary act, but not an illocutionary act. See his (Austin, 1962).

that if a person utters a truth-apt sentence she is conventionally regarded as making an assertion.[50] Of course, even in English there are exceptions from this rule. Uttering a truth-apt sentence ironically, in sleep, or as an actor on stage, merely mentioning it or uttering it as antecedent of a conditional, are cases in which one utters a truth-apt sentence, but does not, at least typically, make an assertion. So we will have to restrict (a') to utterances in normal circumstances. But then, even though (a') is not an analytical truth, it is at least an empirical truth about the conventional use of truth-apt sentences in English. (a') tells us something about the conventional use of truth-apt sentences because without the 'truth-apt'-part, it would not even be an empirical truth about English: one clearly does not, not even typically, make an assertion if one utters any old sentence in English.[51]

One might hope that (a') is not only true of English, but of all languages. In other words, even if (a') is not analytically true, it might nevertheless be necessarily true. But that cannot be. We can easily imagine a language, call it 'Pointish', in which assertions can only be made by *pointing* at sentences from the list of truth-apt formulas in Pointish. So the following is true as a matter of conventional fact about Pointish: a user of Pointish S makes an assertion just in case S points at some truth-apt formula from the list. The Pointers have choosen this way of making assertions, let us assume, because they find it more convenient than uttering those formulas, or perhaps because the are mute or deaf.

That pointing at a truth-apt formula is the conventional way in Pointish to make an assertion, need not mean that it is impossible to make an assertion by using some other than the conventional way. Suppose an English speaker is told that the formulas on the Pointers' list are the truth-apt ones in Pointish. Assume further that this English person instead of pointing at one of their truth-apt formulas, *utters* a truth-apt formula from the Pointers' list of truth-apt formulas (assume for simplicity that they are written in

---

[50]There are other conventional ways to make assertions in English besides uttering. E. g. if a journalist writes a newspaper article or a philosopher writes paper, then her *writing* down truth-apt sentences is regarded as making assertions. Other uses of truth-apt sentences are not conventional ways of making assertions in English. E. g. merely pointing to truth-apt sentences is not a conventional way of making assertions in English. Of course, this does not mean that no unconventional way of using truth-apt sentences in English may ever succeed in making an assertion. E. g. if I ask you about the weather tomorrow in Munich, and you point to a sentence on the whiteboard reading 'It will be mostly sunny in Munich tomorrow', I may take your pointing to the sentence as an assertion of the fact that it will be mostly sunny in Munich tomorrow. But, depending of the conversational background, I may also take it as an request to finally wipe this sentence off the whiteboard, and stop pondering about the weather.

[51]Note that if (a') formulates an empirical truth about the conventional use of English, then Smith's argument might be taken as at least entailing that no moral sentence of English is truth-apt. This would be something, but old-school expressivists wanted more. They did not think that the truth-*in*aptness of moral sentences is a peculiarity of English. They thought that no language could be such that its moral sentences are truth-apt.

English and mean and sound the same as they do in English). Will the Pointers take the foreign utterer to have made an assertion by uttering a truth-apt formula? No, for the Pointers the foreign speaker's action must be unintelligible. This will be so even if we assume in addition that the Pointers know how the formulas of their pointing language would sound if they were uttered. In this case the only thing the Pointers would recognize is that the foreign speaker has uttered a formula from their important list. But perhaps the Pointers are charitable and really good at radical interpretation. In this case they will eventually guess that what the foreign speaker tries to do by his articulating a formula instead of pointing at it, is the same as what they do when they point at a formula from the list. But then this will at most be due to conversational effects; it will not be due to the speaker's playing by the rules, that is by obeying the linguistic conventions of Pointish. That the speaker wants to make an assertion by uttering some formula is not something the Pointers can infer from the fact that the speaker complies with their (semantic) rule for the correct use of truth-apt formulas. It is a convention of Pointish that one makes an assertion if and only if one points to some formula from the list, and the foreign speaker has not done this. I take this little thought experiment to show that (a') is not necessarily true of all languages, but at most contingently true of some languages.

Is there not a way of using these insights into formulating a more general version of (a)? Here is an idea. Our discussion of uttering and pointing as ways of making assertions by following the conventions of particular languages showed that there need not be a specific way of using truth-apt formulas to make assertion. So, if (a') is to tell us something important and general about truth-aptness, then it should not refer to some specific conventional way of using truth-apt formulas in some language. Instead, it should only make the weaker claim that *there is* some such conventional way of use of some formula to make assertion in the language. So I think the following might in fact be a necessary and perhaps even analytical truth for all languages and formulas:

(a") If a formula F is truth-apt in L, then there is a conventional way of use U in L such that if one makes U of F, then one is typically (i. e. in normal conditions) regarded as making an assertion.

This claim links together truth-aptness and assertion, but not via some particular form of usage (e. g. uttering, pointing, singing, writing, dancing...), but by requiring that there at least is *some* conventional way of using in order to make an assertion. It is important that there is some *conventional* way of use, not only that there is some way, because, given the right sort of conversational context it might be possible to use all kinds of non-truth-apt formulas to perform an act that will be regarded as an assertion. Another feature that speaks in favor of (a") it that captures the modality that is

implicit in 'truth-*aptness*' by existentially quantifying over ways of use to make assertions.

Now that we have a clearer understanding of (a), namely (a"), let us turn to (b), namely Holton's claim that "assertions are expressions of belief". Holton claims that this too is not an analytical truth. I disagree. I think that given the explication of expressing in terms of sincerity conditions developed in chapter 2, (b) just is a restatement of Wright's tie, which I regard as stating a constitutive truth or norm about assertion. As such it comes close to being an analytical truth.[52] Recall that Wright's tie says the following:

> assertion has the following analytical tie to belief: if someone makes an assertion, and is supposed to be sincere, it follows that she has a belief (Wright, 1992, 14)

In my view, the focus in this tie is not on sincerity or belief, and in particular it is not on truth-aptness. Instead, it tells us something important about assertion: it tells us that assertion is the speech act the sincerity condition of which is that one holds a belief and not some other mental state. Wright's analytical tie in my view is best understood as stating a rule that is constitutive of the act of assertion. This becomes perhaps more obvious if we reformulate (b) as follows:

(b') S's (speech) act A counts as an assertion, where this is the case only if S's A-ing is sincere, only if S holds a (corresponding) belief.

This sounds pretty analytical to my ear: Anyone who thinks that his assertions could be sincere without believing what he asserts, does not understand what an assertion is. The 'analyticity' of Wright's tie stems from constitutivity of this rule or norm of assertion. Holton's original (b) might be regarded as non-analytical, because it appeals to the notion of 'expressing' which has so many different meanings in ordinary English as well as in philosophical usage. But if we explicate it in the way I have proposed, that is in terms of sincerity conditions it it simply a restatement of Wright's tie which I conceive of as an analytical or constitutive truth about assertion.[53]

---

[52]Compare what Searle says about constitutive rules: "constitutive rules [are] almost tautological. [They] can appear now as a rule, now as an analytic truth" (Searle, 1969, 34). And about expressing he says that "the sincerity condition tells us what the speaker expresses in the performance of the act" (Searle, 1969, 65), and he too thinks that the sincerity condition of an assertion is that the asserter holds a belief.

[53]It might be asked whether (b') really formulates a norm? I think yes. I do not find it absurd to think of 'sincere' as being a normative term in the sense of issuing requirements. Sincerity might be thought to issue requirements in the 'property sense' in which Broome thinks rationality issues requirements (Broome, 2013, Ch. 7). So we can think of (b') also as saying that sincerity requires that you assert p, only if you believe p which then is a statement embedded under the normative operator 'sincerity'.

It is has been a rather long and winding section, it is time for a quick recap. In this section I introduced Smith's argument against the truth-aptness of moral sentences. I also presented Holton's objection that crucial premises of the reformulated argument of Smith, namely (a) and (b), are false. I took a detailed look at these two premises and suggested that if we pay more attention to their formulation they do not appear to be false. On the contrary, (a") and (b') plausibly state analytical or constitutive truth about truth-aptness and the speech act of assertion respectively.

What does this mean for the overall dialectic? If neither (a) nor (b) (or their reformulations as (a") or (b')) are false, does this mean for the quasi-realist expressivists that they have to accept Smith's argument against the truth-aptness of moral sentences? As I have already indicated above, the answer is: no. Even if expressivists accept (a") and (b'), which I think they independently should, then they, as quasi-realists, still should not and will not accept (c) or any version of the claim that moral sentences do not express beliefs. Thus, Smith's argument nevertheless remains unsound.

But this is not the main point of this section. The main point of this section has been to argue that (a") and (b') are true, in order to bring us closer to a plausible theory of truth-aptness. It will be the task of the rest of this chapter to show how expressivists should go about 'earning the right' to truth-aptness, and the insights from our discussion in this section will be central for that.

4.4   Explaining Truth-Aptness by Use

Let me summarize what I have done so far. On the one hand I rejected syntacticism, disciplined syntacticism as well as disagreementism as views about what makes a formula truth-apt. One the other hand I defended a claim of Smith's platitude-respecting theory, namely that expressing a belief is a necessary condition for a formula's being truth-apt against Holton's non-analyticity objection.

This leads me to the idea that expressing a belief is not only necessary, but might even be sufficient for a formula's being truth-apt. So I think that something like the following[54] is the correct explanation of what makes a formula truth-apt:

---

[54]I need to say 'something like the following' because I think that even though this theory of truth-aptness is more precise than many of the proposals that have previously been made, this theory is still pretty vague and a fully spelled out theory of truth-aptness needs to say much more about the details.

Belief Explains Truth-Aptness (BETA): For all $\mathcal{L}$, and all well-formed formulas F of $\mathcal{L}$: F is truth-apt in $\mathcal{L}$ iff and because there is a conventional way of use U in $\mathcal{L}$ such that if someone S makes U of F, then S typically or in normal conditions performs a (speech) act A that expresses a belief.

This says that truth-apt formulas are truth-apt because they express beliefs. Given my preferred explication of expressing in terms of sincerity conditions, this theory of truth-aptness is equivalent to the following theory in terms of assertion:

Assertion Explains Truth-Aptness (AETA): For all $\mathcal{L}$, and all well-formed formulas F of $\mathcal{L}$: F is truth-apt in $\mathcal{L}$ iff and because there is a conventional way of use U in $\mathcal{L}$ such that if someone S makes U of F, then S typically or in normal conditions performs a (speech) act A that counts as an assertion.

This says that truth-apt formulas are truth-apt because they can be asserted.

Which of these two theories is conceptually more fundamental? I am inclined to think that the explanation of truth-aptness in terms of assertion is more fundamental as well as more intuitive than the one in terms of expressing belief. The belief-expressing theory has the problem that the notion of expressing means very different things to different people, and so makes it difficult to assess its intuitive plausibility. The assertion-account does not suffer from the same problem because the notion of assertion is somewhat clearer than that of expressing. Moreover, I think that the assertion-account is particularly appealing if we recall that saying that a formula is truth-apt is just another way of saying that it is an assertoric sentence. Given this, the theory can be restated as the claim that assertoric sentences are those that can be asserted. Formulated thus the view sounds obviously true.

Why should we think that these theories of truth-aptness are more plausible than the ones I have rejected? In particular, why should we think that candidacy for assertion is also sufficient for truth-aptness? One answer is that it is hard to see what more could be required. Moreover, I noted in the introduction that once we investigate into the relations between assertion, belief, propositions, and sentences we may easily lose track of what grounds what. So it would not be surprising if we might simply have been confused about the order of explanation between truth-aptness, assertion, and belief, and language use. Clearly, we use specific parts of language for specific purposes. When we look at what we actually use truth-apt formulas for, we find that we primarily use them for making assertions, that is for expressing our beliefs. We may easily be led to thinking that those formulas are suited for making assertions, *because* there is something special about those formulas,

and that special thing is that they are truth-apt. However, I think the real genealogy of truth-aptness is in fact the other way round. Truth-apt formulas do not enable us to make assertions because they can be true or false. Instead, it makes sense to assess them in terms of truth and falsity because they are used for making assertions, and thus for expressing beliefs. That we use some specific set of formulas for making assertions is the reason why we allow these formulas to be assessed in terms of truth and falsity. It is not that sentences express beliefs because they can be true or false. Rather, sentences can be true or false because they express beliefs.

Several other points speak in favor of the theories of truth-aptness in terms of assertion or expressing belief. First, they are clearly more intuitive than disciplined syntacticism or disagreementism. Second, they do not make it too easy for some formula to be truth-apt. Third, cases such as 'Ouch', imperatives, questions, 'Bob is hiyo', 'Bob is hello', and 'Bob is hooray' are correctly classified as non-truth-apt because they are not used to make assertions or express beliefs. Finally, the theory that belief or assertion explains truth-aptness is a harmonious package: it captures the idea that 'true' is used to *assert*, and it also relates to the widely accepted idea that truth is the aim of *belief*. Furthermore, this theory is also compatible with the key idea of disagreementism: that only sentences that express states with which it is possible to disagree with are truth-apt. Beliefs satisfy this condition because no state is a belief if one cannot disagree with it.

The new theory has all the properties which I assumed in the beginning of this chapter a respectable theory of truth-aptness needs to have. For instance, the theory is decidedly externalist in the sense that whether or not a formula is truth-apt is a matter of conventions and not a matter of some features internal or essential to the formula itself. It also respects the important fact that whether or not a formula is truth-apt depends on the language of which it is part and the conventions that partly constitute that language. Furthermore, I assume that this theory is best understood as a monist theory of truth-aptness, and that truth-aptness is also not a minimal property. Of course, the theory by itself does not rule out the possibility of being a truth-aptness pluralist or distinguishing between minimal and robust truth-aptness. But I have argued that there are general reasons why we should prefer theories that are monist and non-minimalist. Whether this theory of truth-aptness is conservative is complicated for two reasons. A theory of truth-aptness is conservative only if it respects our intuitive verdicts about the truth-aptness of some given formula. The first problem is that there can be all kinds of languages, formal ones in particular, about which we might not have any intuitive verdicts as regards their truth-aptness. Second, there is the problem that even for familiar natural languages like English it is not always clear if some given formula is truth-apt or not. In fact, it is sometimes highly controversial whether some English sentence is truth-apt or not. To name only some cases: formulas containing non-referring

names, vague or gappy predicates, conditionals, probability statements, belief ascriptions, and of course, moral sentences. From the perspective of the expressivist who adopts the new theory, the question which of these kinds of sentences are truth-apt has to be decided on the level of belief. The questions they will have to answer are: Is the state expressed by a probability statement (a subjective probability or credence state) a belief? Can there be beliefs in conditional contents? Can the noncognitive states that expressivists assign to moral sentences be regarded as beliefs? Regarding this latter question, it is clear that what the answer is depends on what states expressivists assign to moral sentence. In the second part of this dissertation I shall argue that certain kinds of noncognitive states can be regarded as beliefs.

Overall, there seem to be good reasons for thinking that what explains truth-aptness is assertion or expressing beliefs. Are there any objections?

One might think that there is an objection from learnability. If truth-apt sentences are those that can be used to make assertions and express beliefs, how is it possible that we are so reliable judges when it comes to the question of whether some formula can be said to be true or false? We clearly do not learn for each and every sentence that this one is used to make assertions, whereas that one is used to make commands. It is obvious that we must use something like syntactic form as a guide to a formula's function.[55] I agree that as children we learn that the order of words is important in our sentences and that different word orders are regularly associated with different purposes. For instance, usually, in a declarative sentence the subject is in front of the verb, whereas in interrogative sentences it is mostly the other way round. This, however, does not show that the fact that in some sentence the subject is in front of the verb is what *makes* that sentence suitable for making an assertion. Rather, what is explanatorily more basic is the fact that we want to make assertions. We have chosen to use language as a suitable tool for doing so. Since we do not use language only for the sake of making assertions, but also for making questions, expressing feelings and much more, it was advisable that we choose some observable marker for those formulas which we want to use for the sake of making assertions. Placing the subject in front of the verb is one convenient way to at least provide some kind of heuristic. However, even though we use form as guide to function, this must not hide the fact that function is explanatory prior to form. The form of a sentence helps us to learn something about its function. But that does not mean that it has this function because it has this form, as the disciplined syntacticists want to make us believe. If I am correct, it is actually the other way round. Here is another possible objection. What

---

[55]Syntactic form is of course not the only marker for function. Conversational context, knowledge of the speaker, tone of voice all play a role in determining what a sentence is used for.

if we design a language the conventions of which are such that 'close the door' expresses a belief, the belief that the door is closed, say? Should we then really regard 'close the door' as a truth-apt formula? To this I reply: why not? When 'Close the door' were conventionally used to express the belief that the door is closed, then, expressivists should hold, 'Close the door' would also no longer mean what it means in English, rather it will come to mean what 'the door is closed' means in English. So it would be fine to think that 'close the door' is truth-apt in such a language.

I think that Boghossian was basically on the right track when he said that requirements for truth-aptness must include requirements for candidacy for assertion.[56] But it was a mistake that he thought that candidacy for assertion requires syntactic sophistication and discipline. In my view the only 'requirement' for candidacy for assertion is the constitutive norm that links assertion to belief which is encapsulated in Wright's tie.

In the next section we will see that the new theory of truth-aptness forces expressivists to abandon what I have called the orthodox strategy of realizing their quasi-realist project.

4.4.1   The Belief-to-Truth-Aptness Argument

I started my discussion of truth-aptness by noting that there is a dominant strategy among expressivists to argue for two of their most important quasi-realist claims, namely that moral sentences are truth-apt, and that they express beliefs. I called this the 'orthodox strategy'. The orthodox strategy involves a two step procedure: first, present a theory of truth-aptness and show that moral sentences are truth-apt. Second, argue from the truth-aptness of moral sentences to the claim that they express beliefs via some linking principle. I said that this strategy is examplified by the following argument scheme:

(1) A sentence is truth-apt iff and because it is such-and-such. (theory of truth-aptness)

(2) Moral sentences are such-and-such. (alleged fact)

(3) Hence, moral sentences are truth-apt. (from 1, 2, via modus ponens)

(4) If a sentence is truth-apt, then it expresses a belief. (Sinclair's principle/minimalism about belief)

(5) Hence, moral sentences express beliefs. (from 3, 4, modus ponens)

It should now be clear why expressivists can no longer pursue the orthodox strategy, if they accept the theory of truth-aptness which I have presented in

---

[56]Cf. (P. Boghossian, 1990, 164).

the previous section. If formulas are truth-apt because they express beliefs, then the following instantiation of the above argument scheme begs the question:

(18) A sentence is truth-apt iff and because it expresses a belief. (BETA)

(5) Moral sentences express beliefs.

(3) Hence, moral sentences are truth-apt.

(4) If a sentence is truth-apt, then it expresses a belief.

(5) Hence, moral sentences express beliefs.

Hence, expressivists need to invert the strategy: instead of arguing from the truth-aptness of moral sentences to their expressing beliefs, they should argue from their expressing beliefs to their being truth-apt. In other words, they should aim to make what I call the 'belief-to-truth-aptness argument':

Belief-to-Truth-Aptness Argument:

(18) F is truth-apt iff and because F expresses a belief. (BETA)

(5) Moral F's express moral beliefs.

(3) Hence, moral F's are truth-apt.

Obviously, for this argument to establish that moral sentences are truth-apt, is must be established first that the states expressed by moral sentences can be regarded as beliefs, i.e. (5). Only then it can be concluded, via the new theory about truth-aptness, that moral sentences are truth-apt. Since I have done my best to make plausible the BETA-theory, (18), in this chapter, the remaining crucial step is, of course, to show that the states expressed by moral sentences can be regarded as beliefs.

So, since expressing beliefs is explanatorily prior to being truth-apt, expressivists need to develop a theory of belief in general, and moral belief in particular, *before* they can use the truth-aptness argument to prove that moral formulas are truth-apt. This, I suggest, must be the *new* quasi-realist strategy.

How hard will it be to pursue this new strategy? That depends on how hard it will be to make plausible that the states expressed by moral sentences can be regarded as beliefs. Perhaps there are easier and harder ways to establish that the states expressed by moral sentences can be regarded as beliefs. But expressivists must be aware that the higher they set themselves the bars for when some mental state can be regarded as a belief, the more justified they will also be in claiming that moral sentences are truth-apt. In other words, the harder expressivists make it for themselves to 'earn the right' to belief, the more earned will their right to belief be. And, consequently, the more earned will their right to truth-aptness be.

Of course, one very easy way to show that there are moral beliefs is precluded by the truth-aptness argument itself, namely minimalism about belief. Minimalism about belief and the theory that Belief-explains-Truth-Aptness differ only in what they take to be definiens and definiendum, as well as in what they take to be explanans and explandandum. Whereas minimalism about belief defines and explains belief in terms of truth-aptness, the proposed theory of truth-aptness defines and explains truth-aptness in terms of belief. Hence we cannot use minimalism about belief to justify the second premise of the truth-aptness argument, since doing so would make the argument circular.

4.5 Where Are We and How Did We Get There?

This is the end of part I. So it is time to wrap things up. So far the major claims of this dissertation have been the following:

Expressing: A formula F expresses a mental state M iff M is the sincerity condition of the speech act performed by conventional usage of F.

Assertion: It is constitutive of the speech act of assertion that it is sincere only if you hold a belief.[57]

Truth-Aptness: A formula F is truth-apt (in $\mathcal{L}$) iff and because there is a conventional way of use U (in $\mathcal{L}$) such that if you make that use of F, then you count as performing an assertion.

From these claims it follows that a formula is truth-apt iff it expresses a belief.

Another important claim of this thesis has been this:

Being For: Expressivists can solve the Frege-Geach problem if they assume that the mental state expressed by moral sentences such as 'murder is wrong' is the state of being disposed to being for blaming for murder (or something relevantly similar).

I also said that two of the key goals of the quasi-realist project are to show how, if metaethical expressivism is true, it is possible that:

Moral Truth-Aptness: Moral sentences are truth-apt, and

Moral Belief: Moral sentences express beliefs.

---

[57]Formulated alternatively: A speech act counts as an assertion only if it is sincere only if one holds a belief.

The general aim of this chapter has been to present a new way of how expressivists should go about arguing for Moral Truth-Aptness. I discussed and rejected several theories of truth-aptness such as syntacticism, disciplined syntacticism, and disagreementism. Finally, I proposed a theory of truth-aptness in terms of assertion or expression of belief. The new theory avoids all the problems to which the previous theories have been subject. I take this theory to be both plausible as a general theory of truth-aptness, and also I take it to be in principle compatible with metaethical expressivism.

An important consequence of this theory of truth-aptness is that it precludes expressivists from further pursuing the orthodox quasi-realist strategy of arguing first for the truth-aptness of moral sentences and then conclude, via minimalism about belief, that they express moral beliefs. The new strategy must be to first argue for the claim that moral sentences express moral beliefs and only after that conclude, via the theory of truth-aptness, that moral sentences are truth-apt.

This means that expressivists need to argue that the states they assign to the sentences of some language can be regarded as beliefs before they can conclude that moral sentences are truth-apt. Given my specific first pass proposal for solving the Frege-Geach problem above this will involve, inter alia, to argue that the state of being for being for blaming for murder can be regarded as a belief, namely the moral belief that murder is wrong. More generally, expressivists have to make plausible that the noncognitive states they take to be expressed by moral sentences can be regarded as moral beliefs. However, at the end of my proposal to solve the expressivists' Frege-Geach problem I suggested an interpretation of the structure of the state expressed by moral sentences in terms of dispositions. So for the dispositional expressivist this means that he must argue that the dispositional states which attitude semantics assigns to sentences can be regarded as beliefs. In the next chapter, I want to do exactly this: I want to justify the dispositional interpretation of attitude semantics by showing that and how it relates to a popular view in the philosophy of belief, according to which beliefs are sets of dispositions. Developing such an expressivist-friendly theory of moral belief is the task to which I turn in the next chapter.

Part II

Moral Thought: Noncognitivism

Chapter 5

Moral Belief

This thesis is about expressing moral belief. Yet I have not said much about belief. This will change now. In the first part I said a lot about how we express mental states by the use of language, and how this might explain the meaning of moral sentences. Moreover, I argued that what makes a formula truth-apt is the fact that it is conventionally used to express beliefs. This has the consequence that in order to earn the right to the truth-aptness of moral sentences, metaethical quasi-realist expressivists must first earn the right to moral belief. That is what I attempt to do in this chapter. So, this chapter is concerned with the nature of belief in general and the nature of moral belief in particular and how it can be squared with metaethical expressivism, especially its noncognitivism.[1] The key idea is to justify the dispositional interpretation of the mental states assigned above by attitude semantics by showing how it matches with an independently motivated view in the philosophy of mind, namely dispositionalism about belief. I argue that by embracing a dispositionalist version of noncognitivism expressivists not only earn the right to speak of moral beliefs, but can also explain what is the distinctive difference between moral beliefs and non-moral beliefs, and also solve some other problems that confront noncognitivism.

---

[1]I will only deal with full or all-out belief in this thesis and not with degrees of belief or credences, even though plausibly moral belief can come in degrees just like any other belief. Dealing with degress of moral beliefs might raise extremely complicated questions for noncognitivists for two reasons. First, dealing with degress of belief, like it is done for instance in formal epistemology is very complicated even if we assume, as most formal epistemologists do, a cognitivist treatment of belief. Second, since noncognitivists identify each moral belief with some noncognitive state, they always have to explain how a moral credence state comes to have precisely those properties it is assumed to have, and how this can be given that it is at bottom a noncognitive state. For a noncognitivist treatment of degress of belief applied to Schroeder's specific version of global noncognitivism in terms of states of being for consult (Staffel, 2019).

5.1   Noncognitivism and Belief

The goal of this chapter is to earn noncognitivists the right to moral belief. Why must noncognitivists earn this right and why can they not simply assume to have it?

Let us start with a common observation. If someone utters 'murder is wrong' in normal circumstances and we take that person to be sincere, we assume her to be in a specific mental state. The ordinary way to ascribe that mental state would be by saying 'She believes that murder is wrong'. In saying this we seem to ascribe the speaker a moral belief. This everyday practice suggests that we ordinarily, and pre-theoretically, take the following to be true:

Moral Belief: Moral sentences express beliefs.

Since I use 'state expressed' and 'judgment' as synonyms this is tantamount to saying that moral judgments are beliefs. The notion of belief in play here is the folk psychological notion of belief, the one we use in everyday life, and not some philosophically refined version of it. It is this ordinary notion of belief which I also take to be definitive and explanatory of the notion of truth-aptness in my above account of truth-aptness in the previous chapter. It is this ordinary notion of belief, not a stronger or a weaker one, earning the right to which is one of the key goals on the quasi-realist's agenda. Metaethical expressivists need to *earn* the right to this notion, and cannot simply take it for granted, because Moral Belief has been viewed as directly conflicting the noncognitivist's traditional view that moral sentences express desires. The problem is that beliefs and desires are considered to be categorically distinct mental states, and noncognitivists are regarded as being committed to this distinctness.

They are regarded as being committed to the distinctness of belief and desire because they are Humeans about motivation, and their alleged ability to explain motivation better than cognitivists which is taken to be one of the key arguments in favor of noncognitivism.[2] It is not perfectly clear what it means for beliefs and desires to be 'distinct'. But, roughly, the idea is that the mind contains two discrete boxes of privileged mental states,[3] a belief-box and a desire-box, and that no mental state can be in both boxes.

Traditionally, noncognitivism has been described as the view that moral judgments are desires, and that descriptive judgments are beliefs. The opposing view is cognitivism which accordingly is described as the view that moral as well as descriptive judgments are all beliefs. Given what I

---

[2]I shall question the claim that noncognitivism is really better at explaining moral motivation than cognitivism in the final chapter of this thesis.

[3]Arguably, there are other boxes for other mental states such as intentions, hopes, or fear etc. as well. I have more to say on distinctness below.

just said about the distinctness of belief and desire, noncognitivism thus becomes the view that moral judgments are in some subbox within the desire-box, whereas descriptive judgments are in the belief-box. So, spelling out noncognitivism in terms of beliefs and desires allows noncognitivists to say that the distinctive difference between moral and descriptive judgments is due to the categorical difference between beliefs and desires. And this is what causes the problem. If noncognitivism is the view that moral judgments are desires, and no desire can be a belief, then Moral Belief must be false. In other words, it seems that the quasi-realist project cannot even get off the ground.

How do contemporary expressivists react to this? Since they want to stay true to their Humean roots, they will want to preserve some difference between moral judgments and descriptive judgments. But since they do not want to preclude right from the start that moral judgments are beliefs, they no longer formulate their view in terms of beliefs and desires. Instead, nowadays noncognitivists do not speak of moral judgments as being desires, but as resembling desires in crucial respects. They are, as it is often said, desire-*like*. Since some state can be desire-like without being a full-blown desire, this might allow noncognitivists to claim that some desire-like states are beliefs while clinging to the claim that full-blown desires cannot be beliefs. But in what way then do moral judgments resemble desires? Often it is said that moral judgments are mental states that resemble desires in that they also have a world-to-mind direction of fit, whereas descriptive judgments are mental states with a mind-to-world direction of fit, and then they assume that no mental state can have both directions of fit. Another way has been to say that whereas moral judgments resemble desires in that they are motivational or conative states, descriptive judgments are representational states. Again, then, it is claimed that no state can be representational and motivational or conative.

Unfortunately, it is rarely made more precise what it is for a state to be a motivational state, and also the metaphor of direction-of-fit is mostly not explicated further.[4]

In this thesis I take noncognitivism to be the following view:

Noncognitivism: Moral sentences express noncognitive states (and not cognitive states).[5]

This is tantamount to saying that moral judgments are noncognitive states. In the course of this chapter, I will give noncognitivism a dispositionalist twist, but it will still stay true to this understanding of noncognitivism in terms of noncognitive states just as the name suggests. The opposing view is

---

[4]I will say more on 'motivational' states and the direction of fit-metaphor in chapter 8 below.

[5]In the following I will mostly leave the qualification 'not cognitive states' implicit.

cognitivism. According to cognitivism all judgments are cognitive states and moral judgments are no exception. In line with Humeanism I will assume that no cognitive state can be a noncognitive state and vice versa.

The notions of a cognitive state and a noncognitive state I do not regard as folk psychological notions. Rather they are classificatory terms of what one might call a philosophical theory of psychology. I will treat the notion of a noncognitive state as a primitive here, just as I will treat the notion of a cognitive state as a primitive. So I will not give any definitions of those terms.[6] For my purposes it suffices to give some examples. For instance, to desire to own an expensive car, to love your wife, to wish that your children have a happy life, to dislike dancing, to disapprove of impoliteness are all examples of noncognitive states. Similarly, to believe that grass is green, to suppose that Goldbach's conjecture is true, to know that you have hands, to see or recognize that it rains are all cognitive states.

Given this understanding of noncognitivism in terms of noncognitive states, the task of the quasi-realist is to answer how moral sentences can express beliefs, if they express noncognitive states.[7] More generally, metaethical expressivists have to make plausible that we can legitimately call some noncognitive states beliefs.

My main goal in this chapter is to make plausible the idea that being disposed to be in certain noncognitive mental states can be regarded as holding a moral belief. Any plausible noncognitive theory of belief must be able to solve the most notorious problems that are typically thought to confront noncognitivist theories. In other words, these typical problems can be regarded as constraints which any plausible noncognitivist theory must satisfy. I will thus start my discussion of moral belief by briefly discussing what these problems are so that I can later refer back to this discussion and show how my own dispositional version of noncognitivism is able to tackle these problems.

5.2   Problems for Noncognitivism

Noncognitivism faces a number of problems besides the one of earning the right to moral belief. In the following I want to briefly present a selection of them. My aim here is not to do full justice to those problems, but rather to give a rough idea of what they are about, so that we can later see how my own version of noncognitivism, dispositional noncognitivism, might help

---

[6]By this I do not mean to suggest that it might be impossible to given an illuminating definition of what it is for a state to be noncognitive. Ultimately, perhaps a definition can be given in terms of a class of specific functional roles. But this is beyond the scope of this thesis.

[7]One possible answer would be the one given by hybrid theorists, namely that moral sentences express noncognitive states as well as beliefs. This answer, of course, cannot be given by metaethical expressivists, since they assume that moral sentences to not express hybrid states, that is combinations of states, but only single unified states.

with these problems.

There are additional problems for noncognitivism besides the ones that will be discussed in a moment. Those additional problems have to do with the rationality of moral belief (see Derek Baker's objection in Chapter 6), its role in reasoning (Cian Dorr's wishful thinking problem, Chapter 7), and how it can motivate action (see Chapter 8). These problems are rather complex and will therefore receive an extended treatment in individual chapters of this thesis. So they are not dealt with in the present chapter.

5.2.1   The One Word Problem

According to noncognitivism, moral judgments and descriptive judgments are two different kinds of states. In the introduction to this chapter, I noted that these two states can be ascribed to agents by the use of 'believes that'. Given the alleged fundamental difference between moral judgments and descriptive judgments, however, it is surprising that both types of states can be ascribed to agents by the use of one single word or phrase, namely 'believes that...'. This gives rise to the so-called 'one word problem'.[8] The problem is that of explaining, in a way that is compatible with expressivism and noncognitivism, how it is possible that one single word, 'believes', manages to ascribe two different kinds of states without thereby making 'believes' turn out to have two different meanings, and thus be ambiguous.

For a solution to be compatible with expressivism we need an analysis of the meaning of sentences containing 'believes that' in terms of the mental states those sentences *express*. For a solution to be compatible with noncognitivism, the analysis must allow that the states we ascribe by the use of 'believes that' can be of the different kinds noncognitivists think moral and descriptive judgments are.

It does not seem overly complicated to offer such a solution. The following has been suggested by Mark Schroeder:

expressivist 'believes': For any sentence F and speaker S, 'S believes that F' expresses the descriptive judgment or cognitive state whose object is the proposition that S is in the mental state expressed by F.[9]

This proposal is expressivistic in that it tells us which state is expressed by 'believes'-sentences, and it is also liberal enough to allow that the state we ascribe to someone by the use of 'believes'-sentences might not be a cognitive state. Moreover, this analysis treats the mental states expressed by belief-ascriptions to be themselves descriptive judgments. This fits with the intuitive thought that even if we use 'believes that' to ascribe someone a moral judgment, the judgment we express by making a belief-ascription is itself not

---

[8]I take the term 'one word problem' from (M. Schroeder, 2010a, 86).
[9]Cf. (M. Schroeder, 2010a, 88).

a moral judgment, but a descriptive judgment about the psychological state of the person to which we ascribe the moral judgment. E. g. the sentence 'Sarah believes that murder is wrong' ascribes Sarah the state expressed by 'murder is wrong' by expressing a descriptive judgment that Sarah is in the state expressed by 'murder is wrong'.[10]

Even though the one word problem does not appear to pose an insurmountable problem for expressivism and noncognitivism, it is not clear if this solution will ultimately be successful. One worry one might have is that it is unclear how the solution could be generalized to work not only for belief-ascriptions, but for attitude-ascriptions more generally. Schroeder's solution translates 'believes' into '*is in* mental state...'. But this solution does not work for ascriptions of mental states other than beliefs. For instance, consider the sentence 'Ana hopes that murder is wrong'. By following Schroeder's proposal for 'believes'-sentences, 'hopes'-sentences express descriptive judgments as well. However, we cannot say that it expresses the descriptive judgment that Ana *is in* the state expressed by 'murder is wrong', since this would be the same state as the one expressed by a sentence ascribing to Ana the belief that murder is wrong. The reason why this does not work is clear: sentences embeddable under 'believes that', 'hopes that', and 'that' more generally, are truth-apt sentences. However, truth-apt sentences, I argued above, express beliefs, but not hopes. Hence, for no sentence embedded under 'believes that' will it be the case that it expresses a hope.

Hence, Schroeder's proposal to solve the one word problem for 'believes' seems promising at first sight, but it is not obvious how it can be generalized into a schema that works not only for belief-ascriptions, but for attitude-ascriptions in general, e. g. the ascription of hopes, intentions, fears, desires.

5.2.2   The Many Attitudes Problem

A problem that is clearly related to the just mentioned complication for a solution to the one word problem, is the so-called 'many attitudes problem'.[11] This is the problem that noncognitivists not only have to provide a noncognitive analysis of moral belief, but also of all other attitudes that we can hold towards moral matters. We cannot only believe that murder is wrong, but also wonder if it is wrong, hope that it is wrong, suppose it to be wrong, fear that it is wrong etc. For all these other attitudes, noncognitivists have to offer an account that is compatible with their commitment that there is no genuinely moral content or propositions. The difficulty is that

---

[10]The analysis is liberal in an additional way: If, for instance, the expressivist is a global noncognitivist of a Schroederian type, he might hold that the descriptive judgment expressed by belief-ascriptions is itself a noncognitive state. Simplifying somewhat, in Schroeder's case, for instance, the sentence 'Sarah believes that murder is wrong' is thought to express the state of being for proceeding as is Sarah is being for blaming for murder.

[11]Again I borrow this useful name from (M. Schroeder, 2010a, 84).

even if noncognitivists have a proposal for how to analyze moral belief as a noncognitive state, they do not automatically have a blueprint for how to deal with those other attitudes. If, for instance, they tell us that to believe that murder is wrong is to disapprove of the action-type of murdering, then what could it be to hope that murder is wrong? It cannot be the hope that murder, since that makes no sense at all. Likewise for all the other attitudes just mentioned.

There are not many attempts in the literature to solve the many attitudes problem. Some suggestions as to how noncognitivists might be able to solve it have been made by Sebastian Köhler in his (Köhler, 2013). His idea is, roughly, that noncognitivists should start with their favourite conception of moral belief, and then try to define all other types of mental states through the relations those states have to moral beliefs. This idea has been picked up and further developed by Bob Beddor recently in (Beddor, 2020). I shall not evaluate this proposal here, but the natural reservation one might have with it is this. Even if one accepts that the many different types attitudes stand in various relations to each other, the idea that all propositional attitudes can be defined by reference to a single attitude, namely belief, seems overly optimistic if not simplicistic. Below we get to learn a radically liberal theory about the nature of attitudes by Eric Schwitzgebel, according to which mental states are characterized by a possibly infinitely large number of dispositional properties.[12] We may expect the truth to lie somewhere between these two extremes of simplicity and complexity.

How might a Schroederian noncognitivist solve this problem? But here is a natural suggestion. Let us suppose that what makes the state of being for blaming for murder a genuinely *moral* judgment is not the state of being for, but rather the mental state or property of *blaming* something.[13] If this it true, then we might say that to hope that murder is wrong is simply the hope that murder is blamed, and the fear that murder is wrong is the fear that murder is blamed. How plausible this account is will depend on a number of things, of course, most importantly on the plausibility of the claim that to blame something is a moral attitude. Moreover, it might be asked why it is that only in the case of belief we need that special attitude of being for? In the case of fear, desire and hope, I just suggested, it seems that we can simply stick to the original attitude. One answer might be that from the perspective of the noncognitivist what is problematic about moral belief is precisely that beliefs are assumed to be a cognitive or representational states. That is what makes the case of moral belief problematic; much more problematic than moral hope, fear, or wondering, since those states are not assumed to function to represent facts in the world, which is what noncognitivists are really worried about.

---

[12](Schwitzgebel, 2002), (Schwitzgebel, 2013).

[13]Recall, that 'blaming' is just Schroeder's toy example. We may replace it with whatever we think fits better as a candidate for a moral attitude, e. g. disapproval.

5.2.3   The Moral Attitude Problem

Noncognitivist theories about moral judgment face another problem, the problem of saying what exactly the noncognitive state is that makes up moral judgment. What about the noncognitive attitude constituting the moral judgment makes it a *moral* attitude? I call this, following Alexander Miller (Miller, 2003, Ch. 4.1), the 'moral attitude problem' for noncognitivism.

Generally it seems there are two possibilities for an answer to the moral attitude problem. Either the moral attitude is an unanalyzable, sui generis kind of ethical feeling or emotion, such as some specifically moral kind of disapproval. Or the moral attitude (or attitudes) are analyzable in terms of non-moral feelings, perhaps feelings we are familiar with from folk psychology, such as desire, disgust, or disapproval.

There is yet no consensus as to whether this problem is solvable or not, and if it is what is the best solution. Miller, for instance, has argued that moral attitudes can neither be unanalyzable, sui generis attitudes nor analyzable in naturalistic terms.[14] Sebastian Köhler disagrees. He has argued in his (Köhler, 2013) that if we use the 'Canberra approach', noncognitivism might be compatible with moral attitudes being either sui generis or fully analysable in naturalistic terms.

I am generally optimistic that something like Köhler's approach might solve the moral attitude problem. Moreover, I also think that some of the proposed solutions which Miller rejects might actually work. Hence, even though I think that noncognitivists must offer some solution to the moral attitude problem, I do not take it to be unsolvable. I shall discuss the proposals rejected by Miller in slightly more detail now because the noncognitivist theory that I shall present below seems to be able to piggiback on those proposals.

According to Simon Blackburn, believing that murder is wrong involves, inter alia, feeling angry about those who murder.[15] Let us abbreviate this as 'B!(murder)', where 'B!' denotes the state of 'booing' or feeling angry.[16] Blackburn is aware that I may feel angry about those who murder, without my anger being a specifically moral attitude. But he adds:

> But I may also feel strongly disposed to encourage others to share the *same* anger. By then I am clearly treating the matter as one of public concern, something like a moral issue. (Blackburn, 1998, 9)

Blackburn here proposes that an attitude is moral when it also involves that others share our attitude. I call this proposal, following Miller, the 'emotional ascent'-approach to the moral attitude problem.

---

[14]See (Miller, 2003, 44).

[15]Cf. (Blackburn, 1998).

[16]The following discussion owes much to Miller's presentation in (Miller, 2003, 88ff).

Miller objects that the emotional ascent-account leads to an infinite regress. According to him, Blackburn defines the relevant moral attitude as the state of being angry at those who murder while liking or 'hooraying' ('H!') those who share that same attitude. More formally, the state of morally booing murder, (i.e. $B!_M$(murder)), is defined as follows:

Emotional Ascent: $B!_M(x) = B!(x) \& (H!(\text{Everyone has the attitude } B!_M(x))$.[17]

Obviously, this is circular since the moral attitude appears in its own definition.

However, Miller misinterprets Blackburn. Blackburn does not say that if I judge that murder is wrong, say, I also want others to also hold the same judgment. If we take Blackburn's above quote literally, his "same anger" refers to the *ordinary* anger, and not to the moral attitude as a whole. More precisely, Blackburn thinks the following:

Emotion Ascent-2: $B!_M(x) = B!(x) \& (H!(\text{Everyone has the attitude } B!(x))$.

This captures the idea of Ayer's emotivism that, as Miller himself says, "a clear function of moral discourse is to convince others, specifically to motivate others, to act in certain ways" (Miller, 2003, 49). Other emotivists such a Charles Stevenson also clearly held that to morally approve of something is roughly equivalent to thinking that oneself approves of something and wanting others to do so as well.[18] Moreover, Miller himself thinks that it is "a conceptual fact that, when I judge that Jones has judged that x is good (bad), I will expect Jones to be disposed, *ceteris paribus*, to demand that I share his non-cognitive sentiments of approval (disapproval) towards x" (Miller, 2003, 49, italics original). All those points suggest that Miller has misunderstood the prescriptivist element in Ayer's, Stevenson's and Blackburn's noncognitivisms: even though moral judgments function to 'urge our attitudes upon others'—and that is what makes them moral attitudes— they do not necessarily function to urge our *moral* attitudes upon others. So, I think that Miller's objection to the emotional ascent-approach fails.[19]

Another proposal that Miller discusses and rejects is in terms of what he calls 'higher-order sentiment' (Miller, 2003, 91). The noncognitivist might suggest, he says, that to believe that honesty is morally good is the attitude

---

[17]Subscript 'M' stands for 'moral', and '&' is taken to "conjoin commitments" (Miller, 2003, 89).

[18]Miller also notes this and cites Kivy about the 'quasi-imperative' force of moral terms: "These terms evince our approval; but they also urge our attitudes upon others. 'I approve; do so as well', was Stevenson's rough analysis of 'good'. Cf. (Miller, 2003, 49) citing (Kivy, 1992, 311) and (Stevenson, 1937, 25).

[19]Miller also considers another proposal on behalf on noncognitivism, which he calls 'stable sentiment'-approach (Miller, 2003, 89). But he thinks this account also falls prey to the same problem of circularity, which if what I just said is correct, is not a problem for noncognitivism, however.

of morally approving of honesty, which in turn is analysed as the higher-order state of approving that one approves of honesty. More formally:

Higher-order sentiment: $H!_M(x) = H!(H!(x))$, where 'H!' stands for 'hooraying' or some other pro-attitude.[20]

Even though it has been objected, for instance by Michael Smith in (Smith, 1994a, 146), that it is not clear at which level of approval an higher-order approval state becomes a state of moral approval, Miller suggests that there is a "non-arbitrary" way of choosing the level by taking into account the cognitive limitations of human agents. For this reason, he suggests, a natural limit might already be reached at the second-order, that is approval that one approves, since beyond this it becomes unclear what a higher-order state is really about.[21] Blackburn also was aware that the higher-order account seems to face the problem of saying where exactly the iteration of approval or disapproval states turns the state into a moral attitude, but rather than trying to find an answer, he suggests that the question need not have a precise answer. He says:

> I do not think it is profitable to seek a strict 'definition' of the moral attitude here. Practical life comes in many flavours, and there is no one place on the staircase that identifies a precise point, before which we are not in the sphere of the ethical, and after which we are. (Blackburn, 1998, 13f.)

I will later return to this proposal by Blackburn and explain why it fits nicely with the version of noncognitivism I favor.

5.2.4   The Tightrope Problem

There is another problem for noncognitivism which stems from the fact that noncognitivists posit multiple kinds of judgments, namely descriptive judgments and moral judgments. This gives rise to what Schroeder calls the 'multiple kinds problem' (M. Schroeder, 2010a, 96). According to him the problem is that descriptive judgments and moral judgments share a number of properties, such as that both can be disagreed with, they have a very similar phenomenolgy (i. e. they feel the same), they have the same functional role, and they can both come in degrees. Noncognititivists

---

[20]Compare (Miller, 2003, 91).

[21]Note the similarity here to my above proposal to treat the state expressed by murder is wrong to be the state of being for being for blaming for murder. Also compare David Lewis' dispositional theory of valuing, according to which to value something is to be disposed to desire to desire it, which he justifies by merely stating that it is more complicated that merely being disposed to desire something, but not to complicated. See (Lewis, Smith, & Johnston, 1989, 116).

have to explain why these multiple kinds have so much in common.[22] Bob Beddor also thinks that it is a "legitimate challenge" for noncognitivists to explain those commonalities (Beddor, 2020, 2803). Even though I think the multiple kinds problem or commonality challenge is an important problem for noncognitivism, I think it is only part of an even bigger problem which noncognitivists are really concerned with. I want to call this bigger problem the 'tightrope problem'. I shall describe what the problem is now.

Noncognitivism is a theory about the nature of moral judgments, that is the states expressed by moral sentences. The chief problem for noncognitivists is that they need to make plausible the counterintuitive claim that some noncognitive states can be regarded as beliefs which is a crucial part of their quasi-realist project. The problem, however, gets further complicated by the fact that noncognitivists are, as I have previously said, committed to the view that cognitive and noncognitive states are 'distinct existences'. As I explained above the main reason for their being thus committed is that if cognitive and noncognitive states were not assumed to be crucially distinct, then the claim that noncognitivism is much better at explaining moral motivation than cognitivism, would not make much sense. In other words, the so-called 'motivation argument' for noncognitivism could perhaps not be used as a justification for noncognitivism.[23]

Due to their commitment to distinctness as well as to quasi-realism, noncognitivist have to walk a fine line. On the one hand they must allow that some noncognitive, desire-like states can be regarded as beliefs. On the other hand they must retain a fundamental difference between beliefs and desires. They must secure the distinctness of belief and desire while arguing that some desire-like states are beliefs. This is the tightrope problem for noncognitivism. Formulated in terms of cognitive and noncognitive states the tightrope problem is the problem of explaining how moral judgments can be enough like noncognitive states and yet enough like cognitive states to merit the folk psychological label 'belief', while retaining a categorical distinctness between cognitive and noncognitive states. The problem is to find an explanatory balance between the commonalities and differences between descriptive judgments and moral judgments on the one hand, and the commonalities and differences between moral judgments and cognitive as well as noncognitive states on the other hand.

---

[22]Schroeder emphasizes that the multiple kinds problem concerns not only belief, but also all other attitudes, since noncognitivists must not only explain the commonalities between multiple kinds of beliefs, but also multiple kinds of hopes, wonderings or fears. In other words the multiple kinds problems can be combined with the previously discussed many attitudes problem to yield an even bigger problem. Cf. (M. Schroeder, 2010a, 97).

[23]I shall have much more to say on the motivation argument in chapter 8 below. For now we just need to know that the distinctness between beliefs and desires, or cognitive and noncognitive states plays a central role in this argument, and for this reason is assumed to be of central importance for noncognitivists.

In order to see whether and how noncognitivists can try to balance this tightrope, as well as contribute to the solution of the other mentioned problems, it will be helpful to have a look at what the major theoretical positions in the philosophy of belief are, and see if any of those might suit the purposes of noncognitivists. The following overview will be nothing more but a sketch, but it will help us to see where the best prospects lie for a theory of belief are that is compatible with noncognitivism.

5.3   Theories of Belief

5.3.1   What is a Belief?

What is a belief? In contemporary philosophy a belief is taken to be a state or attitude of cognitive agents. It is what is called a *propositional* attitude, in fact, many philosophers regard it as the prime example of a propositional attitude. The usual story is that a propositional attitude is the mental state of standing in some specific relation to a proposition. Different types of propositional attitudes, such as beliefs, hopes, intentions, suppositions are distinguished by being different types of relations to propositions, e. g. the believing-relation, the desiring-relation, the intending-relation etc.

Of course, the nature of propositions is controversial, and with it also the nature of propositional attitudes. Indeed, as we already know, part of the motivation for noncognitivism is that noncognitivists reject a certain conception of the nature of propositions, in particular when it comes to the nature of moral propositions. A fundamental commitment of noncognitivism is the rejection of the idea that moral, or more generally, normative propositions, if it is accepted that there are such things at all, are intrinsically representational entities, where 'representational' is understood in a robust manner, perhaps defined in terms of some robust notion of truth such as the one provided by a correspondence theory of truth.[24] This brings us directly to one of the most widely accepted theories of the nature of belief: *representationalism.*

---

[24]It is of course natural to suppose that the underlying reason for this rejection is the metaphysical worries noncognitivists have concerning the nature of moral properties. I suspect that the Mackiean view that moral properties would have to be metaphysically 'queer' entities if they existed, paired with a view of propositions according to which a proposition can only exist if what it represents could exist (e. g. Russellianism), leads noncognitivists to their rejection of moral propositions, and thus the rejection of representationalism about moral belief. For Mackie's famous 'argument from queerness' consult (Mackie, 1977). For more on Bertrand Russell's view on the nature of propositions see (Russell, 1903).

Representationalism

The prevalent theory of belief in philosophy is the view that beliefs are representations of ways the world might be that are somehow stored in the mind.[25] This view can come in two different forms. There are those philosophers who think that those representations or propositions are themselves the beliefs, and there are those who think that the fact that a representation is stored in the mind is the belief. On both forms it is natural to describe beliefs as entities or states of the mind. On the representationalist view of belief if someone acquires a new belief through some input, e. g. through sensory experience, a new representation is added to one mental 'belief box' and it can be recalled when this is necessary. Of course various views about the nature of mental representation are possible and so representationalism can come in a variety of forms. Often when the representationalist view of belief is introduced we are asked to think of the mind as a kind of computer that can receive, store, and manipulate information. One of the most striking representationalist views, for instance, is Jerry Fodor's language of thought hypothesis according to which beliefs are sentences of a universal language of thought stored in the mind.[26] Many representationalists think that the prime function of belief is to track features of the world and thereby help us to guide our actions by providing the information about our surroundings.

Dispositionalism

Another theory of belief is dispositionalism. Instead of treating beliefs as internal states of the mind, dispositionalists think of beliefs as being dispositions of agents. The list of philosophers who have explicitly or implicitly defended dispositionalism about belief or mental states in general is long. Paradigm advocates are (Braithwaite, 1932), (Ryle, 1949), (Armstrong, 1968), (Audi, 1973), (Stalnaker, 1984) (Marcus, 1990), and more recently (Schwitzgebel, 2002), and (Schwitzgebel, 2013). For recent criticism of dispositionalism see for instance (Fassio, 2013), and (Jake Quilty-Dunn, 2018).

Dispositionalism like representationalism can come in a variety of forms. Traditional dispositionalism is the view that to believe a proposition p is to have certain behavioral dispositions with respect to p.

Different forms of dispositionalism may place emphasis on different kinds of behavioral dispositions, such as, for instance, the disposition to assert p in the right sort of situations (e. g. if one is asked whether p, one wants do communicate one's views about p...), or the disposition to exhibit surprise if one learns that p is false, the disposition to assent to q if one learns that p implies q and so on. Often all those dispositions are subsumed under

---

[25]The following cursory presentation of representationalism, dispositionalism, interprationalism and functionalism is based on (Schwitzgebel, 2019).

[26]See (Fodor, 1975).

one general disposition such as the disposition to *act as if p*.[27] It should here already be noted that this closely resembles Schroeder's toy analysis of descriptive judgments as being the states of being for *proceeding as if* something is the case.

Since I will explore a version of dispositionalism below, I also want to mention the two main objections dispositionalist theories of belief face. The first is that traditional dispositionalists pursue a reductionist agenda: it is their aim to *reduce* belief, and other mental states, completely to outward observable behavior. Besides the natural scruples[28] one might have with thinking that to believe something is nothing more than to be disposed to behave in certain ways, the main problem with this idea is that what an agent with some belief will be disposed to do, not only depends on a single belief, but crucially seems to depend on what other beliefs and desires she holds. So it seems that if we reduce a belief to a disposition to act, the activation or manifestation conditions will make reference to other mental states as well which leads the reductionist project into an infinite regress. The second problem is that in many cases the connection between belief and behavior is not only manifold, but also quite unstable. Liars, for instance, will not be disposed to assert *p*, when they believe it, instead they will rather be disposed to assert not *p*. Likewise, paralysed, depressed or weak-willed people might not be disposed to assert *p*, even though they believe *p*. One way to deal with this problem could be to introduce rather specific conditions or appeal to normality conditions. But this might either make the theory utterly complicated or let it look arbitrary or uninformative.

Due to these problems, most contemporary dispositionalists have adopted more moderate forms of dispositionalism, which one its advocates, Eric Schwitzgebel, calls 'liberal' forms of dispositionalism (Schwitzgebel, 2013). Liberal dispositionalists hold on to the claim that to believe something is to possess one or more dispositions, but they allow that those dispositions make reference to other mental states, such as other beliefs, desires or feelings, and they allow that beliefs may not result in outward behavior. Hence, they allow that to believe that *p*, involves the disposition to silently assert to oneself that *p*, or to *feel*, instead of exhibiting surprise if *p* turns out to be false.

Interpretationism

Another theory of belief is *interpretationism*. Interpretationists like liberal dispositionalists allow that believing something makes reference to other mental states, but their emphasis, like the traditional dispositionalist's, is

---

[27]See (Engel, 2005).

[28]Behaviorist theories of mental phenomena always have a problem with the fact that we are conscious of many mental states of ours and that they have a certain feel or phenomenology to them. Behaviorism seems to completely ignore the phenomenology of our mental life.

more on the interpretable outward behaviour. Famous interpretationist are for instance Daniel Dennett (Dennett, 1978), (Dennett, 1987) and Donald Davidson (Davidson, 1984). According to Dennett, an agent, or any complex system, can be viewed from three different perspectives: the physical-, the design-, and the intentional stance. A system can be said to have beliefs, when its behavior matches a pattern that might be best explained by taking the intentional stance towards it, that is by attributing to it intentional states like beliefs and desires. Davidson's theory is very similar in spirit. Dennett and Davidson allow that belief attributions are indeterminate in the sense that it might sometimes be possible to interpret one and the same system or agent as holding different sets of belief. Morever, a common theme in interpretationist theories is that for a system to have a belief is nothing over and above matching a particular pattern of interpretable behavior. A claim one often finds is that believing something is identical to matching a belief-pattern. Hence, many interpretationists reject that whether or not an agent believes something is an objective matter of fact. This theme will reoccur below in my discussion of Schwitzgebel's dispositionalism.

Functionalism

Another widely accepted theory of belief is functionalism. Functionalism about belief, or mental states more generally, is the view, roughly, that what makes something a mental state of a particular type is the causal or functional role it plays. In particular it is the causal relations it actually, potentially, or typically has to sensory input, behavioral output, and other mental states. Many contemporary functionalists think that what it is to be a belief can be captured by a complex definite description. Belief, they say, is *the* state, whatever it is, that plays such-and-such a causal-functional role. This implicit characterization of what belief is is then transformed into an explicit definition of belief in the same way as Ramsey and Lewis have proposed to treat theoretical terms.[29]

Unfortunately, functionalists about belief rarely make precise what specific causal-functional roles they take to be definitive of belief. A commendable exception is (Loar, 1981), and also Hannes Leitgeb who lists five "constitutive functional properties of belief" (Leitgeb, 2017, 6). The first is this:

Assumption 1: Belief is a propositional attitude of cognitive agents: an agent's belief that p is a mental state that has as its content the proposition that p.[30]

We have already encountered this assumption in my presentation of representationalism above. Leitgeb also emphasizes what he calls the 'epistemic' function of belief:

---

[29]See (Lewis, 1970).

[30]This and the following assumptions are from (Leitgeb, 2017, 2-6).

Assumption 2: Belief is an agent's representation of what the world is like; it aims at the truth.

Leitgeb cashes this out by using Elisabeth Anscombe's famous direction of fit metaphor which in turn he spells out in normative terms via a truth norm for belief, according to which a belief that p is correct (or fitting) iff p is true.[31]

Like many other functionalists, Leitgeb also notes a 'pragmatic' function of belief, that is the role it plays in causing action:

Assumption 3: If combined appropriately with an agent's desires (and subject to a ceteris paribus clause), belief should commit an agent to rational *action*. (Leitgeb, 2017, 5, italics original)

Leitgeb thinks of assumption 2 and 3 as "input and output functions" of belief (Leitgeb, 2017, 5). But he also thinks that there are functions having to do with how beliefs interact with each other that concern their coherence on which he cites Michael Bratman:

Assumption 4: "An agent's beliefs are subject to an ideal of integration. Other things equal one should be able to agglomerate one's various beliefs into a larger, overall view; and this larger view should satisfy demands for consistency and coherence". (Bratman, 1999, 17)

Finally, he emphasises the "social" or "linguistic" side of belief:

Assumption 5: If an agent is capable of linguistic discourse, then what is expressed by the agent's sincere *assertions* should be her beliefs. An agent ought: to assert sincerely that X only if she believes that X.[32]

---

[31] The notion of a direction-of-fit is due to (Anscombe, 1963). A critical discussion of this notion is in (Ortiz-Millán, 2018).

[32] Even though those claims clearly relate to what I have said above about expressing and assertion and Wright's tie, there seem to be crucial differences. For instance, it is not clear if Leitgeb allows *insincere* assertions to express beliefs. Here and on some other occasions he at least implicates that only *sincere* assertions express beliefs. For instance, he says: "Assertions are speech acts; if they are sincere ..., they express beliefs" (Leitgeb, 2017, 6). In footnote 21, however, he clearly embraces what we above called the 'same content-account of expressing', namely that "[t]he descriptive sentence that gets uttered expresses a proposition, that proposition gets asserted, and the act of asserting that proposition expresses the speaker's belief in that proposition." This suggests that Leitgeb thinks that by either sincerely or insincerely asserting a proposition one expresses a belief. However, Leitgeb explicitly says that he will "understand the term 'assertion' to be restricted to sincere and serious assertions from the start" (Leitgeb, 2017, 275). Given this understanding, we again get the impression that Leitgeb thinks that only sincere assertions express beliefs; a view which I argued above need not be implausible, but is at least incompatible with the purposes of expressivism. This, of course, is not an objection to Leitgeb's understanding of expressing, since his interest is not in expressivism.

I want to note that not all functionalists about belief formulate what they take to be constitutive functions of belief in normative terms as Leitgeb does. For a completely non-normative formulation of such functionalist assumptions see Eric Schwitzgebel's formulations in (Schwitzgebel, 2019, 1.4). Since noncognitivists have a rather vexed relationship with normativity, as they have with morality, they are likely to prefer non-normative versions of functionalism.

If we have a look at the above assumptions, it must be noted that it is not totally clear that all assumptions describe what might properly be called functions of belief, and so it is not clear, why we should think that they mark a form of functionalism. For instance, the first and the second assumption do not obviously say something specifically about the causal or functional role of belief. Rather they are assumptions that could also be made by a representationalist.

This is of course not objection to the above list. Rather it highlights what should already have become visible: many of the different kinds of theories of belief presented in this section are or can be made compatible with each other, and in fact most philosophers of belief select and combine claims from representationalism, dispositionalism, interpretationism, and functionalism. Even though it might seem that representationalism and functionalism are two very different takes on belief, many contemporary philosophers seem to hold views that combine claims from both camps, as is apparent from Leitgeb's listing assumption 1 as a constitutive functional property of belief. The question representational functionalists have to answer is how a mental state can represent something by playing a certain functional role? One familiar answer is to embrace a form of what has been called 'functional or conceptual role semantics'.[33] According to this view what *kind* of mental state as well as what *content* it has is completely reducible to facts about its functional role, that is to what is actually, potentially, or normally causes and is caused by.

Another difference worth mentioning concerns Leitgeb's normative understanding of assertion. He treats the normative dimension to come from the 'ought' whereas I think that if there is any normativity involved here at all it comes from the constitutive rule governing assertion which can be formulated by the use of 'sincerely': Sincerity requires that you make an assertion only if you hold a corresponding belief. One might hesitate thinking that sincerity is normative. But we can conceive of sincerity like I think we should conceive of rationality, morality, prudence or even etiquette. Rather than providing us reasons for being rational, moral, prudent or sincere, you simply do not count a rational, moral, prudent or sincere, if you do not comply with the rules of rationality, morality, or sincerity. For instance, you simply do not count as polite (in the case of etiquette) when you put your feet on the dining table, and you do not count as rational if you believe p while also believe not p. Likewise, you do not count as sincere if you make an assertion, but do not hold a corresponding belief. This is a constitutive norm or requirement governing the speech act of assertion. Compare what said earlier about this in chapter 2.

[33]For more on conceptual role semantics see for instance (Harman, 1982).

Dispositionalism is also compatible with functionalism and might even be regarded as a specific form of it. How might dispositionalism be a form of functionalism? Schwitzgebel, for instance, proposes that one way is by distinguishing between backward- and forward-looking dispositions, where the backward-looking are those that emphasize what causes the belief in question, and the forward-looking are those that emphasize what the state can cause. Dispositionalism, he suggests, can be seen as a form of functionalism that even though it accepts that the backward-looking functions are important, only takes the forward-looking functions as being definitive of belief. So, even though dispositional functionalists may take assumptions 1 and 2 to be important, they may take only 3, 4, and 5 as definitive.

Even though the above classification of theories of belief is helpful for getting a rough overview, we see that the boundaries are not sharp. Consequently the names of the above theories are used very differently by different authors.

5.3.2   Noncognitivism and Belief

Which of the aforementioned theories of belief suits the purposes of noncognitivism? I have already mentioned that even if some noncognitivists accept that moral judgments are propositional attitudes, for instance, by embracing a non-standard view about propositions, they clearly reject that moral judgments are representational states in a robust sense. For this reason it is natural to suspect that noncognitivists will not normally be representationalists about belief.[34]

Indeed, most contemporary noncognitivists instead embrace forms of functionalism, interpretationism, dispositionalism, or mixtures thereof, though a completely worked-out noncognitivist theory of belief has yet to be developed.

5.4   Dispositionalism and Noncognitivism

Noncognitivism, I said above, traditionally has been regarded as the claim that moral judgments are not beliefs, but desires. Formulated thus, the view is highly counterintuitive. Some contemporary noncognitivists have proposed to reformulate the thesis of noncognitivism more cautiously not in terms of belief and desires, but in terms of desire-like and belief-like mental states. Many contemporary noncognititivsts "hold that moral beliefs are

---

[34]Of course it remains open to them to be representationalists about moral belief by embracing a form of minimalism about propositions or representation. But representationalism, as it is normally understood, does not distinguish between robust and minimal notions of representation. So I find it more natural for a noncognitivist to embrace one of the other theories of belief, e. g. a functionalist one, and at most add to that the claim that moral beliefs can be minimally representational. Claiming to be a representationalist about belief while being a minimalist about representation is misleading and rather eccentric.

much more similar to desires than they are to prosaic [i. e. non-moral] beliefs" (Beddor, 2019, 1) It is natural to spell out the similarities in broadly functionalist or dispositionalist terms. Beddor, for instance, proposes to understand noncognitivism explicitly as a claim about the functional role of moral belief which can be formulated as follows:

Functionalist Noncognitivism: the functional role of moral beliefs resembles the functional role of desires much more than it resembles the functional role of non-moral beliefs.[35]

But functionalist and dispositionalist theories of belief can be developed in a variety of ways. Which ones are suited for the purposes of noncognitivism? In order to answer this question I shall in the following discuss two broadly functionalist-dispositionalist theories, and finally evaluate which of their elements might help noncognitivists, and which elements they might have to reject.

5.4.1 Schwitzgebel's Dispositional Theory of Belief

Let us start with a dispositional theory that has been proposed without noncognitivism in mind. In a series of papers, Eric Schwitzgebel has proposed and defended a theory of belief in particular, and of mental states more generally, which is dispositionalist, but also includes functionalist and interpretationist elements.[36] In this section I want to introduce Schwitzgebels so-called 'phenomenal, dispositional account of belief' (Schwitzgebel, 2002) and evaluate whether it is an account suitable for the needs of noncognitivists, in particular as regards the above described tightrope problem.

Schwitzgebel's theory is primarily dispositionalist, that is it holds, very roughly, that beliefs are dispositions of agents. This links back to the old pragmatist idea that to believe something is to be disposed to act in certain ways. For a historic treatment of this idea see (Engel, 2005).

Here is Schwitzgebel's central claim about belief, which he also takes to provide a schema for all other propositional attitudes:

Beliefs as Dispositions: To have a belief (or any other mental state M) is, primarily, to have a dispositional profile that matches, to an appropriate degree and in appropriate respects, a dispositional stereotype for that belief (or M), typically grounded in folk psychology.[37]

In order to understand Schwitzgebel's theory of belief, we need to understand the terminology he uses. A dispositional *profile*, according to Schwitzgebel, is a set of dispositional properties, or short: dispositions, that

---

[35]Cf. (Beddor, 2019, 1).

[36]The papers I refer to are (Schwitzgebel, 2002) and(Schwitzgebel, 2013).

[37]Cf. (Schwitzgebel, 2013, 75).

an agent has. A dispositional *stereotype* for some mental state (like the state of believing that there is beer in the fridge) is a cluster of dispositions that would be regarded as characteristic of something that possesses that mental state.

For present purposes having a rough idea of this theory will suffice. The rough idea is this: on the one hand there are agents with their own individual dispositional properties. The complete set of those dispositional properties is that agent's dispositional profile. On the other hand, there are, perhaps culturally determined, stereotypes associated with specific general properties of agents, such as their character traits, like being courageous or hot-tempered. Schwitzgebel thinks, that mental states, like beliefs or desires, are similar to character traits in that they have stereotypes. The underlying thought here is that communities might have found it useful to distinguish certain sets of dispositional properties of agents, perhaps in order to make it easier to predict their behavior. So certain typical combinations of dispositions have been grouped together and received names for better reference, such as 'belief', 'desire', 'hope' etc, just like we invented the labels 'courageous' and 'hot-tempered' to roughly classify people which helps with coordinating action. Those sets of dispositions do not have sharp boundaries though, but are rather fuzzy; a property which Schwitzgebel thinks is beneficial for various reasons, as we will see below. Due to the fuzziness, Schwitzgebel prefers not to speak of sets, but of 'clusters' of dispositional properies that make up a stereotype for each specific mental state. He invites us to think "of the dispositional stereotype for belief that p [...] as consisting of the cluster of dispositions that we are apt to associate with the belief that p" (Schwitzgebel, 2002, 251).[38] In each such cluster, some dispositional properties are "central" or "characteristic", while others are only "marginal" or "peripheral" to the stereotype for some specific mental state (Schwitzgebel, 2002, 252). But the dispositional properties in the stereotyptical cluster of belief cannot only be central or peripheral. They can also be further divided into three categories, namely 'behavioral', 'phenomenal' and 'cognitive' dispositions. Schwitzgebel illustrates this with his favorite example which is the belief that there is beer in the fridge:

> The dispositional properties belonging to belief stereotypes fall into three main categories. The most obvious, perhaps, are behavioral dispositions, the manifestations of which are verbal and nonverbal behavior, such as, in the present case, the disposition to say that there is beer in the fridge in appropriate circumstances, and the disposition to go to the fridge if one wants a beer. Equally important, though rarely invoked in dispositional accounts of any sort, are what may be called phenomenal dispositions, dispositions

---

[38]Unfortunately, Schwitzgebel does not further explicate the modality signaled by 'apt' here.

to have certain sorts of conscious experiences. The disposition
to say silently to oneself, "there's beer in my fridge," and the
disposition to feel surprise should one open the fridge and find no
beer are phenomenal dispositions stereotypical of the belief that
there is beer in the fridge. Finally, there are dispositions to enter
mental states that are not wholly characterizable phenomenally,
such as dispositions to draw conclusions entailed by the belief
in question or to acquire new desires or habits consonant with
the belief. Call these cognitive dispositions. (Schwitzgebel, 2002,
252)

There are several things to note here. First, since Schwitzgebel allows
conscious experience and other mental states, the 'phenomenal' aspect, to
play a central role in his dispositional theory of belief, he escapes the above-
mentioned anti-behaviorist worries against dispositional accounts of belief.
Schwitzgebel is explicit that he does not pursue the reductionist agenda
of the behaviorists. His aim is not to reduce mental states to observable
outward behavior. Second, it must be noted that some of the dispositions
Schwitzgebel mentions here clearly echo some of the core functionalist as-
sumptions mentioned above, such as the disposition to assert (Assumption
5), the disposition to act when combined with a desire ('wants') (Assumption
3), or the disposition to draw conclusions entailed by the belief (Assumption
4), and he also seems to accept the view that belief is a propositional atti-
tude (Assumption 1). It must also be noted that none of the dispositions
Schwitzgebel mentions seems to capture the second functionalist's assump-
tion (Assumption 2), that belief is necessarily a representational state. This
suggests that he does not take our ordinary concept of belief to include
the idea that belief is a representational state. In fact, he conceives of his
dispositional theory as an alternative to representationalism. He explicitly
rejects any view that "treats having an attitude [e. g. belief] as a matter
of possessing some particular internally stored representational content, a
content perhaps poised to play some specific set of cognitive roles depending
on the attitude type" (Schwitzgebel, 2013, 75).[39]

---

[39]Even though Schwitzgebel often sounds as if he blatantly rejects representationalism, it
is not clear that he really does or must reject the idea that belief is a representational state.
Some philosophers have proposed to switch from talk about internally stored representations,
or intrinsically representational entities like propositions, to talk about representing as true
or predicating something of something, where representing and predicating are understood
not as semantic relations between quasi-linguistic entities (e. g. propositions) and the
world, but as mental acts. Cf. (Soames, 2013). On such an understanding nothing seems
to speak against the view that to believe that p involves the disposition to *represent-as-true*
that p, which Schwitzgebel might count to the category of phenomenal dispositions because
it involves the mental state of representing-as-true.

According to Schwitzgebel, for an agent to believe that there is beer in the fridge it is not required that he has all of the above-mentioned dispositions or all of the dispositions we would normally expect someone with that belief to have. In other words, in some cases an agent can be said to believe something even though he has only some, but not all of the dispositional properties that are stereotypical for holding that belief. More generally, holding a belief, or any other mental state, is a matter of how good an agent's dispositions match some dispositional stereotype. Schwitzgebel explains:

> the greater the proportion of stereotypical dispositions a person possesses, and the more central those dispositions are to the stereotype, the more appropriate it is to describe her as possessing the belief in question. (Schwitzgebel, 2002, 252)

This statement, like many others, in Schwitzgebel's writings has a clear interpretationist ring to it. He thinks that if someone is appropriately describable as possessing some belief, then he holds the belief. According to Schwitzgebel, there is nothing more or less to believing than being interpretable as believing, that is as matching some belief stereotype to an appropriately high degree. In his view, whether or not someone holds a belief is not an objective matter of fact over and above, or below, someone's dispositional profile. When in doubt whether someone 'really' believes something, it is best to give a detailed description of his dispositions.

One of the major points in favor of his dispositional theory of belief, thinks Schwitzgebel, is that it easily handles what he calls 'in-between cases'. An in-between case is a case where an agent is neither accurately describable as holding some attitude or not holding it.[40] He thinks that such cases are abundant, and his writings are full of colorful and detailed examples. Representationalist theories of belief, he argues, have a hard time explaining such cases because they always have to treat holding some belief as a matter of Yes-or-No: either some state is in the belief-box or it is not. In contrast, the dispositional stereotype approach has no problem explaining in-between cases because whether or not an agent's dispositions match the stereotype for some mental state is a matter of degree and interpretation. More precisely, the ability to handle in-between cases comes from at least four places in the dispositional theory. First, stereotypes to do not have sharp boundaries. So it is not even clear when there is a perfect match between an agents dispositions and a belief's stereotype. Second, an agents dispositional profile can match a dispositional stereotype without the dispositions in the agent's profile all having to be in the stereotype. Moreover, stereotypes do overlap, that is the same disposition can belong to more than one stereotype. And finally, all those matters depend on interpretation.

---

[40]Schwitzgebel insists that even though some in-between cases of believing can be handled by appealing to degrees of belief, the cases he is interested in cannot be so handled (Schwitzgebel, 2002, 261).

5.4.2 Dispositionalism and Noncognitivism?

Now that we are familiar with the cornerstones of Schwitzgebel's dispositional theory of belief, we can ask whether this theory might generally be suited for the purposes of noncognitivism. In order to suit the purposes of noncognitivism it should at least have some promise to solve the traditional problems for noncognitivism mentioned above. Most importantly, however, it should be compatible with the idea that moral beliefs are noncognitive states of mind.

At first sight, nothing seems to speak against the compatibility of noncognitivism and dispositionalism, in fact they seem to fit quite nicely. One thing that will clearly be music in the ears of the noncognitivist is Schwitzgebel's rejection of the idea that belief is a representational state.[41] Another major attraction of Schwitzgebel's brand of dispositionalism for noncognitivism is its liberalism with respect to when an agent can count as holding a belief. One of the reasons for its liberalism, I have said above, is the fact that stereotypes have fuzzy edges and can overlap. In other words, one and the same disposition can belong to the stereotypes of more than one attitude. This implies that if an agent satisfies one stereotype she might simultaneously nearly match the stereotype for some other attitude. This might help with the noncognitivist's quasi-realist project of justifying the claim that some noncognitive states deserve to be called beliefs. This would be the case if the stereotypes for some noncognitive states would overlap with the stereotypes for moral beliefs.

Interestingly, Schwitzgebel thinks that this is indeed often the case, even though he is not concerned with noncognitivism. He notes that since the dispositional stereotypes of different attitudes can overlap, one and the same agent might accurately be described either as holding the *belief* that something is better than something else, or as being in the state of *valuing* something more than something else, or as *desiring* something more than something else. He treats this as a further strength of his theory and writes:

> On a dispositional stereotype approach to the attitudes, we can treat the stereotypes associated with these somewhat different attitudes as largely overlapping, though with different centers and peripheries. Believing and desiring and valuing would seem on the surface to be very different attitude types, and are often treated as such—beliefs are 'cognitive', desires are 'conative', they have different 'directions of fit' etc.—and yet [in some cases] belief, desire, and valuation seem only subtly different. (Schwitzgebel, 2013, 90)

---

[41] Actually, it is not clear that noncognitivists want to be anti-representationalists about all beliefs, especially descriptive beliefs. This relates to my worry, below, that Schwitzgebel's dispositionalism as he construes it stands in conflict with the noncognitivists' commitment to the Humean distinctness claim.

Quasi-realist noncognitivists might be inclined to cheer at this because, in a sense, this is pretty close what they want to establish, namely that some noncognitive, desire-like states have so much in common with beliefs that they are appropriately interpretable as beliefs, and thus deserve to be called beliefs.

Thus, the liberalism of Schwitzgebel's dispositionalism might be viewed as an attraction for noncognitivists, because if mental states do not have clearly defined boundaries, this makes room for arguing that some noncognitives states can be classified as moral beliefs. If the clusters of dispositions associated with certain noncognitive states resemble the clusters of dispositions associated with beliefs, that is, if they overlap in crucial respects, then this might justify the noncognitivists' claim that moral judgments can be regarded as beliefs even though they are noncognitive states.

Nevertheless, I think that Schwitzgebel's dispositionalism at least in its current form, is unsuited for the purposes of noncognitivism. The reason is that Schwitzgebel's dispositionalism is *too* liberal. Schwitzgebel's theory may allow that in some cases there is no difference between some noncognitive states and what we think of as being moral beliefs. But Schwitzgebel's theory allows more than that, more than noncognitivists want, namely to completely collapse the distinction between states that are cognitive (i.e. belief-like) on the one hand, and states that are noncognitive (i.e. desire-like) on the other. Schwitzgebel is well aware of this implication of his theory:

> On a dispositional stereotype approach, there is no sharp division between these attitude types, though the attitudes seem to cross the cognitive-conative divide. (Schwitzgebel, 2013, 90)

The extreme liberalism of Schwitzgebel's dispositionalism might be regarded as an attraction, but it also raises a problem. The problem is, as I have explained above, that noncognitivists are committed to the claim that some mental states are categorically distinct. This distinctness is crucial, I repeatedly said, for the noncognitivists so-called motivation argument (on which I will have much more to say in chapter 8) which relies on the Humean theory of motivation according to which both of these categorically distinct state are necessary in order to generate motivation.

Schwitzgebel himself does not view the clash of his liberalism with distinctness as a problem because for him the Humean view of "the kinematics of belief P, copulating with desire Q to beget intention R" is all "representationalist imagery" (Schwitzgebel, 2013, 90). Schwitzgebel, wants to completely reject this 'imagery' and with it the underlying idea that there are any categorical differences between different types of mental states.[42]

---

[42]It would of course be interesting to know how Schwitzgebel thinks the 'kinematics' of belief and desire to generate intention or motivation can be spelled out in terms of dispositions. But as far as I know he is completely silent on this topic. I will return to this 'kinematics' or 'mechanics' in the second half of chapter 8.

For him there are no sharp boundaries to be found in the mind—there are no belief- and desire-boxes. In his view, the representationalist philosophical psychology is "an optimistic promise or simplistic cartoon sketch of the mind" (Schwitzgebel, 2013, 94). Rather, he thinks, "the mind is a weird, kludgy chaos of dynamic agonisms and antogonisms" (ibid.). The claim that there are categorical distinctions between some mental states does not sit well with Schwitzgebel's extreme liberalism.

Noncognitivists attracted to Schwitzgebel's dispositionalism might reply to this worry that, literally speaking, Schwitzgebel does reject that there is any division between attitude types. Instead, it might be replied, he merely rejects that the division is clear-cut. That seems correct. At some places, Schwitzgebel is careful enough to say that different types of attitudes might be very similar yet not identical:

> It is in general, I think, an appealing feature of the dispositional stereotype approach that, through the mechanism of overlapping stereotypes, it naturally handles the fact that possession of one attitude seems *nearly but not quite tantamount* to the possession of related attitudes, both within and across general attitude types. (Schwitzgebel, 2013, 91, italics added)

Nevertheless, Schwitzgebel himself is clearly attracted to the idea that there are no boundaries at all in the mind, and that if there are any differences between different types of attitudes such as beliefs and desires, cognitive and noncognitive states, they are totally conventional or practical. Sometimes describing someone as believing something might be more suited to highlight particular dispositions of agents, where in other situations describing the very same person as desiring something might help highlight other dispositions of the agent that are more relevant to the this situation. But as useful as distinguishing between different types of states might be for coordinating our interaction with others, they do not track any real difference in the mind, or so Schwitzgebel thinks.

Still I think dispositionalism should be highly attractive to noncognitivists and even if Schwitzgebel's theory does not perfectly fit the purposes of noncognitivism, it might be modified so as to fit it better. For dispositionalism to suit the noncognitivists purposes the key requirement is that it must enable them to draw a sharp boundary between cognitive and noncognitive states. In order to do so we need to restrict its liberalism. How?

What is particularly problematic about Schwitzgebel's theory is that even though he assumes that for any specific belief there are some central and some peripheral dispositions in its stereotype, he thinks that "[n]o one disposition is either necessary or sufficient for the possession of any belief" (Schwitzgebel, 2002, 252). The natural idea for the noncognitivists would be to reject this claim and sharpen Schwitzgebel's rather vague and metaphorical

talk of central and peripheral dispositions. They might instead assume that each mental state has a specific set of dispositions that are *essential* to or constitutive of its being that mental state in the sense that if an agent does not have that disposition, he does not hold that mental state. More generally, it should be argued that there is some restricted set of types of dispositions that is definitive not only for some specific mental state (like the belief that there is beer in the fridge), but definitive of specific *types* of mental states (like the mental state type of belief). We might additionally allow that there are some dispositions that are typically associated with the mental state type of belief, and moreover some dispositions that are typically associated with some specific belief though those dispositions are not necessary for holding a belief or a specific belief.

So, more generally, noncognitivists can generally agree with Schwitzgebel that to hold a mental state is to have certain dispositional properties. But they should reject Schwitzgebel's extrem liberalism, in particular they should reject the claim that no disposition is either necessary or sufficient for holding any type of belief. Instead they should claim that certain types of dispositions are necessary and perhaps even sufficient for holding specific types of attitudes. For instance, they should claim that there is a set of dispositions that is definitive of the general attitude-type denoted by the folk psychological term 'belief', and another set of dispositions definitive of the general attitude-type denoted by the folk psychological term 'desire', and so on for each type of mental state.

It should be obvious how this would help with the noncognitivists' claim that moral judgments are beliefs while being noncognitive states. First, noncognitivists would now be in a position to claim that belief and desire are distinct in that their sets of essential dispositions are distinct. Second, if the mental states noncognitivists identify with moral beliefs share some, most, or perhaps all of its essential dispositions with the dispositions definitive of belief, then those noncognitive states might legitimately be regarded as belief. The distinctness between cognitive belief states and noncognitive desire state might then either consists in their having some different essential dispositions, or at least some different peripheral dispositions.

Much more has to be said to make this modified version of dispositionalism plausible and show how it would help with noncognitivism's quasi-realistic claim that moral judgments are can be regarded as beliefs though they are noncognitive states. In particular we would need to know which dispositions are definitive of belief. Then, if we have a clear grasp of the dispositions that constitute belief, noncognitivists would have to show that some noncognitive states are actually constituted by the same dispositions that are definitive of belief, so that they would be justified in claiming they are worth to be called beliefs. I turn to this task after the next section.

In the next section I want do discuss another type of functionalist-dispositionalist theory of belief, which has explicitly been proposed as a philosophical psychology suited for the purposes of noncognitivism. Even though this theory uses a different terminology than Schwitzgebel's, comparing both theories will be instructive and help us to see what might be dispositional properties essential to belief.

5.4.3   Köhler's Conceptual Role Expressivism

Recently, several noncognitivists moved to formulating their theory explicitly as a claim about the functional role of moral belief. One of them is Sebastian Köhler (Köhler, 2017). According to him, moral judgments function to motivate action, whereas descriptive judgments function to represent features of the world. In his terminology, moral judgments are 'conative' states, whereas descriptive judgments are 'representational' states. In this section, I want to give a brief overview of Köhler's noncognitivism about moral judgments. The main point of this overview is not to criticize Köhler's view, with which I am generally sympathetic, but rather to give us some hints as to what noncognitivists might take to be the dispositional-functional properties essential to belief.

According to Köhler, noncognitivism is view that "there is a distinctive difference between normative and descriptive judgments" (Köhler, 2017, 189). We already know that, traditionally, this difference has been spelled out by saying that normative judgments are desires or at least desire-like, whereas descriptive judgments are beliefs. Köhler is well-aware that this way of drawing the distinction is problematic for the quasi-realist who in the end wants to establish that moral judgments also are beliefs. For this reason he, like me, reserves the terms 'belief' and 'desire' for our ordinary, folk psychological notions of belief and desire. His suggestion is that the distinctive difference between normative or moral judgments on the one hand and descriptive judgments on the other, should not be drawn on the level of the "folk psychology" of belief and desire, but on the level of what he calls a "robust philosophical theory of psychology" (Köhler, 2017, 199).

The philosophical theory of psychology Köhler thinks is best suited for helping noncognitivists with drawing their distinction, is "conceptual role semantics", which he regards as a school within functionalism about the mind (Köhler, 2017, 199).[43] He writes:

---

[43]Köhler is interested in conceptual role semantics primarily in order to develop a 'deflationary' account of the propositional content of moral beliefs. But I ignore his broadly Sellarsian proposal here, since my focus is on what he has to say with regard to the functional role of belief, and not with content. Moreover, Köhler explicitly says that his conception of propositional content is not the only one that might be compatible with expressivism or noncognitivism. I myself, however, am not convinced that his proposal will ultimately successful, since I think that his way of dealing with different types of attitude-ascriptions is somewhat piece-meal.

> According to *functionalism about the mind*, mental states are
> dispositional states that are fundamentally characterized by their
> 'functional roles.' The *functional role* of a mental state is its
> causal role within a mental economy. (Köhler, 2017, 199, italics
> original)

What is conceptual role semantics? Conceptual role semantics is a theory
about how and why mental states have the contents they happen to have.
Conceptual role semantics relates to functionalism by its claim that mental
states have their contents in virtue of properties of some parts of their
functional roles.[44] Though mental states may have many functional roles,
not every part of the functional role of a mental state concerns the content
it has. But some parts, according to conceptual role semantics, do, and so
those parts of the functional role that determine the contents of a mental
state might be called the '*conceptual* role' of a mental state.

In the literature on the conceptual role of mental states it is common to
distinguish three different kinds of conceptual role relevant for determining
the content of a mental state. Those are usually called 'mind-entry', 'mind-
to-mind', and 'mind-exit' conditions. Köhler describes them as follows:

> 'Mind-entry' conditions specify the state's role in the procession
> of sensory stimuli. 'Mind-to-mind' conditions specify the state's
> role in the processes of reasoning. And 'mind-exit' conditions
> specify the state's role in the production of actions. (Köhler,
> 2017, 200)

Köhler uses those conditions to draw a general distinction in his philo-
sophical theory of psychology between what he calls *representational* states
and *conative* states, where he explicitly understands 'representational' in
a 'robust', or non-minimal sense. According to him representational states
have a rich set of mind-entry conditions, and a specific set of mind-to-mind
conditions, but only poor mind-exit conditions, if at all. Conative states, on
the other hand, have a poor set of mind-entry conditions, a rather weak set
of mind-to-mind conditions (such as drawing attention to certain things),
but a rich set of mind-exit conditions.[45]

But Köhler not only uses those conditions to draw a general distinction
between representational and conative states. He also thinks that specific

---

[44]This is a rather narrow understanding of conceptual role semantics. This understanding
is not uncommon, but there is also a broader understanding in the literature according
to which it applies not only to the contents of mental states, but also to the meanings of
linguistic item like sentences and that focuses not only on the functional role of mental
states, but more generally on the use of certain entities. On this broad understanding,
every use theory of meaning, and thus also expressivism, might be viewed as as conceptual
role semantics. For more on conceptual role semantics in the broad sense consult (Harman,
1982).

[45]Cf. (Köhler, 2017, 203)

instances of mental states are characterized by specific conditions. More precisely, according to Köhler each mental state is characterized by specific sets of mind-entry, mind-to-mind, and mind-exit conditions. For instance, the belief that grass is green is characterized specific mind-entry conditions, such as being responsive to visual appearances of greenness. Other beliefs are characterized by different sets of conditions. For instance, the belief that snow is white is characterized not by a responsiveness to greenness, but to whiteness. This allows to distinguish very specific instances of mental states. But Köhler believes not only that the specific instances of a mental state are determined by its conceptual role (such as the belief that snow is white and the belief that grass is green). He also believes that what *type* of mental state some mental state is (such as beliefs and desires) are characterized by the conceptual roles they play. In particular he thinks that what type of mental states some state is depends on the *types* of mind-entry, mind-to-mind, and mind-exit conditions that are typical for the mental states of those types. So, for instance, the mental state we ordinarily call 'belief' shares with all other beliefs that they have a common core of conditions. Likewise for other mental states like desires.

With the distinction between representational and conative states in terms of their functional roles, the important question for the quasi-realist noncognitivist becomes whether the conceptual role of the mental states which we ordinarily call beliefs includes only states that have the conceptual role of representational states, or whether some of those states have the conceptual role of conative states. In order to answer this we need to know two things. First, what is the conceptual role of those states we ordinarily calls beliefs? Second, are there some conative states that have this conceptual role of beliefs.

So, what is the conceptual role characteristic of belief? According to Köhler, who is a disciplined syntacticist about truth-aptness and a minimalist about belief, it must be the following:

> On a minimalist conception of belief, mental states are beliefs just in case they have a conceptual role that would make them suitable for expression by a disciplined declarative sentence. This puts some restrictions on the kinds of conceptual role that could characterize beliefs: only conceptual roles with a sufficiently rich set of mind-to-mind conditions to account for the inferential relations necessary to solve the Frege-Geach Problem could do so. (Köhler, 2017, 204)

So what makes a mental state a belief, according to Köhler, is the fact that it is characterized by a set of mind-to-mind conditions that are necessary to solve the Frege-Geach problem. In other words, beliefs, first and foremost, are those mental states that play a certain role in reasoning. Does this answer the first question?

This at least somewhat answers the first question. I say 'somewhat' because I want to mention a potential worry about Köhler's account. Köhler clearly pursues what I have above called the orthodox quasi-realist strategy which was that of arguing from the truth-aptness of moral sentences to their expressing beliefs. In determining which inferential relations (mind-to-mind conditions) a state must have in order to qualify as a belief, Köhler makes explicit reference to the Frege-Geach Problem. It is clear that he means the embedding problem for *truth-apt* sentences and not some other embedding problem for non-truth-apt sentences (like, for instance, imperatives). This would all be fine provided disciplined syntacticism were a viable theory of truth-aptness. But I have argued in chapter 4 that it is not, and that instead truth-apt sentences are those and only those that express beliefs. So one might worry that a quasi-realist who wants to argue from belief to truth-aptness is caught in a circle if he wants to follow Köhler's conceptual role expressivism. One cannot say that a state is a belief if it has the inferential relations necessary to solve the Frege-Geach problem for truth-apt sentence, that is sentences that express beliefs. But I think the circle can be escaped. Instead of giving a linguistic criterion (e. g. "inferential relations necessary to solve the Frege-Geach Problem"), we should give a epistemological criterion like for instance 'having the inferential relations we expect ordinary beliefs to have' or 'the inferential relations epistemologists think beliefs have'. Noncognitivists who want to follow Köhler's conceptual role suggestions, but pursue the new, instead of the orthodox quasi-realist strategy, can then regard any mental state as a belief that has a conceptual role that can account for the inferential relations we expect beliefs to have. Thereby the circularity worry is blocked, and so the first question is answered.

What about the second question from above? Are there some conative states that have this conceptual role of beliefs? If having the right kind of mind-to-mind conditions is sufficient for being a belief, then the second question becomes: are there some states falling on the conative side of the divide that also have this kind of mind-to-mind conditions? Köhler himself does not provide an answer in his (Köhler, 2017), but he is clearly optimistic about answering this in the positive, since he believes that expressivistic theories such as Schroeder's a capable of solving the Frege-Geach problem in terms of, in his terminology, conative states such as being for.[46]

5.4.4   Comparison

There are surely important differences between the general claims of functionalism about belief from above (see assumption 1-5), the phenomenal dispositionalism by Schwitzgebel, and Köhler's conceptual role functionalism. But instead of focusing of the differences, I want to highlight some similarities here.

---

[46]See his (Köhler, 2012).

Many of the above functionalist assumptions seem to reappear in a different fashion and different terminology in the theories of Schwitzgebel and Köhler. For instance, that beliefs are propositional attitudes (assumption 1 above) is something Schwitzgebel also takes for granted.[47] Noncognitivists cannot take that for granted. Contemporary noncognitivists like Köhler, however, take pains to argue that there is a sense in which moral beliefs nevertheless can be regarded as propositional states. Even though much of Köhler's effort is devoted to showing how conative attitudes might be beliefs, he also wants to make sense of their contents, and develops a "deflationary" account of propositions along Sellarian lines.[48]

Furthermore, that at least some beliefs are representational (assumption 2 above) is reflected by Köhler's claim that representational states have rich sets of mind-entry conditions which help them track features of our external environment, and of course he takes some representational states to be regarded as beliefs as well. Schwitzgebel, though he wants his theory to contrast with representationalism, says nothing that would rule out the idea that some beliefs involve the disposition to keep track of features of the world.

Moreover, both Schwitzgebel and Köhler, emphasize that beliefs in conjunction with desires play a crucial role in the generation and motivation of action (assumption 3 above).

As mentioned above, Leitgeb regards assumptions 2 and 3, the representation and motivation functions, as 'input and output' functions of belief. This is clearly captured by the mind-*entry* and mind-*exit* conditions of Köhler. And Schwitzgebel's dispositionalism also includes dispositions to react to changes in the world as well as to produce changes, what he called 'forwards' and 'backwards-looking' dispositions (Schwitzgebel, 2019, 1.4).

Moreover, both Schwitzgebel and Köhler identify or at least associate certain sets of properties with belief they either take to be 'stereotypical' (Schwitzgebel) or 'characteristic' (Köhler). Where Schwitzgebel prefers to speak of dispositional properties, Köhler speaks of functional or conceptual roles, but they do not seem to be so far from each other. Furthermore, Schwitzgebel holds that some dispositions are central to all beliefs, and Köhler likewise thinks that what unites all different kinds of beliefs, descriptive and moral alike, is that they have a certain set of mind-to-mind conditions. So Köhler also thinks that certain properties are central to belief.

---

[47]Schwitzgebel often speaks of the belief that p that is the cluster stereotypical for the belief that p, and that it is stereotypical for the belief that p, that agents are disposed to assert p. Compare (Schwitzgebel, 2002, 264).

[48]See (Köhler, 2018).

5.4.5   Revisiting the Tightrope

Lets now put together what I have said about dispositionalism and function-
alism and revisit the noncognitivist's tightrope. At the end of the section
on Schwitzgebel's dispositionalism, I asked if it might suit the noncogni-
tivists purposes. I raised the problem that his theory might be too liberal
to be suitable for the noncognitivist. The problem was that even though
Schwitzgebel thinks that some dispositions are central to certain beliefs, there
might be exceptional cases in which it is appropriate to describe someone
as holding some specific belief, even though he lacks a disposition that is
central to its stereotype. This led Schwitzgebel to the quite radical claim
that no disposition is either necessary or sufficient for any belief. Some
philosophers might like this extreme liberalism. However, it threatens to
contravene the quasi-realist noncognitivist's claim that descriptive and moral
judgments both merit the label 'belief', and yet are interestingly different
kinds of mental states.

   Köhler's theory, in contrast, offers a solution to the noncognitivists
tightrope problem precisely by providing at least a necessary condition for
belief, namely that the states must have the right kind of mind-to-mind
conditions. Might Schwitzgebel not do the same?

   Even if there is clearly no straightforward translation of Schwitzgebel's
theory into Köhler's theory, it is pretty natural to think of Köhler's mind-
to-mind conditions for beliefs as the set of central dispositions, and his
mind-entry and mind-exit conditions as somewhat peripheral properties. So
what noncognitivists sympathetic to Schwitzgebel's dispositionalism should
say is that being disposed to perform certain kinds of inferences or reasonings,
is not only central but necessary for belief.

   But noncognitivists need more, they also need a sufficient condition for
a set of dispositions to count as a belief in order to establish that some
noncognitive states can be regarded as moral beliefs. I think that even
though having the right set of mind-to-mind conditions, that is playing the
right role in reasoning, might be necessary for a state to count as belief, it is
not by itself sufficient. What additional conditions could be required to make
it sufficient? It cannot be that a state in addition has certain mind-entry
and mind-exit conditions because then we would again threaten to collapse
the difference between moral and descriptive beliefs. But we could say that
in order for a state to count as a *descriptive* belief is that it must have a
specific set of mind-to-mind as well as a specific set of mind-*entry* conditions.
Moreover, we could say that for a state to count as a *moral* belief it must have
the right mind-to-mind relations and a specific set of mind-*exit* conditions.
So, in each case having the right mind-to-mind conditions would be necessary
for holding a belief, but insufficient. In order to be sufficient it must either
in addition have the right mind-entry or the right mind-exit conditions.
Borrowing from Schwitzgebel's terminology, noncognitivists could say that

though all beliefs have the same center, descriptive and moral beliefs have different peripheries.

Köhler makes a similar suggestion and concludes that such a functional-dispositional theory of psychology is compatible with the noncognitivists claim that moral and descriptive judgments are fundamentally different, even though they are both beliefs.[49] But he also notes that an important task remains for the noncognitivists, namely that of establishing that *there are* mental states that have the set of mind-to-mind conditions necessary for belief, but also the mind-exit conditions characteristic of conative states.

5.4.6 Putting Everything Together

In order to avoid Schroeder's global noncognitivism, I suggested above (section 3.7) a particular interpretation of the structure of the mental states invoked by attitude semantics. I suggested to replace the outer 'being for' with 'disposition', with the consequence that we no longer have to accept Schroeder's counterintuitive claim that *all* judgments are noncognitive states. Instead we received the much more common view that they are all dispositions.

We now see that this proposal fits nicely with general views about the nature of belief as defended by functionalists and dispositionalists. This provides dispositional expressivists with an independent justification to treat the dispositional states expressed by moral sentences as moral beliefs. Moreover, this proposal not only connects to familar theories in the philosophy of belief, but it also suits the purposes of noncognitivists, and might even help with some of its most notorious problems in the philosophy of mind and epistemology as I shall argue.

Against the background of what I have said about Schwitzgebel's and Köhler's theories of belief, I want to highlight several points about my proposal to replace global noncognitivism with global dispositionalism.

My above proposal treats the mental state expressed by 'murder is wrong' as a certain disposition. It avoids the anti-behaviorist worries against traditional dispositionalist views of belief by being decidedly 'phenomenal' in Schwitzgebel's sense: to believe that murder is wrong is not just a disposition to act in some way, but a disposition to be in some other mental state, namely the disposition to *be for* that murder is blamed.

It also helps with balancing the tightrope. On my proposal there is a clear difference between the states expressed by descriptive and by moral sentences. Whereas 'grass is green' expresses the disposition to occurrently *believe* (or to proceed as if, or to judge, if you prefer) that grass is green, 'murder is wrong' expresses the disposition to *be for* that murder is blamed. The difference can now be explained by the difference between occurrently

---

[49]Cf. (Köhler, 2017, 204).

believing and being for, where the former is a cognitive and the latter a noncognitive state which I have assumed to be categorically distinct.

Yet both states might be regarded as beliefs. This is so because the dispositional state expressed by 'murder is wrong' has the mind-to-mind conditions suitable for solving the Frege-Geach Problem. This of course was one of the major goals of my semantic proposal in chapter 3. Just as a reminder, attitudes semantics allows us to say that a moral sentence like 'murder is wrong' is inconsistent with its negation because it expresses a state that is rationally incoherent with the state expressed by its negation, and it explains the logical relations it can stand in to other (logically complex) sentences. Likewise for descriptive sentences like 'grass is green' and all other sentences of the language. The only assumption noncognitivists now need to make is that it is rationally incoherent to be disposed to belief that grass is green while also being disposed to believe that grass is not green, and *mutatis mutandis* for being disposed to be for that murder is blamed. But that assumptions seems fairly plausible. Nevertheless, there might be some problems, which I will mention in a moment.

Above I suggested, in line with what Köhler says, that having dispositions with the right mind-to-mind conditions and the right mind-*entry* conditions is necessary and sufficient for a state's being regarded as a *descriptive* belief. Moreover, I suggested that having dispositions with the right mind-to-mind conditions and the right mind-*exit* conditions is necessary and sufficient for a state's being regarded as a *moral* belief. Noncognitivists can now claim that the reason why descriptive belief is different from moral belief is that there is no state that has the right mind-to-mind conditions as well as the right mind-entry and mind-exit conditions. In Köhler's terms no state is both a representational and a conative state, even though both might be what folk psychology classifies as belief in virtue of those state's having the right properties to feature in pieces of rational reasoning.

If what I have argued so far is correct, the following is how dispositionalist noncognitivists should describe their theory. They should start with their theory of moral belief according to which holding a moral belief is to be disposed to be in a certain noncognitive state such as being for. They could, for instance, say that to believe that murder is wrong is to be disposed to be for blaming murder. They should then argue that this disposition has all the rational properties we assume the folk psychological belief to have. For instance, it is incoherent with the state we think of as the belief that murder is not wrong. The reason that those beliefs are incoherent is that this latter belief is the disposition to be for not blaming for murder, which is incoherent with being disposed to be for blaming murder. Likewise, for descriptive beliefs such as the belief that grass is green which is the disposition to occurrently believe that grass is green. By the same token, noncognitivists can even explain why it is rationally incoherent to believe the premises of a piece of valid theoretical reasoning (such as modus ponens

reasoning) while not believing the conclusion. Even though all beliefs are dispositions, there is a fundamental difference between descriptive beliefs and moral beliefs. Whereas descriptive beliefs are dispositions to occurrently *believe* a proposition, moral beliefs are dispositions to *be for* a proposition. Both occurrently believing and being for involve the right set of mind-to-mind conditions. Additionally, to occurrently believe involves a rich set of mind-entry conditions, and to be for something involves a rich set of mind-exit conditions. That is why even though both types of dispositions are different, they can both be called beliefs. But moral beliefs though they share with conative or noncognitive states the property of having a rich set of mind-exit conditions, also differ from paradigmatic noncognitive states, such as desires, in that the latter do not have the mind-to-mind conditions of necessary for a state's being regarded as a belief.

What does all this mean? Now that I have argued that the dispositions which attitude semantics assigns to sentences can be regarded as beliefs, we know that they can legitimately figure as the sincerity conditions for assertions. In other words, it is now possible for expressivists to claim that the sentences of the formal language in chapter 3 can be used to make assertions. By the account of truth-aptness developed in chapter 4 it follows that the formulas to which the expressivist semantics assigns those dispositions can be regarded as truth-apt. This completes what I have above called the belief-to-truth-aptness argument.

The arguments in this chapter suggest that noncognitivists should embrace the following view:

Dispositional Noncognitivism: Moral judgments are dispositions to be in noncognitive states.

By choosing a suitable noncognitive state, such as the state of being for, dispositional noncognitivists are able to solve the Frege-Geach problem as well as earning to right to speak of moral judgments as being moral beliefs. Since dispositional noncognitivists have thus earned the right to speak of moral beliefs, they have also earned the right to treat sentences expressing those states as being truth-apt because every sentence that expresses a state that can be regarded as a belief, is truth-apt.

5.4.7 Prospects and Problems for Dispositional Noncognitivism

I argued that dispositional noncognitivism allows noncognitivists not only to solve the Frege-Geach problem, but also to earn the right to moral belief and moral truth-aptness. Now is the time to see how it fares with respect to the above-mentioned problems for noncognitivism.

Start with the *one word problem*. The problem was to give an expressivist semantics for 'believes that' that is compatible with the idea that descriptive beliefs are different kinds of states from moral beliefs. Here dispositional noncognitivism can use the solution proposed by Schroeder above: sentences of the form 'S believes that P' express the descriptive belief that S is in the state expressed by P. For instance, if P is 'murder is wrong' then 'S believes that murder is wrong' expresses the descriptive judgment that S is disposed to be for blaming for murder. Since to hold a descriptive judgment is also a disposition, someone who accepts the belief-ascription must be disposed to occurrently believe (or proceed as if) that S is disposed to be for blaming for murder. This is a mouthful, but that might also be one of the reasons why language has evolved in a way that it enables 'belief'-talk.

Above I noted that this way of dealing with the one word problem has a problem with being generalizable to other kinds of attitude-ascriptions, desire- or hope-ascriptions, for instance. Part of the problem is that noncognitivists also have to solve the many attitudes problem. How does dispositional noncognitivism fare with respect to the many attitudes problem?

The *many attitudes problem* is that noncognitivists not only have to offer a theory of moral belief, but of many other moral attitudes as well such as moral desires, intentions, hopes, fears... Dispositional noncognitivism might offer a surprisingly simple solution to the problem. To believe that murder is wrong, I suggest, is something like being disposed to be for that murder is blamed. Why not say that to hope, say, that murder is wrong is being disposed to hope that murder is blamed, and to fear it is to be disposed to fear that murder is blamed and so on? This would yield a perfectly general solution to the many attitudes problem.

I do not want to investigate the plausibility of this proposal in greater detail here.[50] Even if this proposal would prove not to work for some reason, dispositional noncognitivism is compatible with the alternative solutions to the many attitudes problems that have been proposed in the literature.[51]

What about the *moral attitude problem*? Nothing I have said about dispositional noncognitivism answers the question as to what makes an attitude a moral attitude. I have prefered to remain neutral as to what makes the dispositions expressed by moral sentences genuinely moral attitudes. For the sake of simplicity I decided to stick as closely a possible to Schroeder's

---

[50]This proposal surely raises a bunch of questions. Why, for instance, is it that only in the case of moral belief we have to interpret belief as a disposition to be *for*, whereas in the case of moral hopes or moral fears we can interpret it as a disposition to *hope*, and a disposition to fear, respectively? What, it might be asked is so special about belief, that noncognitivists have to analyze it away, but not hope or fear? One possible answer might be that for the noncognitivists belief is the really problematic state because belief is generally regarded as being a representational state, and representation of moral facts is problematic. Hopes and fears etc, are not in the business of representing moral facts and are therefore not equally problematic as beliefs.

[51]For some such recent proposals see (Köhler, 2013) or (Beddor, 2020).

toy proposal to treat being for as the general attitude and the blaming relation corresponding to 'wrong', by saying that to believe that murder is wrong is to be disposed to be for blaming for murder. I prefer not to take a stand on whether this proposal provides a good or the correct analysis of moral belief or whether there is some better alternative. A crucial question noncognitivists have to answer about this specific proposal is whether the state of being disposed to be for blaming for something is a moral attitude due to that fact that being for is a moral attitude, or blaming is a moral attitude, or whether what makes it a moral attitude is the fact that they are combined in the way they are.

Though I prefer not to dig deeper into this problem, I want to mention one advantage dispositional noncognitivism might have with respect to the moral attitude problem that more traditional noncognitivists theories lack. Traditional theories often hold that to have a moral belief like the belief that murder is wrong, is to hold some noncognitive attitude like disapproval towards what appears to be its immediate object, for instance the state disapproving of murder. This raises the question why this should be a moral belief. We can disapprove of all kinds of things, but this does not turn every disapproval state into a moral belief. This is where dispositional noncognitivists might be better off, since they can make moderate use of Schwitzgebel's liberalism. They might say that the disposition to be for blaming for murder is central, in fact, necessary for the belief that murder is wrong in virtue of the fact that it has the right mind-to-mind and mind-exit conditions. However, they might also say that *typically* moral beliefs come with a suite of additional dispositions as well, such as for instance the disposition to feel guilty or ashamed when you have committed murder, to be disposed to prevent others from committing murder, to be disposed to be angry with those who murder, to be disposed to approve of those who also are for blaming murder and so on and so forth. Those who have raised Moorean open questions about the noncognitivist analyses of moral belief might perhaps be somewhat tamed by this maneuver.

Köhler, however, disagrees. In a footnote he writes that analyzing moral belief as a "complex dispositional state" does not close off Moorean questions (Köhler, 2013, 486). His own proposal to solve the moral attitude problem uses the Ramsey-Lewis method which, I have remarked above, is also used by functionalists about belief. However, even though I shall not investigate this matter further here, nothing in Köhler paper suggests that he thinks the above dispositional claims might not be a able to figure as 'platitudes' in the "folk theory of moral thinking" which then ramseyfies into a philosophical theory of moral belief (Köhler, 2013, 504). So the idea of allowing some liberalism about moral attitudes seems compatible with Köhler's functionalist approach of making all the above-mentioned dispositions part of the long definite description that characterizes the moral attitude which noncognitivists take to constitute moral beliefs.

Moreover, it should be noted that dispositional noncognitivism can be viewed as a version of what Miller (see section 5.2.3) called the 'higher-order sentiment'-approach to the moral attitude problem, since they are dispositions to be in attitudes (e.g. being for) towards attitudes (e.g. blaming). Above we have seen that Blackburn thinks this might be a viable approach if we allow that there is no 'precise point on the staircase to the ethical'. This fits nicely with the liberalism of Schwitzgebel's dispositional theory of belief according to which there similarly is no precise set of dispositional properties that clearly mark the difference between someone's believing that x is better than y and someone's valuing x over y.

Dispositionalism also seems to be a fruitful idea when it comes to problems the noncognitivist has in the theory of action, though in the final chapter I shall argue that the idea that noncognitivism is better at explaining moral motivation than cognitivism is a false dogma of metaethics. Take the thesis of motivational internalism, which is the view that there is an internal conceptual connection between moral belief and motivation. Michael Ridge has remarked that "the conventional wisdom is that that the natural dialectical beneficiary of the truth of motivational internalism is expressivism [noncognitivism]" (Ridge, 2015, 135). If holding a moral belief is *identical* with being in some desire-like motivational state, then, it is obvious, at least prima facie, why the connection between normative belief and motivation is so tight.[52]

It has, however, recently been objected by Caj Strandberg (Strandberg, 2012) that expressivism threatens to make the connection between moral belief and motivation *too* tight, thereby ruling out cases of amoralism where an agent holds a moral belief, but lacks any motivation to act according to his belief. The worry is that if moral beliefs *are* motivational states, then how is it possible that one can hold a moral belief, and not be motivated? One idea to deal with this objection, Strandberg notes, is to identify moral belief not with some desire-like state, but with a disposition to be in some desire-like state. The idea is to utilize the fact that dispositions do not have to manifest. So, if holding a moral belief is identified only with a *disposition* to be in a desire-like state, but not with the desire-like state itself, it is clear that the relation between moral belief and motivation can be internal without being necessary because one can be disposed to be in a motivational state (like being for), yet not be in it. So, dispositional noncognitivism makes room for amoralism.[53]

---

[52]In chapter 8 of this dissertation I shall argue, against 'conventional wisdom', that noncognitivism is actually worse off than cognitivism at explaining the relation between moral belief and motivation.

[53]Strandberg's goal in (Strandberg, 2012) is to show that the dispositional strategy to handle amoralism fails because the problem to explain the possibility of amoralism reappears on the level of occurrent moral belief. However Eriksson (2014) convincingly replies that noncognitivists can resist Strandberg's argument if they deny that being in a desire-like state necessarily involves being motivated. I fully agree with Eriksson that

All in all, dispositional noncognitivism opens up a number of promising possibilities for dealing with the above-mentioned problems that bother more traditional forms of noncognitivism.

5.4.8 Specific Problems for Dispositional Noncognitivism

Although dispositionalism might help to deal with some notorious problem of noncognitivism, dispositional noncognitivism also has some problems of it own. I shall mention two of them.

One rather subtle problem specifically for dispositional noncognitivism is that it, unlike its Schroederian relative, no longer offers an analysis of disbelief, in particular of occurrent disbelief. Here is why. According to dispositional noncognitivism the states expressed by sentences are certain dispositions, and those dispositions have all the marks of belief. Following Robert Audi (Audi, 1994), it has become standard to distinguish between dispositional beliefs and occurrent beliefs, where the former are regarded as rather stable or permanent states, and the latter are episodic events in the minds of agents. It is natural to think that the disposition which the noncognitivist thinks is expressed by a sentence like 'murder is wrong' is a dispositional belief. This raises the question of what according to the dispositional noncognitivist is an occurrent belief. The natural answer seems to be this: if to dispositionally believe that murder is wrong is to be disposed to be for blaming for murder, then to occurrently believe that murder is wrong is just to occurrently be for blaming for murder. So far so good. But if to dispositionally disbelieve that murder is wrong is to be disposed to *not* be for blaming for murder, then by the same token to occurrently disbelieve it is to occurrently not be for blaming for murder. But this simply means that you occurrently lack some being for state. But lacking a state is different from disbelieving something. Hence, by moving from global noncognitivism (all beliefs are states of being for) to global dispositionalism, noncognitivists ruin their neat account of disbelief. What is particularly weird is that while it is still possible to dispositionally disbelieve something, it is not possible to occurrently disbelieve something. Perhaps this problem can be solved, but I think it is also not much of an issue if dispositional noncognitivists simply accept it. After all, their goal was not to provide an account of disbelief, but an account of moral belief.

There is another problem which is more pressing. It also has to do with the distinction between dispositional and occurrent belief. It is a version of what has been called the 'disjunction problem'.[54] The disjunction problem

---

desire-like states are not necessarily motivating. In chapter 8 below I emphasize the even stronger claim that desires do not motivate.

[54]Despite its name the disjunction problem arguably applies to other connectives to. For the disjunction problem see (M. Schroeder, 2011, unpublished) and (Silk, 2015).

is that the noncognitivists' theory of belief must show that it is possible to believe a disjunction without believing either disjunct. Take as an example some purely descriptive belief in a disjunctive content such as the belief that p or q. It is surely possible to hold this belief without believing either p or believing q. According to dispositional noncognitivism to believe that p or q is to be disposed to pai p or pai q.[55] It is clearly possible to be so disposed without either being disposed to pai p or be disposed to pai q, which are the belief that p and the belief that q, respectively. But what about occurrently believing p or q? Again following the natural assumption of the previous paragraph, to occurrently believe p or q, according to the dispositional noncognitivist, is to simply pai p or pai p. But it is not possible to pai p or pai q without either pai p or pai q, which are the occurrent belief that p and the occurrent belief that q. In other words, if dispositional noncognitivism is true, it seems impossible to occurrently believe a disjunction without occurrently believing either disjunct. Strangely, as before with the problem of occurrent disbelief, the phenomenon concerns only occurrent disjunctive belief, not dispositional belief. The task for the dispositional noncognitivist is either to provide a solution to the problem of occurrent disjunctive belief, or provide an explanation of why occurrent belief is different in this regard and hence claim that the problem is not a problem. I leave this as an open problem.

5.4.9   General Problems for Dispositionalism

Due to the fact that dispositional noncognitivism is a dispositionalist theory of belief, it also inherits the problems dispositionalism already has. One such problem it that all dispositionalists have to tell us something about the nature of the dispositions involved in their theory. I have in my discussion completely ignored the important question of what dispositions are. The nature of dispositions is highly controversial, and I wanted to avoid having to take a stand on whether disposition-claims are best understood as conditionals or counterfactuals or what Fara (Fara, 2005) calls 'habituals', how to deal with masked and finkish dispositions, what are the manifestation conditions of the relevant dispositions are, and the complicated issues surrounding categorical bases. It is pretty likely that some conceptions of dispositions may suit the needs of dispositionalists better than others, and that perhaps a suitable conception will be rather technical and not have much in common with our ordinary understanding of dispositional properties such as a vase's fragility. One difference, for instance, is the following: though some of our beliefs may be fairly robust and hard to change, beliefs also can change quickly. For example, when I look outside the window and see that the sun is shining I will instantly and automatically revise my previously held

---

[55]Note that it is not being disposed to pai p or q, in which case the disjunction problem would not arise.

belief that it is raining outside. This does not quite fit with our ordinary understanding of dispositions as being fairly stable properties of things.

Another problem one might see is the following. Beliefs can stand in rational relations because rationality supervenes on the mind, and beliefs are states of the mind. But it is not clear what conception of dispositions is compatible with our thought that the dispositions making up the beliefs are mental states and so stand in rational relations. But if dispositions cannot stand in rational relations, how can they be beliefs?[56]

A number of additional problems against dispositionalism about belief in general and against Schwitzgebel's specific version of dispositionalism in particular, have also been discussed for instance by (Carruthers, 2013, 150ff). There is also a very general objection made by (Fassio, 2013) who argues that dispositionalism about belief has what he calls 'blind spots'. According to him if dispositionalism is true, then there are some beliefs that cannot be manifestly or occurrently believed. This, he argues, is problematic because every disposition must be manifestable at least in principle. Unfortunately, I lack the space here to deal with all those interesting and challenging objections, and so I have to leave them for some future occasion.

At the end of chapter 3 I made a proposal for an interpretation of the structural solution to the Frege-Geach problem. The proposal has been to replace 'being for' with 'disposition'. My motive there has merely been to avoid the radical view that all beliefs are noncognitive states of being for. The aim of the present chapter has mainly been to provide an independent justification for that seemingly ad hoc dispositionalist proposal by showing how it can be related to a view about the nature of belief that many philosophers, independently of noncognitivism, have found and still find plausible, namely the view that to believe something is to have certain dispositions. Before I end this chapter, I want to mention two other sources of justification for noncognitivism to 'go dispositional'. The first is that the idea of dispositional noncognitivism does not seem to be new, but has already been suggested by one of the great figures of noncognitivist metaethics: Charles L. Stevenson. The second is that the idea that to think something valuable or of value is to be disposed to be in a noncognitive state has been advocated even by clearly realist-inclined metaethicists like John McDowell and David Lewis.

5.4.10   The Emotional Aura of Words

Dispositionalism noncognitivism not only connects to general views about the nature of belief, but also has a historic precursor in one of the most influential figures of expressivism or noncognitivism, namely Charles L. Stevenson. Stevenson wanted to give a 'psychological' definition of meaning.

---

[56]For an argument that requirements of rationality should be regarded as governing only occurrent attitudes see (Lee, 2018).

The meaning of a word, according to him is a "tendency (causal property, dispositional property)" to have certain "psychological causes and effects" (Stevenson, 1937, 22). The kind of meaning he was most interested in was, of course, what he called 'emotive' meaning, which he defined as follows:

> The emotive meaning of a word is a tendency of a word, arising through the history of its usage, to produce (result from) *affective* responses in people. It is the immediate aura of feeling which hovers about a word. (Stevenson, 1945, 23, italics original)[57]

Stevenson explicitly defended a dispositional theory of *meaning* according to which "meaning is taken as a disposition of a sign to produce psychological reactions".[58] What distinguishes, in his view, *descriptive* from *emotive* meaning is the nature of the particular psychological reaction a sign is disposed to produce. According to him, "descriptive meaning is the disposition of a sign to affect cognition" or "produce *cognitive* mental processes". Likewise for the emotive meaning of a sign which is "a disposition to evoke attitudes" or produce *noncognitive* mental processes.[59]

But, Stevenson not only suggested a dispositionalist semantics, but also a dispositional theory of mental states. He favors analyses which take mental states to be "dispositions to *action*" (Stevenson, 1945, 66, emph. orig.). A belief or attitude, according to him, is a "complicated conjunction of dispositional properties" (Stevenson, 1945, 60). Combining Stevenson's dispositionalist theory of meaning with his dispositionalist theory of mental states, gives his theory an interesting twist: if the psychological responses that a sign is disposed to produce are themselves dispositions, then "meaning is in part a second-order disposition" (Stevenson, 1945, 60). This suggests that Stevenson, even though he did not offer a formal semantics, had a theory of meaning in mind that comes very close to what I have called dispositional expressivism, even structurally. With Stevenson the dispositional expressivist can say that the meaning of 'murder is wrong' consists in this sentence's disposition to produce an affective response, namely the being for that murder is blamed, where the being for is itself a disposition to produce the attitude of blaming for murder.

Stevenson suggests that the task of a "psychological (or 'pragmatic')" semantics consists in a systematic assignment of dispositions to signs of a language (Stevenson, 1945, 79). Stevenson did not attempt to provide a

---

[57]Stevenson takes this idea from a passage from Ogden and Richards who speak of the "emotional aura" of words. They claimed that 'good' in 'this is good' has a "purely emotive use [...] it serves only as an emotive sign expressing our attitude to *this*, and perhaps evoking similar attitudes in other persons, or inciting them to actions of one kind or another" (Ogden & Richards, 1923, 125, italics original).

[58]The following citations in this paragraph are from (Stevenson, 1945, 77,67,62,66,60).

[59]Note the parallel here to Schwitzgebel's phenomenal dispositions, as well as to Köhler's mind-to-mind conditions.

solution to what he thought of as "one of the most difficult problems that meaning-theory includes—that of explaining how separate words, each one with its own meaning, can combine to yield sentence-meanings" (Stevenson, 1945, 67). But even though he did not attempt to solve the compositionality problem, and thus the Frege-Geach problem, he thinks that it is feasible "to take each word as having a disposition to affect cognition, just as the full sentence does. The problem reduces, then, to one of explaining the interplay of the dispositions of several words, when realized conjointly" (Stevenson, 1945, 67).

I take it that *dispositional expressivism* with its attitude semantics provides what Stevenson only promised: it solves the expressivist's problem of compositionality by showing how the dispositions expressed by complex sentences can be functions of the dispositions expressed by their parts.[60]

5.4.11   Cognitivist Dispositionalism

With Stevenson we have a central figure of noncognitivism who already embraced the idea that meaning as well as mental states should be understood in dispositional terms. But not only noncognitivists find that idea attractive. Dispositional noncognitivism also has striking similarities with views proposed by philosophers not normally associated with expressivism, noncognitivism, or anti-realism. For instance, John McDowell's view in (McDowell, 1985) that moral properties must be understood as secondary qualities in analogy to color properties also surely admits of a dispositional reading, especially if dispositions like secondary qualities are analyzed in terms of counterfactuals.[61]

Another example for a rather cognitivist dispositionalist is David Lewis. In his Lewis et al. (1989), he defended a dispositional theory of value. There he proposes an analysis of value according to which, roughly, "values are what we are disposed to value" (Lewis et al., 1989, 113). Valuing something, he goes on "is some sort of mental state, directed toward that which is valued", and "valuing is a *favorable* attitude" just like desiring (Lewis et al., 1989, 113f.). If Lewis had stopped here, then he would have proposed a theory of value that identifies some thing's being of value, or good, with disposing one to desire it. This idea would have been structurally identical with Schroeder's theory.[62] However, Lewis immediately adds that "we'd

---

[60]More on Stevenson's dispositionalism can be found in (Boisvert, 2016).

[61]For more on dispositions, secondary qualities and counterfactual analyses of them see for instance (Bird, 2007).

[62]Schroeder does not offer a proposal for 'good', but it is clear that since the semantic contribution of the predicate 'wrong' is the blaming-for relation, and 't is wrong' expresses being for blaming for t, 't is good' will express something like being for *praising* for t. If we replace 'being for' with 'being disposed' and 'praising' with 'desiring' a Schroederian dispositionalist account of 'good' could say that believing that something is good is being disposed to desire it, which closely mirrors what I have just said about Lewis' account.

better not say that valuing something is just the same as desiring it" (Lewis et al., 1989, 115). Rather, in his view, "valuing is just desiring to desire" (Lewis et al., 1989, 115), which leads him directly to the view that "to be a value—to be good, near enough—means to be that which we are *disposed*, under ideal conditions, *to desire to desire*. Still more complicated, still more plausible"(Lewis et al., 1989, 116, italics added). So, even though Lewis' theory seems to lean more toward realism than toward anti-realism,[63] he independently of noncognitivism argued for the view that to think that something is good is being disposed to desire to desire it, which structurally resembles the noncognitivists proposal for moral belief that I have tried to defend. Moreover, Lewis believes that his theory answers Moore's open question argument and explains motivational internalism, which, I have explained at the beginning of this thesis, are standardly cited as two of the main reasons for being a noncognitivist.

5.4.12 Conclusion

My aim in this chapter has not been to develop and defend a full-blown dispositional noncognitivist theory of moral belief, but rather to sketch what I generally take to be a promising direction for such a theory. A full-blown dispositional noncognitivist theory would have to say much more about what the relevant dispositions are, which involves saying something about the nature of dispositions in general, and the specific dispositions of belief and moral belief in particular. Moreover, a more worked-out theory would have to be much more specific about the nature of the attitudes moral believers are disposed to hold, e. g. if it is being for, desire, some other familiar pro-attitude, or some sui generis moral sort of approval.

In this chapter, I have argued that dispositional noncognitivism, the view that moral beliefs are dispositions to be in noncognitive states, has many attractions and is worth to be investigated further. It connects to a central position in the philosophy of belief, and has precursors in noncognitivism as well as in positions associated with moral realism. Moreover, many of the traditional problems appear in a new light when approached from dispositional noncognitivism, most importantly what I have called the tightrope problem. Yet there also remain several problems. Some of those are problems specifically for dispositional noncognitivism e. g. occurrent disbelief and disjunctive belief, internalism. Others are problems generally for dispositionalism e. g. the nature of dispositions, blind spots etc.[64]

Even though I dealt with a number of problems for noncognitivism, there are still several problems which I ignored in this section. The reason why

---

[63]Lewis himself finds the question whether his theory of value is realist or antirealist "hard" (Lewis et al., 1989, 114).

[64]Interestingly, most of the just mentioned problems have to do with what happens to occurrent belief.Perhaps there is a single solution for all of them, namely by dropping the

I ignored them is that I think that those problems are pretty complex and therefore deserve a more detailed treatment. In the next three chapters I will discuss three different problems for noncognitivism.

In the next chapter, (chapter 6), I will discuss an objection that has been leveled by Derek Baker against Schroeder's specific form of noncognitivism in terms of being for. *Baker's objection* is that, roughly, Schroeder's noncognitivism cannot explain certain rational properties of moral belief and must therefore be false. I shall argue that the objection fails on grounds that it rests on mistaken premises. I discuss Baker's objection because my dispositional noncognitivism is closely modeled on Schroeder's noncognitivism. By showing that Baker's problem does not arise for Schroeder's version of noncognitivism, I can show that it is also highly unlikely that it could be problem for my dispositional version of noncognitivism.

In chapter 7 I discuss what is conceived of as one of the most thorny problems for noncognitivism, Cian Dorr's *wishful thinking objection.* I suggest a solution to this problem which is independent of dispositional noncognitivism. I will argue that sometimes one is epistemically justified in believing something on the basis of a desire.

In chapter 8, finally, I will change my tune. Up to this point I will have tried to defend metaethical expressivism and noncognitivism against various objections. Then, however, I shall raise two novel problems for noncognitivism having to do with how moral beliefs motivate action. Both *motivation problems* are surprisingly simple, but have yet been completely unnoticed in the literature. I suggest some ways to solve those problems, but leave it as an open question whether these solutions are ultimately convincing.

---

seemingly natural assumption, mentioned above, that the belief that is identified with some set of dispositions is a dispositional belief. If there is some other way of distinguishing between dispositional belief and occurrent belief than merely taking the latter to be the manifested dispositions, then all of the above problems might disappear.

Chapter 6

Expressivism and Rationality

In this chapter I, reply to an objection leveled against Mark Schroeder's specific form of noncognitivism in terms of the attitude of being for as he developed it in (M. Schroeder, 2008a). The objection, in a nutshell, is that given the noncognitive states Schroeder's semantics assigns to sentences, the noncognitivist cannot explain why the states expressed by certain special sentences are rationally incoherent. The objection fails, I will argue, because it rests on mistaken assumptions about what is and what is not a requirement of rationality. As a result of this discussion we shall see that it is also highly unlikely that Baker's objection could successfully be leveled against my structural solution to the Frege-Geach problem and the dispositionalist interpretation of it.

6.1   Baker's Objection from Rationality

In his paper 'Expression and Guidance in Schroeder's Expressivist Semantics' (Baker, 2018) Derek Baker raises an objection to expressivism as it has been developed by Mark Schroeder in his (M. Schroeder, 2008a). Baker argues that Schroeder's expressivist (1) is committed to certain sentences expressing rationally incoherent states of mind, and he objects (2) that the expressivist cannot explain why these states would be rationally incoherent. The aim of this chapter is to show that Baker's argument for (1) is unsound, and that (1) is unlikely to be true. This obviates the need to explain the alleged rational incoherence, and so Baker's objection, (2), to Schroeder's expressivism is undermined.

Baker's objection is based on a particular argument for (1). Baker argues that if the view Schroeder developed on behalf of expressivism is true, then the following sentence "must be expressing inconsistent psychological commitments" or an "inconsistent state of mind" (Baker, 2018, p. 838):

(S) 'Murder is wrong, but not blaming for murder makes sense'.

So, Baker's argument is supposed to establish the claim (henceforth: Baker's claim) that the mental state expressed by (S) is, as I prefer to say, *rationally incoherent*, and his objection is based on that claim. As before I stipulate that a mental state or set of mental states is rationally incoherent iff by simultaneously holding these states an agent violates some requirement of rationality.

Baker's objection to Schroeder's expressivism is *not* that the above sentence, (S), intuitively does not express a rationally incoherent mental state.[1] Rather Baker accepts the incoherence and objects that Schroeder's expressivism is unable to explain the rational incoherence of the state expressed by (S), at least without raising serious problems elsewhere in the expressivist theory.

The aim of this chapter is to show that Baker's argument is unsound, and so does not establish Baker's claim about the incoherence of the state expressed by (S). Moreover, I shall argue that even more promising arguments for that claim also fail and so it is unlikely that the state expressed by (S) is rationally incoherent. But if Baker's claim is false, this undermines his objection to expressivism: one cannot object to expressivism that it is unable to explain why a state is rationally incoherent, when in fact it is not rationally incoherent.

6.2   Baker's Objection

Let us start by taking a closer look at Baker's objection and Schroeder's expressivism. On Baker's understanding, (S) is a conjunction of two atomic first-order language sentences. The first conjunct, 'murder is wrong', contains the normative predicate 'wrong' applied to 'murder'. The second conjunct, 'not blaming for murder makes sense', contains the normative predicate 'makes sense' applied to 'not blaming for murder'.[2] Baker primarily, though

---

[1]This is surprising because Baker, I think correctly, says that "the claim is not that the utterance 'Murder is wrong, but one ought not to blame for murder' sounds infelicitous; I am happy to acknowledge it does not" (Baker, 2018, p. 838). Baker suggested to me (in personal communication) that despite its felicity (S) may nevertheless express an incoherent state, but that the incoherence need not be obvious. However, if the expressed incoherence is non-obvious, this clearly shifts the burden on Baker to provide a convincing argument for the incoherence which, if my arguments in this chapter are correct, Baker does not do.

[2]We should not treat 'not blaming for murder' (or 'notblamingformurder') as a singular term in a formal first-order language. At least, we must be aware that if we treat it as a singular term, then the alledged rational failure of believing that notblamingformurder makes sense while also believing that blamingformurder makes sense cannot be explained by the logical meaning of 'not', since 'not' does not contribute to the meaning of the term 'notblamingformurder' in the usual way. But if 'not blaming for murder' is not a singular term, however, this is a problem for Baker's objection because his whole discussion assumes that 'makes sense' functions like a predicate in order to be able to use Schroeder's semantic clause for atomic formulas (see chapter 3 above, and below). I will ignore this problem because I am primarily concerned with Baker's argument on which his objection rests,

not exclusively, uses 'makes sense' in his paper as the normative predicate in the second conjunct, but he suggests that it could be replaced by 'is the thing to do', 'is rational', 'is reasonable', 'is fitting', 'is appropriate', or even 'one ought to'.[3]

The heart of Schroeder's metaethical expressivism, as we know from chapter 3, is a formal semantics for a first-order language containing moral or normative predicates like 'wrong'. This expressivist semantics provides a recursive mapping from every sentence of the language to the mental state it expresses. Schroeder, recall, stipulates that in his particular form of expressivist semantics, all sentences express noncognitive states of *being for*, even non-normative sentences, since he believes that otherwise expressivists will not be able to provide a constructive solution to the Frege-Geach problem. Being for is supposed to be a noncognitive, desire-like, motivational state that takes complex mental properties as its content. For atomic sentences, Schroeder offers the following semantic clause:

Atom For every predicate F and singular term t, sentences of the form $F(t)$ express $\text{FOR}(\psi(t))$, where $\psi(t)$ is a complex property one is for having, namely the property of standing in the ψ-relation to t.[4]

Schroeder suggests that 't is wrong' expresses being for blaming t, or slightly more formally: $\text{FOR}(\text{blaming}(t))$. The sentence 'murdering is wrong' then expresses the state of being for blaming murdering. In the following I will assume, Schroeder's expressivist has already earned the right to say that the state expressed by the sentences of his language express beliefs. So as a quasi-realist, Schroeder's expressivist will hold that to believe what a sentence says is to be in the state expressed by it. So he will hold that to believe that murder is wrong is to be for blaming murder, and likewise for all other beliefs.

Another important stipulation of Schroeder's theory, which we have also already come across in chapter 3, is that being for is capable of what he calls *A-type inconsistency*: A set of mental states is A-type inconsistent iff the states are all of the same genus (e. g. belief, intention, or being for...) and the set of states is such that being in those states is rationally incoherent if their contents are (logically) inconsistent. It follows, that a set of states of being for is rationally incoherent if their contents are (logically)

---

and not with his objection. Baker's argument, unlike his objection, does not require that 'makes sense' is a predicate.

[3]Standard semantic theorizing treats 'ought' as a deontic modal understood as a sentential operator. Baker cannot treat either 'makes sense' or 'ought' as a modal, however, because Schroeder's semantics does not contain a semantic clause for deontic modals, and so Baker cannot use such a clause to construct the mental state expressed by (S).

[4]Cf. (M. Schroeder, 2008a, 78). As before, I follow Schroeder in using small caps to denote mental states. For simplicity I also ignore that only closed sentences express states, and I ignore Schroeder's use of λ-abstractions in the metalanguage.

inconsistent. This implies that FOR(α) is rationally incoherent with FOR($\neg\alpha$), where α is a schematic letter standing for an arbitrary metalanguage formula describing a property one is for having. It is natural to reformulate A-type inconsistency as a requirement of rationality: rationality requires that one is not being for (logically) inconsistent contents. It follows from my definition of rational incoherence as violation of a rationality requirement that being for inconsistent contents is *one* way of being rationally incoherent, but it need not be the only way.[5]

According to Baker, Schroeder's semantic clause for atomic sentences together with his clause for conjunction (see chapter 3 above) determines that (S) expresses the following state:

(U) FOR$\big($blaming(murder) and $\psi(\neg$blaming(murder))$\big)$.

---

[5]An anonymous reviewer of Erkenntnis objects that I do not make clear what should count as rational incoherence. This is incorrect. Above I said that I treat any violation of a rational requirement as giving rise to rational incoherence. Since I think Schroeder's A-type inconsistency, as well as what he calls B-type inconsistency (for these notions see (M. Schroeder, 2008a, 43,48)), are classifications of types of requirements of rationality, it follows from my definition of rational incoherence that any way of being A- or B-type inconsistent is a way of being rationally incoherent. For instance, if it is a requirement of rationality that one is not being *for* something while also being *against* it (which would be a form of B-type inconsistency), then violating this requirement makes one rationally incoherent. The same holds for any other requirements of rationality.

The same reviewer criticizes that I do not mention that Schroeder and Baker (in joint work with Woods, (Baker & Woods, 2015)), disagree about how expressivists can and should explain logical relations such as logical inconsistency between sentences in terms of the rational relations between the states they express. In my view, all sides agree that expressivists want to be able to say that a set of sentences is logically inconsistent iff the rational incoherence between the states expressed by them survives substitutions. But Schroeder believes, whereas Baker and Woods do not, that expressivists can guarantee that rational incoherence survives substitutions in a *constructive* and non-*ad hoc* way, only if they appeal to forms of rational incoherence which are due to relations between the *contents* of states (A-type inconsistency), and not due to relations between *types* of states (B-type), because he thinks B-type accounts cannot be generalized. (In (M. Schroeder, 2015b), however, Schroeder argues expressivists can have a constructive higher-order account without requiring A-type inconsistency. For Schroeder's notion of constructiveness see for instance his (M. Schroeder, 2010b, 133) or the précis to a PEA-soup discussion (Paakkunainen, 2015)). It is not totally clear whether Baker and Woods disagree with Schroeder's idea that all such explanations need to be constructive, or if they disagree with the idea that constructiveness can be achieved only by appeal to A-type inconsistency. But whatever their disagreement ultimately consists in, I must admit that I do not see why this debate should have any bearing on the issue being explored in Baker's paper to which this chapter offers a response. Whether or not expressivists can or should explain logical inconsistency by appeal to specific forms of rational incoherence, such as A-type inconsistency, does not tells us which further requirements of rationality expressivists should accept, which is the topic of this and Baker's paper. In particular it does not imply that they should accept the principles which Baker needs in order to make his argument against Schroeder (e. g. Forcia or Direct Enkrasia below). The debate about how expressivists can account for logic, even though they do so by appealing to rational incoherence, is orthogonal to the issue discussed here.

Baker here uses ψ as a placeholder because Schroeder's semantics does not include a lexical entry for the predicate 'makes sense'.

Now, if Baker's claim is correct, which says that the state that is expressed by (S) must be an incoherent or inconsistent state of mind, and Schroeder's theory says that (U) is the state expressed by (S), then (U) must be an inconsistent state of mind. This raises the question: "does Schroeder's theory predict an inconsistent state of mind?" (Baker, 2018, 838).

The answer to this will obviously depend on what ψ is, and the largest part of Baker's paper is concerned with the question of whether 'ψ' can be replaced with something that allows Schroeder to predict the rational incoherence of (U). In fact, Baker thinks there *are* ways to predict the incoherence, for instance, if 'ψ' is replaced with 'being for'. But he argues that every replacement that allows Schroeder to predict the incoherence creates problems elsewhere in the expressivist theory (e. g. having to do with the inexpressibility of some states). So his objection is that, *ultimately*, expressivists are unable to explain the incoherence.

I think that expressivists could say some things in reply to Baker's various replacements of 'ψ', and the problems those replacements cause. But it is unnecessary that I go into the downstream details of Baker's objection here, since I want to suggest that Baker's claim about incoherence is false anyway. I do so by showing that Baker's argument for the claim is *unsound* (Sect. 6.4), and that better arguments for it can and should be resisted by Schroeder's expressivist (Sect. 6.5). This undermines Baker's objection to expressivism, because if there is no incoherence to be explained in the first place, then it cannot be objected that expressivism is unable to explain the incoherence.

6.3   Baker's Argument for Incoherence

Why does Baker think that the sentence 'murder is wrong, but not blaming for murder makes sense' must express a rationally incoherent mental state? He offers the following argument:

> the agent who utters the above sentence [(S)] is expressing a *commitment* [...] to blame for murder, while simultaneously the same agent judges it makes sense not to blame for murder—and since 'makes sense' is normative, she must also be expressing a commitment *not* to blame for murder. So it seems like we should want to say that [a person who utters] 'Murder is wrong but not blaming for murder makes sense' is semantically incompetent or she is irrational—she's committing to doing something and simultaneously committing to not doing it. (Baker, 2018, 837, italics original)

Here is another version of the argument:

> We can put the point this way: either the agent does not blame
> for murder, in which case she fails to have an attitude she is
> committed to having; or she blames, in which case her emotions
> don't conform with her own better judgment, her judgment about
> which reaction would make the most sense. So she is guilty of
> inconsistency in her psychological commitments. (Baker, 2018,
> 838)

This argument relies on two principles. The first is:

> it seems like an implicit commitment about the nature of the
> being-for attitude is that failing to have the target attitudes that
> one is for having is a failing of rational coherence. Presumably
> it was always part of the story that an agent who is for blaming
> for murder but does not in fact blame for murder is guilty of
> irrationality (all else being equal). (Baker, 2018, 843)

I shall reformulate this as a requirement of rationality, and call it:

Forcia: Rationality requires: if one is FOR(α), then one is α.

The argument also relies on a second principle which is:

> [There] is the widespread intuition that an agent who sincerely
> judges that something is the thing to do will be motivated to do
> it, barring weakness of will or other forms of defective agency.
> Someone who judges that some feeling is the one that makes
> sense suffers from a case of irrationality (of recalcitrant emotion)
> if they do not in fact feel it. (Baker, 2018, 837)

Applied to his second conjunct he says:

> If I judge 'Not blaming for murder makes sense,' what this
> judgment should intuitively regulate is my attitude of blame (or
> my lack of such an attitude). If I continue to blame for murder,
> I am acting, by my lights, irrationally. ... [that judgment] does
> rule out the rational coherence of continuing to blame for murder.
> (Baker, 2018, 839)

We can also reformulate this as a requirement of rationality, and call it,
following Broome (Broome, 2013):

Direct Enkrasia: Rationality requires that if one believes one ought to $\neg\alpha$,
then one is $\neg\alpha$.

Two points on my reformulation of Baker's argument. First, Direct Enkrasia is formulated in terms of 'ought', instead of 'makes sense' or 'is the thing to do', because this is the way Enkrasia (see below) is standardly formulated. This should be fine for Baker, since he explicitly says that he "treat[s] 'One ought to x' as equivalent to 'x-ing makes sense' or 'x-ing is reasonable' " (Baker, 2018, 831). Second, in the quote above Baker says that someone who jugdes that something is the thing to do will be *motivated* to do it. My reformulation does not mention motivation because, if instead of Direct Enkrasia we interpret Baker as saying that it is rationally required that if one believes one ought to $\neg\alpha$, then one is *motivated* to $\neg\alpha$, then my reformulation of Baker's argument would become invalid, and I do not what to ascribe Baker an invalid argument.

So here is Baker's argument in a more explicit form:

Forcia: Rationality requires that if one is FOR(α), then one is α.

Direct Enkrasia: Rationality requires that if one believes one ought to $\neg\alpha$, then one is $\neg\alpha$.

Corollary: one necessarily violates a requirement of rationality (i. e. one is rationally incoherent) if one is for α and believes one ought to $\neg\alpha$.

Belief as Being For: believing that t is wrong *is* being for blaming for t.

Baker's Claim: Hence, one necessarily violates some requirement of rationality if one believes that t is wrong and believes that one ought not blame for t. (From Corollary, Belief as Being For, and substitution of α with 'blaming for t')

This argument is valid. It is impossible to satisfy both Forcia and Direct Enkrasia, if one believes one ought to $\neg\alpha$ while also being for α, which means that one cannot be in both states and be fully rationally coherent. By replacing α with 'blaming for t', we get that it is rationally incoherent to believe one ought not to blame for t while also being for blaming for t. Since, according to Schroeder's noncognitivist "analysis of belief" (M. Schroeder, 2008a, 97) (i. e. Belief as Being For), being for blaming for t is believing that t is wrong, it follows that the belief that t is wrong is rationally incoherent with the belief that one ought not blame for t. Hence, Baker's argument entails his claim about incoherence.

But here is the problem. Even though Baker's argument is valid, it is not sound. Contrary to what Baker assumes, Schroeder's expressivist *does* not, *need* not, and *must* not accept either Forcia or Direct Enkrasia. I shall argue for each of these claims in turn.

Schroeder's expressivist does not assume Forcia or Direct Enkrasia, at least not explicitly. Schroeder nowhere in his book or his papers[6] assumes "that an agent who is for blaming for murder but does not in fact blame for murder is guilty of irrationality (all else being equal)" (Baker, 2018, 843). What Schroeder in fact does assume is "that being for has the motivational property that someone who is FOR(α) will tend to do α, all else being equal" (M. Schroeder, 2008a, 93). But this is neither formulated as nor intended to be a requirement of rationality.[7] Since Schroeder's expressivist does not assume Forcia or Direct Enkrasia, he will also not accept Baker's argument.[8]

Schroeder also does not *implicitly* assume Forcia or Direct Enkrasia because his theory does not *need* these principles. Neither of them plays a theoretical role in his theory. The only principle Schroeder actually needs to establish his results, especially about logic, is A-type inconsistency.

Perhaps Schroeder would have to assume additional requirements besides A-type inconsistency, if he wanted to explain not only logic, but also how it is possible that certain requirements of rationality hold, if normative or moral beliefs really are noncognitive states of being for. But whatever additional requirements expressivists need in order to explain some widely accepted requirements, they *must not* accept either Forcia or Direct Enkrasia, since they are false, at least this is what I shall argue now.

Direct Enkrasia is false. Rationality does not require that rational agents *succeed* in doing what they believe they ought to do. Rational agents may be omniscient, but they need not be omnipotent. That one sometimes fails to do what one believes one ought to do, does not make one irrational, because it is sometimes, in fact often, beyond our control to secure that we succeed in doing what we believe we ought to do. This line of argument is rooted in the widely accepted principle that rationality supervenes on the mind. This principle implies that rationality is a relation between mental states, and not between mental states and facts, events or actions.[9]

Direct Enkrasia will be compatible with the supervenience principle, and may in fact sound more plausible, if it is restricted to α's that are not actions or states of the world, but mental states.[10] It has, however, been

---

[6]A collection of some of Schroeder's papers on expressivism can be found in his volume (M. Schroeder, 2015a).

[7]Even if it were formulated as a requirement, it would require only that one *tends* to do α, not that one does α, which is what Baker's argument needs.

[8]As a reviewer of Erkenntnis correctly points out, Schroeder's expressivist might be committed to Baker's claim even if he does not assume Forcia, and even if Baker's argument is unsound. It is for precisely this reason that I discuss and reject other possible arguments for the incoherence below.

[9]See for instance (Broome, 2013, 89), (Kiesewetter, 2017), (Wedgwood, 2017).

[10]Baker's talk of 'emotions', 'feelings' and 'attitudes' in the above quotes clearly suggests that he implicitly restricts Forcia to mental states. But it is not clear that in the case of being for blaming for murder, blaming is a mental state, or that it could not be replaced by actions like sanctioning, punishing, or outlawing.

convincingly argued by John Broome that "we have to reject the Direct Enkratic Condition, even restriced to mental states. Its apparent attraction is spurious, and the counterexamples show it is false" (Broome, 2013, 96).[11]

Similar points can be made about Forcia. If Forcia is not restricted to mental states, then it violates the supervenience constraint for rational requirements. One cannot be accused of irrationality just because one fails to succeed to do what one is for doing. Bad luck is not the same as bad reason. But even if Forcia is restricted to mental states, it seems false. Teemu Toppinen, to who's paper Baker refers in a footnote, argues against 'Schroeder-style views' that there is "*no appropriate rational connection*, at all, between desiring to desire φ and the desire to φ", and likewise that "it is possible for me to be fully rational, to be for desiring φ, and yet not to desire to φ" (Toppinen, 2015, 160f., italics original). Moreover, it seems that Baker himself should be skeptical about Forcia, given that he says about desire that "it is an open question whether there is any breakdown in rational coherence in failing to have an attitude one desires to have" (Baker, 2018, 840). I do not see why this should be any different in the case of being for. Hence, Forcia, just as Direct Enkrasia, is not plausibly a requirement of rationality.

To summarize: Schroeder's expressivism does not assume Forcia or Direct Enkrasia. Moreover, Schroeder does not need to assume these principles in order to establish the results of his expressivist semantics and logic. And more importantly, these principles must not play a role in his theory, since they do not appear to be requirements of rationality at all. So I conclude that, since either of the crucial premises of Baker's argument is false, the argument does not establish the truth of his claim about incoherence.

6.4 Better Arguments for Incoherence?

In the previous section I argued that Baker's argument for incoherence is unsound. But showing that an argument is unsound does not show that its conclusion is false. So might Baker's claim nevertheless be true? In this section I offer some better arguments for Baker's claim of incoherence, but argue that even these better arguments can and should be resisted by expressivists. Here is a better argument:

No Contradictory Intentions: Rationality requires that if one intends α, then one does not intend $\neg\alpha$.

Enkrasia: Rationality requires that if one believes one ought $\neg\alpha$, then one intends $\neg\alpha$.

---

[11]For discussion of (Restricted) Direct Enkrasia and the counterexamples see (Broome, 2013, Ch. 6) who concludes that "Versions of Direct Enkrasia have an initial appeal, but they are no real alternative to Enkrasia itself" (Broome, 2013, 173).

Corollary: One necessarily violates a requirement, if one intends α while believing one ought $\neg\alpha$.

Belief as Intention: Believing that t is wrong is *intending* to blame for t.[12]

Baker's Claim: it is necessarily rationally incoherent to believe that t is wrong while believing that one ought not blame for t.

This is valid. No Contradictory Intentions and Enkrasia cannot jointly be satisfied if their antecedents are true, because their consequents are inconsistent (Corollary). What we need in order to derive Baker's claim from this is the additional noncognitivist premise which says that to hold a moral belief is to *intend* something (Belief as Intention), instead of being for something. This premise is Schroeder's Belief as Being For from above, simply with 'being for' being replaced by 'intention'. Baker's claim then follows straightforwardly from Corollary, Belief as Intention, and by substituting 'blame for t' into 'α'.

This argument is better than Baker's original one because it relies on the widely accepted requirements No Contradictory Intentions and Enkrasia.[13] I will not debate their truth here, and simply assume, for the sake of argument, that expressivists accept them.[14] But even if Schroeder's expressivist accepts these as requirements, the argument has an obvious flaw: Schroeder does not accept Belief as Intention, but Belief as Being For. So, this argument does not apply to Schroeder's expressivism, at least as it is currently formulated.

So is Schroeder's expressivism safe from this new argument? One might think that there are two problems, both deriving from the fact that being for and intention share some of their main features.

Here is the first problem. What is important about the being for attitude is that it is a noncognitive, desire-like pro-attitude that is rationally incoherent to hold towards inconsistent contents. But these things seem to be true of intention as well, and Schroeder has not told us enough about the being for attitude to clearly see important differences with intention. This suggests that Schroeder could perhaps have formulated his theory in terms of intention, instead of being for. If he did, he would accept Belief as Intention. But then the better argument would apply to his theory and commit him to Baker's claim about incoherence which, Baker objects, he cannot explain.

Are there any reasons that speak against replacing being for with intention? I think the fact that identifying moral belief with intention allows one

---

[12]A reviewer rightly emphasizes that just like the attitude of being for is only part of Schroeder's toy expressivist theory, the relation of blaming for is also just part of the toy example, and not essential to his theory. That is correct. But since nothing in the better argument hinges on the blaming for-relation we can safely ignore this.

[13]That they are widely accepted of course does not mean they are uncontested. Famously, Kolodny (2005) denies that there are *any* requirements of rationality.

[14]It is an underexplored question which, if any, requirements of rationality expressivists accept or are committed to accepting.

(via the better argument) to derive Baker's claim (which is intuitively false) is itself a reason not to identify moral belief with intention.

Another reason against the identification with intention is that if the belief that t is wrong were the *intention* to blame for t, then the belief that one ought to blame for t would require (via Enkrasia) to believe that t is wrong. This does not sound correct to me in general. It seems rationally fine to believe that blaming for t is the thing one ought to do even though one does not believe that t is morally wrong.[15]

Another consideration that speaks against replacing being for with intention is the following. Many think that moral belief provides motivation that is *pro tanto*. This means it can be overridden or outweighed by stronger motivation, perhaps provided by some other moral belief, or a belief about what one *prudentially* ought or ought *all things considered* to do. Plausibly being for just like desire, also provides *pro tanto* motivation. In contrast to desire and being for the motivation that is provided by intentions seems rather *pro toto*. The motivational force of an intention seems to be the result of weighing the motivational forces of competing desires or other *pro-tanto-*motivation-providing attitudes. This is also supported by the observation that if one intention is stronger than some other incompatible one, then we will normally drop the weaker one completely, which is less frequently the case with desires: we often stubbornly stick to our desires even if they are incompatible and of very different strengths. All this suggests that moral belief is more desire-like, and hence rather like being for, than like intention.

These considerations suggest that the Schroederian expressivist should not accept *Belief as Intention*.[16] In any case, Schroeder can escape the better argument by simply sticking to his choice of letting being for be the central attitude in his theory.

This brings us to the second problem mentioned above. If being for and intention are similar in important respects, then perhaps being for is also governed by requirements that are structurally similar to those governing intention. Suppose, for instance, that the following variation of Enkrasia governs the being for attitude:

Forkrasia: Rationality requires that if one believes one ought $\neg\alpha$, then one is *being for* $\neg\alpha$.

---

[15]Maybe Gibbard would disagree with me here. He proposes to analyse what is morally good in terms of which feelings one rationally ought to have. So perhaps he would accept that believing one ought to blame for t rationally requires believing that t is wrong.

[16]Another reason against identifying believing that something is wrong with intending to blame for it is that it is not clear that it is even possible to intend to blame, if we think of blaming as a certain sort of attitude or mental state. The impression that our attitudes, like our beliefs, are not under our control in the same way as our actions are, is shared by a number of philosophers such as, for instance, (Broome, 2013, 220f.), (Hieronymi, 2006), (Wedgwood, 2017). I thank an anonymous reviewer of Erkenntnis for raising this point.

This principle together Schroeder's Belief as Being For and with the following principle (implied by A-type inconsistency)

No Contradictory Being Fors: Rationality requires that if one is being for α, then one is not being for $\neg\alpha$.

again entails Baker's claim.

This version of the better argument has two premises that are actually accepted by Schroeder, namely Belief as Being For and No Contradictory Being Fors. However, Forkrasia is not part of Schroeder's theory, at least not an explicit part, and I do not think anyone has defended this particular principle.

But is Forkrasia implicitly accepted by Schroeder? Perhaps this is so. In a short paragraph of his book Schroeder investigates the idea of explaining the motivational character and inconsistency-transmittingness of intentions on the basis of the motivational character and inconsistency-transmittingness of the attitude of being for.[17] In this context he assumes, though quite hypothetically, that "intending to do A ... involves or commits to being for doing A" (M. Schroeder, 2008a, 101). Call this assumption: *Intention-For*. This bridge-principle, together with Enkrasia would entail Forkrasia. But this means that Intention-For, Enkrasia, Belief as Being For, and A-type inconsistency would also jointly entail Baker's claim. Should expressivists accept this argument?

Since Schroeder accepts Belief as Being For, and A-type inconsistency, this argument crucially rests on Enkrasia and Intention-For. I have already noted that it is unclear if Schroeder's expressivist accepts Enkrasia or not, but I assumed, for the sake of argument, that they accept it.

So this novel argument hinges completely on Intention-For. So what about Intention-For, the view that intention involves or commits to being for? Interestingly, there is a very similar thesis accepted by many philosophers of mind, namely the thesis that intention involves *desire*. For instance, Kieran Setiya says that it

> is a matter of consensus in the philosophy of intention that intending to do A entails wanting to do A, in the motivational sense for which the 'primitive sign of wanting is trying to get' (Anscombe, 1963, 68). Doubts about this entailment are attributed to ambiguities in 'desire.' When I intend to do A reluctantly, from the motive of duty, I may deny that I want to do it, but what I lack is 'appetite' not 'volition'. (Setiya, 2018)

---

[17] An attitude, according to Schroeder, is said to be 'inconsistency transmitting' if pairs of this attitude are rationality incoherent just in case their contents are (logically) inconsistent.

Most authors understand the notions of 'wanting' or 'desire' here rather narrowly, as they figure, for instance, in discussions of the so-called Humean theory of motivation.[18] The entailment thesis is mostly taken to be intuitively true or defended indirectly by showing that apparent counterexamples to it fail. The only direct argument for the entailment I know of is this. If a person has or expresses an intention that she is going to do something, it is always appropriate or legitimate to ask that person "Why do you *want* to do that?"[19] It seems that the question would not always be appropriate, if what it presupposes, namely that one wants something, would not be made true by that person's having an intention. This suggests that it is impossible to intend something without wanting or desiring it. This is supposed to justify the view that intention entails desire.

In the present context the question is: Does intention also entail being for? Answering this is difficult because we would have to know more about the being for attitude in order to answer it. For instance, is it a decidedly moral attitude or not? If we think of it as being some *sui generis* attitude of moral approval, then the entailment does not seem to hold: intending to go swimming does not imply that one morally approves of swimming.[20] If, on the other hand, being for is not a decidedly moral attitude, then it could just be the attitude that goes along with our ordinary notion of being for something. After all, 'to be for something' or 'to be in favor of something' are familiar English expressions that contrast with 'being against' (see the OED), and it is natural to think of voting as a way of indicating what we are for.

But this ordinary notion of being for something also does not support Intention-For. Consider the following case. Former Prime Minister Theresa May intended to implement Brexit.[21] But even if, in some sense of 'want', she wanted or was motivated to implement Brexit (because doing so was her

---

[18] For an overview on desire see (T. Schroeder, 2020). For a list of properties of desires see (Audi, 1973), or more recently, (Sinhababu, 2017). However, some authors in the debate have a very broad conception of the wants or desires that are entailed by intentions so that various kinds of noncognitive or motivational states count as desires (see (Hyman, 2014, 84), (Tuomela, 1977, 128)). These authors seem to take the true core of the principle that intention involves desire to be the following: "someone intends to do something only if he is *motivated* to do it" (Davis, 1984, 46). But this principle is much weaker than the claim that intention entails the particular mental state of desire in the narrow sense.

[19] See (Davis, 1984), (Thompson, 2008, 104).

[20] I fully agree with a reviewer that the attitude of being for, as Schroeder thinks of it, cannot be a *sui generis* attitude of moral approval. If it were, then the belief that grass is green (which on Schroeder's proposed analysis is the state of being for proceeding as if grass is green) would also have to count as a moral belief, which is absurd. It might be replied that what turns being for blaming for murder into a *moral* belief is not the being for attitude, but the relation of (morally) blaming for something, or it is the combination of being for with blaming. Compare what I above said in the context of the moral attitude problem.

[21] I am of course bracketing conspiratorial doubts about whether the real-world Mrs May really intended or only pretended to intend to implement Brexit.

political duty or mandate), it appears incorrect to describe her as *being for* Brexit. Famously, she voted Remain, and so she was for Britain's remaining in the EU, not for leaving it. Moreover, if Mrs May had on some occasion declared that she intends to implement Brexit (perhaps by saying 'I shall implement Brexit'), it may have been appropriate to ask why she wants it (she wanted it because she promised to deliver on the 'will' of the British people), but it would have been inappropriate to ask her why she is *for* Brexit, since everyone knows she was not for it, but against it, and perhaps she would even have voted Remain again, if there had been a second referendum.[22]

Cases like this one cast doubt on the thesis that intention entails being for. For this reason, Schroeder's expressivist should not accept it. I conclude that even the better arguments for Baker's incoherence claim fail. Thereby, Baker's objection to Schroeder's expressivism is undermined.

6.5 Conclusion

The main aim of this chapter has been to argue against Baker's claim that it is rationally incoherent to believe that murdering is wrong while believing that one ought not blame for murdering. I have done so by arguing that Baker's original argument for the incoherence rests on false premises, and is thus unsound. I showed that there is a better argument available for the incoherence of these beliefs which, however, does not apply to Schroeder's expressivism as the view is currently formulated, and there are reasons for Schroeder not to reformulate his view in terms of intention or to assume that intention entails being for. I take my arguments to provide inductive evidence for the falsity of Baker's claim. Consequently, if Baker's claim about incoherence is false, he cannot object that expressivism is unable to explain the incoherence.

I have argued that Baker's objection is unsuccessful against Schroeder's specific form of noncognitivism. What about my structural proposal, and the dispositionalist interpretation of it in particular? I think that it is highly unlikely that if Baker's objection does not apply to Schroeder's noncognitivism, it will apply to an even more complex version of it, namely my dispositional noncognitivism. The reasons I have given against Baker seem to work especially if the states are assumed to be more complex, in

---

[22]The same reviewer who agrees with me that being for is most certainly not a *sui generis* moral attitude of approval, also thinks that Schroeder's attitude of being for is not the attitude that goes along with our ordinary notion of being for something. As a reason for this the reviewer cites that Schroeder stipulated that his being for attitude can only take (mental) properties as objects, whereas we can ordinarily be for taking certain ways of acting. If that is true, then perhaps the Brexit example, and with it all appeals to the ordinary meaning of 'being for something', will not be able to show that intention does *not* involve being for. Anyways, it also does not show that intention *involves* being for, which is the claim a defender of the better argument would have to make plausible.

fact, there the reasons are even more convincing. For instance I argued that Forcia is not a requirement of rationality, that is rationality does not require that if you are being for blaming for murder, then you are blaming murder. For the same reasons that this is not a requirement, it also does not seem to be a requirement that if you are disposed to be for blaming for murder, you are blaming for murder. This, however, would be needed in order to get Baker's objection going.

There are better objections against noncognitivism having to do with the rationality of beliefs than Baker's. For this reason I prefer not to investigate in detail whether there might be a better Bakerian objection against dispositional noncognitivism. The burden of proof is shifted to Baker.

In the next chapter, I turn to what is thought to be one of the most difficult problems for noncognitivism in epistemology, namely Cian Dorr's wishful thinking objection.

Chapter 7

Expressivism and Reasoning

Metaethical expressivists cannot really claim to have earned the right to moral belief unless they can solve what is considered to be one of the hardest problems, if not the hardest problem, for noncognitivism in epistemology: Cian Dorr's *wishful thinking* problem. In this chapter, I offer a solution to this problem.

7.1   Introduction

Cian Dorr in his paper on *Non-Cognitivism and Wishful Thinking* (Dorr, 2002) argues that if noncognitivism is true, then forming a descriptive or factual belief on the basis of moral belief can never be epistemically justified because believing something on the basis of a desire (more generally: noncognitive state) can never be justified. Doing so is wishful thinking. However, intuitively, reasoning from moral beliefs to factual beliefs can be justified. Hence, Dorr concludes, noncognitivism must be false.

Dorr's argument rest on the seemingly plausible assumption that believing something on the basis of a desire can never be justified. I will argue that this is false. Believing something on the basis of a desire can be justified. In fact it can be epistemically, not just pragmatically, justified. I agree that paradigmatic cases of wishful thinking never epistemically justify a belief. But the case which Dorr considers is crucially different from paradigmatic wishful thinking, even if we assume noncognitivism. The most important difference is that moral thinking can be rational, whereas wishful thinking cannot.

I argue that even if moral beliefs are noncognitive states, moral thinking is not wishful thinking. Moral thinking can give one reasons for beliefs whereas wishful thinking cannot. More precisely, I argue that a piece of reasoning gives one a (right kind of) reason for (or against) a (factual) belief if (and only if) (i) the reasoning is rational, (ii) one holds the premise states, and (iii) one holds the premise states for (the right kinds of) reasons. I defend

this principle and show that it is possible for noncognitivists to satisfy all three conditions, especially condition (iii), and so rational reasoning involving moral beliefs can give one the right kind of reasons, that is truth-related or epistemic reasons for factual beliefs. Hence, noncognitivists are not wishful thinkers.

7.2   Dorr on Wishful Thinking

In a famous paper (Dorr, 2002), Cian Dorr argues that noncognitivists must hold that accepting[1] the premises of the following argument cannot give someone reason to accept the conclusion:

P1 If lying is wrong, the souls of liars will be punished in the afterlife.

P2 Lying is wrong.

C Hence, the souls of liars will be punished in the afterlife.

Dorr, however, takes it to be common-sensical that accepting P1 and P2 can give someone reason to come to accept C. Therefore, noncognitivism must be false.

   In order to make his argument plausible, Dorr considers the case of Edgar which he describes as follows:

> It seems quite easy to tell a story in which acceptance of P1 and P2 gives someone reason to come to accept C. Thus: Edgar confidently accepts P1, having been taught to do so by teachers whom he takes as authorities on this kind of matter. Since these teachers have not inculcated any views in him as regards the morality of lying, he is inclined to think that lying is perfectly all right provided you get away with it. So he does not accept P2. As regards C he is undecided, and has good reason to be. No-one whose opinion he trusts has told him anything one way or the other; nor does he have evidence of his own that bears on the specific nature of the afterlife. Later, after reading a philosophy book, he is led to reflect on his moral commitments. The immediate result of this process is just this: he comes to accept P2. Holding fast to his acceptance of P1, he does as coherence demands, and revises his expectations about the nature

---

[1]Wherever possible I follow Dorr's use of speaking of acceptances of sentences. To accept a sentence is understood here as a term of art for being in whatever state that is expressed by that sentence. For instance, cognitivists will say that to accept P2 is to be in the cognitive state of believing that lying is wrong, whereas noncognitivists will say that accepting P2 is to be in some noncognitive state such as disapproving of lying, or in the case of dispositional noncognitivism being disposed to be for blaming for lying.

of the afterlife. He takes P1 and P2 as his reasons for initially coming to accept, and continuing thereafter to accept, C. If he were later to reconsider or reverse either of those attitudes, he would thereupon cease to accept C. I say that Edgar could well have been rational throughout the story. Before he changed his mind about P2, it would have been unreasonable for Edgar to come to believe C. But that was relative to the rest of his beliefs about the world as they were at that time. Once his beliefs had changed so as to include the belief that lying was wrong, they could easily have become such as to support C over its negation. (Dorr, 2002, 98)

Edgar's case shows that some change in his epistemic situation can bring him from a situation in which he does not have reason to accept C to a situation in which he has a reason to accept C. Lets us refer to those two epistemic situations as e_1 and e_2 respectively. Some of the changes between e_1 to e_2 make the features of his epistemic situation sufficient to give Edgar a reason for accepting C.

For now what is important is that according to Dorr, there cannot be a change in one's reasons for believing C unless at least one of two other changes has occured. He embraces the following principle:

Only a change in one's cognitive states, or in one's evidence, can make a difference between a case in which it would be irrational [i. e. one lacks reason] to believe something and one in which it would be rational [i.e. one has reason] to do so. (Dorr, 2002, 99)

Now, here is the problem:

According to the non-cognitivist *all* that happened when he [i.e. Edgar] came to accept P2 was a change in his non-cognitive attitudes. (Dorr, 2002, 99, italics added)

If this is true, then it follows (via modus tollens from Dorr's change-principle) that there can be no difference concerning the reasons Edgar has for accepting C between e_1 and e_2. If the *only* change that happened was that Edgar adopts a noncognitive attitude, and Edgar does not have reason to accept C in e_1, then it follows that in e_2 Edgar still lacks reason to accept C. Therefore, Dorr concludes, if noncognitivism is true, then Edgar's accepting C in e_2 would be just as unjustified as it was in e_1. Since, however, in e_2 Edgar intuitively does have a reason to accept C, it follows that noncognitivism must be false.

7.3 Moral Thinking is Not Wishful Thinking

In the following sections I shall argue that Dorr's objection can be blocked and that even if moral thoughts are like wishes as the noncognitivist claims, moral thinking is not wishful thinking.

7.3.1 Accepting P2 is Not the Only Change

Must noncognitivists accept Dorr's argument? I think, no.[2] Noncognitivists should not accept Dorr's claim that *all* that happened to Edgar was a change in his non-cognitive attitude, and they should also not accept Dorr's change principle. It is not clear why Dorr thinks they must accept those claims. It seems part of Edgar's case, as Dorr has set it up, that Edgar's accepting P2 is not the only relevant change. At least two other important changes occur as well. I will discuss those two changes now.

Second Change: Rationally Committed to C

One thing that changes in Edgar's epistemic situation from e_1 to e_2 is that he comes to accept P2. If noncognitivism is true, this means that he comes to hold a new noncognitive state. But this is not the only relevant change.

A second changes is this. Since Edgar already accepts P1 and then comes to accept P2, he thereby becomes rationally committed to accepting C. This is not the case in e_1, since there he does not accept P2. This raises the question: Can his mere becoming rationally committed to accept C give Edgar a reason for accepting C?

Some philosophers might say: yes. They might say that rationality is normative and being normative means providing reasons. They claim that rationality provides us with what has been called 'coherence-based' (Budolfson, 2011, 248) or 'attitudinal' (Broome, 2013, 75) reasons.[3] Philosophers who accept that there are such reasons will say that if rationality requires you to accept the conclusion of a valid argument if you accept its premises,

---

[2]I am not alone in thinking that Dorr's argument can be blocked. Dorr assumes that if noncognitivism is true, the change from not accepting P2 to accepting P2 involves only a change in one noncognitive states. James Lenman's and David Enoch's responses to Dorr challenge this assumption by claiming that even if by coming to accept P2 one comes to hold a noncognitive state, this does not show that this is not also accompanied by a change in ones cognitive states involved, such as for instance the higher-order belief that one now is in that noncognitive state. See (Enoch, 2003) and (Lenman, 2003b). They claim that the changes in Edgar's cognitive states that non-coincidentally accompany his change in noncognitive state when he comes to accept P2, allow the noncognitivist to claim that accepting P1 and P2 gives him reason to accept C. To me, this line of argumentation is not very convincing, and has been criticized at some detail by Schroeder, so I will not restate the problems for Lenman and Enoch here. Cf. (M. Schroeder, 2010a, Ch. 9).

[3]I am not claiming that Budolfson or Broome hold that rationality provides coherence-based or attitudinal reasons. They just offer those helpful terms.

then accepting the premises gives you at least some reason to accept the conclusion. This is so, it is sometimes said, due to the fact that everyone has reason to be rationally coherent. If this is true, then Edgar's accepting P1 and P2 gives him at least some (defeasable) reason to accept C.

It is, however, highly controversial that rationality by itself provides reasons. Niko Kolodny's classic paper 'Why be rational?' (Kolodny, 2005) denies that rationality provides reasons. Kolodny assumes that requirements of rationality are narrow-scope, that is they are of conditional form in which 'rationally required' does not govern the whole conditional, but only the consequent. By assuming that rationality is normative, in the sense of providing reasons, it straightforwardly follows that if in accepting the premises of a valid argument one becomes rationally required to accept the conclusion, this gives one reason to accept the conclusion. This, according to Kolodny and many other philosophers, is highly implausible because it amounts to bootstrapping reasons into existence out of nothing. Merely coming to accept the premises of a valid argument cannot generate a reason for a conclusion-belief, for this would make it much too easy to justify our beliefs. Kolodny's bootstrapping worry sounds very similar to Dorr's complaint that if noncognitivism is true, Edgar could get a reason for C by merely coming to hold a noncognitive state.

In fact, Dorr also seems to reject the view that rationality provides reasons for accepting the conclusions of valid arguments: "Not every valid argument is such that accepting its premises could give one reason to accept its conclusion", he says (Dorr, 2002, 98). So, Dorr also seems to hold that the reason for thinking that Edgar has reason to accept C cannot be that he believes that Edgar's accepting P1 and P2 rationally requires him to accept C, and rationality itself provides reasons to believe.

But let us ignore this for a moment and suppose for the sake of argument that Dorr really thinks that Edgar's reason for accepting C stems from the fact that due to his accepting the premises he becomes rationally committed to accept the conclusion and this is where the reason to accept C comes from.

The problem with this line of argument is that noncognitivists could take it as well. This is because Dorr grants the noncognitivists that "the combination of the attitudes expressed by P1 and P2 with any attitude incoherent with acceptance of C ... is an incoherent combination. We can put this by saying that P1 and P2 *imply* C" (Dorr, 2002, 97, italics original)[4]. So, if rationality gives one reason to accept what is implied by what one already accepts, then Edgar's accepting P1 and P2 could give him reason to accept C even if some of the states are noncognitive as the noncognitivist claims.[5]

---

[4]Dorr's last sentence shows that he is familiar with the noncognitivists' idea to explain logical relations, such as implication, entailment etc. in terms of rational incoherence. Cf. Chapter 3 above.

[5]Note that this line of argument would conflict with Dorr's change principle according

Hence, it seems right to hold that by accepting P2 (and P1) Edgar becomes rationally required to accept C by itself is insuffcient give Edgar a reason to accept C.

Third Change: Reason for Accepting P2

There is another important change that occurring in Edgar's epistemic situation from e_1 to e_2. Dorr's tells us that when Edgar comes to accept P2 this is the result of a newly occurring event. This event is his reading a philosophy book and reflecting on his moral commitments. Dorr tells us that the "immediate result of this process" is that Edgar comes to accept P2. Unfortunately, Dorr does not tell us *what* Edgar read in the book nor what exactly is involved in the process of 'reflecting on one's moral commitments'. But it seems that whatever Edgar read or what is involved in reflecting on one's moral commitments, doing those things not only brought him to accept P2, but provided him a reason for accepting P2. So another important change that seems to have occurred in Edgar's epistemic situation is that he has acquired are reason for accepting P2 which he previously lacked. Might this explain where the reason for accepting C comes from?

7.3.2 Reason-Transmission

In order to answer this, let us switch away from Edgar and Dorr's moral example, and consider a purely non-moral argument:

Reasoning for resurrection

P3 If God exists, then Jesus has risen from the dead.

P4 God exists.

C2 Jesus has risen from the dead.

Now imagine the case of Bertrand. Bertrand, let us assume, accepts P3 and he has reason to accept it: he has been taught to accept P3 by teachers whom he takes as authorities on this kind of matter. However, Bertrand is undecided with regard to P4, he has no reason to accept it, and he does not accept it. The same holds with regard to C2. However, one day, all of a sudden, Bertrand comes to accept P4. He does so for *no* reason—just out of the blue.

So what happens to Bertrand's epistemic situation concerning C2? Since Bertrand comes to accept P4, and accepting P4 is coming to hold a cognitive attitude, there is change in cognitive attitude. Hence Dorr's change principle

---

to which there either must have been a change in reason for the premises or a change in cognitive state.

is not violated. It seems, however, that in Bertrand's case, his accepting P3 and P4 does not give him reason to accept C2. Even though it is rationally incoherent for him not to accept C2, he is unjustified in accepting C2 merely on the basis of his accepting P3 and P4. So, Bertrand's case suggests that Dorr's change principle at best states a necessary condition, but not a sufficient condition for giving an agent a reason to accept the conclusion of some argument.

Now consider a slightly different case, the case of Anselm. Anselm like Bertrand accepts P3, and for the same reason. Moreover, like Bertrand, Anselm also comes to accept P4. However, unlike Bertrand, Anselm does so for a reason. Anselm's reason, let us suppose, is that he experienced a true miracle, namely that the Holy Spirit descended on him in bodily form like a dove, and a voice came from heaven telling him that Jesus is his son.[6] What is crucial about this case is that Anselm, unlike Bertrand, seems to have a reason for accepting C2.[7]

I take these considerations to show that we get reasons out of a piece of reasoning only if we put reasons in. More generally I take the foregoing consideration of cases to suggest the following principle:

Reason-Transmission A piece of reasoning gives an agent a reason for holding the conclusion-state if and only if

(i) the reasoning is rational,

(ii) the agent holds the premise-states, and

(iii) the agent holds the premise-states for some reasons.

I want to note here that a similar principle has been formulated in justification logic.[8] There it is generally assumed that if r_1 is a reason or justification for (if A, then B), and r_2 is a reason for A, then the so-called 'application' of the first and the second reasons is a reason for B. More formally:

$$r_1 : (A \rightarrow B) \rightarrow (r_2 : A \rightarrow [r_1 \cdot r_2] : B)$$

Here the ':' stands for 'is a reason for', and '$\cdot$' denotes the application operation on reasons. This has the interesting feature that even though the application of r_1 and r_2 *is* a reason for B (i.e. $[r_1 \cdot r_2] : B$), this reason is the result of the 'application' of reasons r_1 and r_2, but it is not an independent reason, r_3 say, for B, as, for instance, a new perceptual experience that B

---

[6]Cf. Luke 3:22, Matthew 3:17.

[7]Perhaps we also have to assume that Anselm does not have too strong a reason against any of the premises or against the conclusion. In what follows I take this assumption as being implicitly made.

[8]Cf. (Artemov & Fitting, 2021, Sec 1.1).

would be. This can be viewed as capturing the fact that it is the reasoning that partly constitutes the new reason and not a reason stemming from an external source.

I think that Dorr would deny that the above Reason Transmission-principle lists all the individually necessary and jointly sufficient conditions. His change-principle suggests that he would hold that in the case where the conclusion-state is a cognitive state, (iii) must be modified as follows:

(iii2) the agent holds the premise-states for some reasons *or* the premise-states are cognitive states.[9]

But are (i), (ii), and (iii2) jointly sufficient for giving an agent a reason for holding a cognitive conclusion-state?

I take Bertrand's case to show that the answer is: no. If (i) and (ii) are satisfied, and the premise-states are cognitive-states (which satisfies (iii2)), this is still not sufficient for giving him a reason for believing that Jesus has risen from the dead. Simply coming to hold the premise-states, but for no reason what so ever, even if they are all cognitive, still seems insufficient for giving an agent a reason for holding the conclusion-state of a piece of rational reasoning.[10]

Anselm's case, in contrast, suggests that in addition to that the premise-states must be held for a reason. Dorr might therefore suggest that we should accept the following stronger principle instead:

A piece of reasoning gives an agent a reason for holding a cognitive[11] conclusion-state if and only if

(i) the reasoning is rational,

(ii) the agent holds the premise-states,

(iii) the agent holds the premise-states for some reasons, *and*

(iv) the premise-states are cognitive states.

---

[9]More accurately I should say the *relevant* premise-states, since some piece of reasoning could contain a (noncognitive) state as premise that does not play any role in the reasoning, e. g. take the case where an agent believes that if p, then q, believes p, intends r, and on this basis infers the belief that q. In what follows I shall leave the qualification to relevant states tacit.

[10]This clearly also relates to what I have previously said about bootstrapping.

[11]Note that the above Reason-Transmission principle was not restricted to cognitive conclusion states, but was more general. The reason for the restriction here is that it is highly implausible that if we considered a piece of practical reasoning it would be necessary that the premise-states must all be cognitive states. So if the following principle should be at least somewhat plausible, it must be restricted to theoretical reasoning which ends in a cognitive conclusion state.

It should be noted that this principle entails, but is not entailed by, Dorr's claim that (iii2) is a necessary condition for a piece of reasoning to give an agent a reason for accepting the conclusion. So Dorr's reason for accepting (iii2) might be that he actually implicitly accepts this stronger principle.

If this stronger principle is true, noncognitivists seem to have a problem, since it is not clear how (iv) can be satisfied in cases in which a premise state is a moral belief. Noncognitivists might try to get around this problem by weakening condition (iv) in the following way:

(iv2) holding the premise-states is non-coincidentally accompanied by cognitive states.[12]

(iv2) is weaker than (iv) by requiring only that being in the premise states is accompanied by cognitive states while allowing that some premise states are themselves noncognitive states. This condition would still be compatible with Dorr's claim that there can be a change in the reasons for the conclusion-state only if there is either a change in the reasons for the premise-states or a change in the agents cognitive states.[13]

For noncognitivism to comply with this principle noncognitivists would have to show that even though coming to accept a premise involves coming to hold a noncognitive state, holding this state is non-coincidentally accompanied by cognitives states. In this case a change in noncognitive attitude also results in a change in the accompanied cognitives states. I will, however, not pursue this line here.[14]

Instead I suggest another line of defence. I propose that rather than weakening (iv) to (iv2), noncognitivists should reject (iv) altogether, and stick with the weaker Reason-Transmission principle above, that is (i), (ii), and (iii).

Suppose that Dorr accepts that the Reason-Transmission principle is correct. He might then still object that noncognitivism implies that Edgar's reasoning cannot give him a reason to accept C because noncognitivists are unable to satisfy (iii). In fact, part of the reason why Dorr thinks that

---

[12]Lenman and Enoch seem to accept something like this condition also. Cf. (Enoch, 2003) and (Lenman, 2003b).

[13]Note that even if holding the relevant noncognitive states would non-coincidentally be accompanied by cognitive states, a change in noncognitive state not necessarily also leads to a change in a cognitive state. This is so if an agent would already be in the cognitive state. Suppose, for instance, that the noncognitive state N is accompanied by the belief that one is in N. If it is possible to hold this latter belief without holding the noncognitive state, then if one comes to adopt the noncognitive state, there still need not be any change in cognitive attitude. This means that in order for (iv2) to be really compatible with Dorr's change principle some further refinement would be needed.

[14]This possible line of reply to Dorr's objection is pursued by David Enoch and James Lenman on behalf of noncognitivism, but with only moderate success. For a critical discussion of Enoch's and Lenman's replies to Dorr's problem consult (M. Schroeder, 2010b). I am not convinced by Enoch's and Lenman's proposals.

noncognitivists cannot explain how accepting P1 and P2 can give Edgar reason to accept C seems to be precisely that he implicitly assumes that if accepting P2 is a noncognitive state, it is impossible that to accept it for a reason.

But why should we think that noncognitivists must hold that one cannot have reasons for accepting P2? It seems clearly possible to have reasons for noncognitive states. For instance, the fact that hungry sharks are dangerous might be a reason for fearing to swim in front of them. Hence, there is room for noncognitivists to claim that what changes in Edgar's situation is not only that he adopts a noncognitive attitude, but he also receives a reason for accepting P2. If what changes in Edgar's situation is that he receives a reason for accepting P2, then the second disjunct of Dorr's original change principle will not be violated, which means that Dorr's modus tollens argument against noncognitivism is thereby blocked. Moreover, if it is possible to have reasons for noncognitive states, and thus for accepting P2, then by Reason-Transmission, it follows that even if noncognitivism is true, Edgar in e_2 has a reason for accepting C.

I think that Dorr would not accept this as a reply to his objection. According to him everyone "should agree that the state of accepting C is a belief", where by 'belief' he means 'cognitive state'. So, Dorr might object that even if it is possible to have a reason for accepting P2 when accepting P2 is a noncognitive state, and this might give Edgar some reason to accept C (via Reasons-Transmission), where this is understood as a cognitive state, it does not give Edgar an *epistemic* reason for believing C, but at best a practical reason for believing C. Dorr's talk of "evidence" clearly suggests that he has something like this in mind.

What is it for a reason for a belief to be an epistemic reason? Here I want to make a suggestion that might help the noncognitivists with solving the wishful thinking problem. My suggestion is that we embrace a slightly boarder understanding of 'epistemic reason' than might be standard, an understanding that will allow us to say that even non-cognitive states can be held for epistemic reasons. I suggest that we identify what it is to have an epistemic reason for some mental state with having what is called a 'right kind of reason' for that state, or short: RKR.[15]

It is widely agreed that there are reasons of the right kind and reasons of the wrong kind (short: WKR) for being in some mental state. Furthermore, it is generally assumed that if a reason is not a reason of the right kind for some attitude, then it is a reason of the wrong kind, and vice versa. I want to emphasize that the fact that something is a reason of the right kind

---

[15]There might be reservations to use the term 'epistemic' with respect to non-doxastic states. In this case I want to note that the issue might be purely verbal. Nothing in my arguments hinges on the term 'epistemic reason'. The only thing I need in order to make my argument is the distinction between reasons of the right kind and reasons of the wrong kind.

for some mental state need not mean that it is a good reason, and that something is a wrong kind of reason need not mean that it is a be bad reason or no reason at all for being in some state.[16] In other words, the distinction between right and wrong kinds of reason is assumed to be orthogonal to the distinction between good and bad reasons.[17] Moreover, the reasons of the right kind for beliefs are of another kind than the reasons of the right kind for other attitudes such as, for instance, intentions or other pro-attitudes. Thus, though there might be exceptions (as I shall argue below), a reason that is of the right kind for a belief need not, and in fact typically will not be a reason of the right kind for an intention and vice versa.

What makes a reason a reason of the right kind for holding some mental state? It is controversial if there is a general answer to this, and if there is one, what it is. It is less controversial what the right kind of reason for specific types of mental states, such as beliefs, intentions and other pro-attitudes, are. For instance, it is generally assumed that a right kind of reason for a belief is a reason that somehow concerns the truth (or falsity) of the belief. For this reason they are called *truth-related* reasons (short: TRR). A truth-related reason for the belief that p, it is also often said, is a reason that 'bears on' the question of whether or not p. There are also right kinds of reasons for other states besides belief. For instance, it is also widely assumed that the right kinds of reasons for intentions are *action-related* reasons (short: ARR) and that reasons of the right kind for other pro-attitudes, such as desire, admiration or disapproval, are *value-related* reasons. An action-related reason for some intention is a reason provided by facts that also provide reasons for actions, and a value-related reason for some pro-attitude is a reason provided by features of something that make it valuable. In other words, an action-related reason is a reason that bears on the question whether or not to *do* something, and a value-related reason is a reason that bears on the question of whether or not to value it.[18] I take the distinction between ARR and VRR not to be as sharp as the distinction between ARR and VRR on the one hand and TRR on the other. For this reason I will often mention ARR and VRR in one breath.

With these remarks in place, we might reformulate Dorr's central claim as follows:

> If noncognitivism is true, Edgar in e_2 does not have a right kind of (or epistemic) reason, that is a truth-related reason, for believing C.

---

[16]This point has also been made by (Jacobson, 2013).

[17]This is similar to what is assumed in discussions of moral worth. An action might be morally right, but still lack moral worth, and an action might be morally wrong, but have moral worth. For more on moral worth see for instance (Arpaly, 2002).

[18]I take this way of drawing the distinction from (Gertken & Kiesewetter, 2017), and from (McCormick, 2018) and (Jacobson, 2013).

The Reason-Transmission principle from above, is too weak to establish this claim, since it at most would allow us to establish that Edgar has a reason for accepting C, but it does not allow to establish the stronger claim that Edgar has a RKR for accepting C, and thus TRR for believing C.

What could noncognitivists say in reply to this stronger version of Dorr's wishful thinking objection?

7.3.3 WKR-Transmission

As I have just said, even if (i)-(iii) of the above Reason-Transmission are satisfied, noncognitivists cannot conclude that Edgar has a reason of the right kind for accepting the conclusion. What is needed is a principle for accepting conclusions for RKRs. What could such a principle look like?

In order to answer this, take again the reasoning for resurrection from above and consider the case of Pascal. Pascal like Anselm, accepts P3 and he has reason to accept it: he has been taught to accept P3 by teachers whom he takes as authorities on this kind of matter. Pascal also comes to accept P4, that God exists, and he like Anselm has a reason for accepting it. But Pascal's reason for accepting P4 differs from Anselm's (which, recall, was that he experienced a true miracle): Pascal accepts P4 because he is convinced that doing so is more beneficial than not doing so. We may assume that his famous wager-considerations provide him a good reason for accepting P4. Yet, many philosophers think that even if it is a good reason for accepting P4, it is the wrong kind of reason (WKR) for accepting P4.[19] It is, as it is often called, a 'practical' reason for belief, rather than a 'theoretical', 'evidential', or 'epistemic' one. The crucial question is what happens with regard to C2 in Pascal's case? It seems that even if Pascal's reasoning might give him a reason for C2, it does not give him a RKR, but a WKR at best. So in the case of theoretical reasoning, that is reasoning with beliefs only, if we put a WKR in, we will not get a RKR out, but only at best a WKR. Basing a factual belief on beliefs that are held for merely practical reasons, does not seem to give a RKR or epistemic reason for C.

This finding is not limited to theoretical reasoning. It is supported by practical reasoning as well, that is reasoning that involves beliefs as well as intentions. Consider the following:

Reasoning for crime

P5 Believing that robbing a pharmacy is a means of drinking the toxin.

P6 Intending to drink the toxin.

C3 Intending to rob a pharmacy.

---

[19]Some philosophers think it is primarily a good (and in fact right kind of) reason for *wanting* to accept P4, and thus only secondarily a good reasons for accepting P4.

Consider the case of Kavka. Kavka accepts P5,[20] and he does so for a reason, in fact for a reason of the right kind, let's assume. Kavka accepts P6 too, but his reason for doing so is of the wrong kind. His reason for intending P6 is that an evil demon threatens him with severe pain unless he intends to drink the toxin. This is a WKR because it is no fact that would make drinking it valuable (such as its being tasty). What happens with C3 in Kavka's case? Arguably, his believing P5 and his intending P6, and having reasons for both of them, gives him a reason for intending to robb a pharmacy. However, just as before in the case of Pascal, it seems that the reason this gives him for being in the conclusion-state, in this case for intending to robb a pharmacy, is a wrong kind of reason at best.[21]

This suggests that in the case of theoretical as well as practical reasoning if we put in a WKR, we get out at most a WKR. Hence, it seems that not only get reasons transmitted from premises to conclusion, but also the *kind* of reason. More precisely, if one of the relevant premises of a piece of reasoning is held for a wrong kind of reason, then the reason this gives us for the conclusion is also a wrong kind of reason at best.

7.3.4 RKR-Transmission

In the previous section I argued that even though reasons are transmitted from premise-states to conclusion-states in a way similar to what is assumed in justification logic, Reason-Transmission does not help noncognitivists to establish Dorr's stronger claim that Edgar in e_2 has a reason of the right kind for believing C, that is a TRR. The reason why it does not help is that even if Edgar has some reason for accepting P2, where this is a noncognitive state, does not guarantee that the reason his reasoning generates for accepting C is a RKR. In particular we have found that if the reason for which one holds one of the relevant premise-states is a WKR, the reason generated for

---

[20]I said above that I follow Dorr in speaking of acceptance of sentences wherever possible. We see here that this makes sense only for declarative sentence, since there is no declarative sentence that expresses an intention (in the sense of 'express' as developed in chapter 2). I here take the liberty to speak of 'accepting P6' while what I mean is that someone is in the state described (not expressed) by P6. Speaking so makes it easier to refer to the premise and conclusion states of pieces of reasoning, but makes visible if we are concerned with a piece of theoretical or practical reasoning.

[21]I must admit that intuitions are not so strong here. A paradigm right kind of reason for intending to robb a pharmacy would be an action-related reason for so intending, a reason speaking in favor of robbing a pharmacy. It is not clear that in Kavka's case he does not have an action-related reason for intending C3, and therefore a right kind of reason for it. If this were so, then it would show that at least in some cases being in the premise-states of a piece of rational reasoning can give one a right kind of reason for the conclusion-state even if some reason for the premise-state is not of the right kind. This could be of help for noncognitivists because if they could make sense only of wrong kinds of reasons for moral beliefs, it might be possible that this gives one a right kind of reason for the conclusion state. Whether this line of defence is plausible I shall, however, not investigate here.

accepting C is also at best a WKR. So we found that if we put WKR in, we get WKR out.[22]

But since, intuitively, it seems that Edgar in e_2 not only has some reason or WKR, but it seems that he has a RKR for accepting C, we need to know under which conditions reasoning provides a RKR for its conclusion. The discussion of Pascal's and Kavka's case have already shown that we do not get a RKR out unless we put in a RKR. So having RKR for the premise-states is necessary for getting a RKR for the conclusion-state.

The crucial question is: Is it also sufficient? An affirmative answer is strongly suggested by noting that if in Pascal's and Kavka's case we change only one thing, namely we replace the WKR with a RKR for the second premise, this is the only change necessary for also giving us a RKR for the relevant conclusion. For instance, Anselm's epistemic situation with regard to P3 and P4 is just like Pascal's except for one thing: Anselm believes in God's existence for a right kind of reason, whereas Pascal believes it for a wrong kind of reason. Pascal's weighing considerations do not bear on the question whether or not God exists (that is, they do not bear on its truth), but rather on the question of whether or not to believe that God exists, that is it bears on an action. We can assume that Anselm's reason is that he has a deductively valid proof of God's existence or if you wish that he has experienced a true miracle. In this case it seems that the reasoning for resurrection gives Anselm a right kind of reason for believing that Jesus has risen whereas it seems that Pascal only gets a wrong kind of reason for this belief.

Likewise in the case of practical reasoning. Take the case of Sokrates. Sokrates' situation is like Kavka's with regard to the reasoning for crime except for the fact that, let us assume, Sokrates intends to drink the toxin for a right kind of reason. Let us, for the sake of argument, assume that Sokrates' reason is that he finds the toxin tasty.[23] Together with the other conditions

---

[22]Hannes Leitgeb, in personal communication, asked me to clarify why what I am discussing in this chapter is different from the question at issue in the ethics of belief debate between W. K. Clifford (Clifford, 1877 [1999]) and William James (James, 1897). The question at issue between Clifford and James concerns the question whether it is 'wrong always, everywhere, and for anyone to believe anything on insufficient evidence.' Clifford thought that it is, whereas James thought that in some rare cases it might be ok or even right to believe something on insufficient evidence. Translated into the terminology of this chapter the question is whether a reason for believing something can sometimes be a good reason even though it is of the wrong kind for belief. James thinks that it can, but only if some specific conditions are satisfied. (For those conditions see James famous "Will to believe" (James, 1897).) Clifford's question, however, is different from the question at issue in this chapter which is whether a piece of reasoning can give one a right kind of reason for a conclusion-belief even though the right kind of reason for a premise-state is not a truth-related reason. I think that James would agree with Pascal that even though it might be rationally permissible to believe in God on the basis of practical considerations, these considerations are not truth-related. Hence, the debate between Clifford and James is orthogonal to the issue discussed here.

[23]Of course I do not want to claim that the historic Sokrates intended to drink the

that are satisfied in his case the reasoning for crime seems sufficient to give him a reason of the right kind to intend to rob a pharmacy.

So, on the one hand my discussion of cases like Pascal's and Kavka's suggests that at most what one can get out if one puts a WKR in, is a WKR. Or, equivalently, one cannot get a RKR out, if one puts a WKR in. On the other hand, the cases of Anselm and Sokrates suggest that we can get a RKR out if all the reasons we put in are RKR.

This suggests that the following might be a general principle:

RKR-Transmission

> An agent has a RKR for the conclusion-state of a piece of reasoning (theoretical, practical-instrumental...) if and only if

  (i) the reasoning is governed by a requirement of rationality,

 (ii) the agent holds the premise-states (beliefs, intentions...), and

(iii3) the agent holds all the premise-states for RKR.

If RKR-Transmission is in fact a general principle, then what noncognitivists have to show in order to show that Edgar has a RKR for accepting C is that it is possible to accept P1 and P2 for RKR. This raises the question of what are, intuitively, RKR for accepting P1 and P2.

7.3.5 What is the RKR for Accepting P2?

If noncognitivism is true, is it possible to accept P1 and P2 for a RKR? Let us ignore P1 here, since Dorr also does not problematize it.[24] In order to answer this question we need to know what *intuitively* is a RKR for P2. Unfortunately, it is not clear what a right kind of reason for P2 is, because it is generally not obvious what could be reasons for a moral beliefs. This holds especially for foundational moral beliefs, such as the belief that pain is bad or that pleasure is good in itself. For foundational non-moral beliefs, such as the belief that this thing in front of me appears red, it is less problematic to tell a story about what justifies them. One thing that justifies such a belief is the fact that I see that it is red or I have a visual experience of redness, when looking at it. In the case of moral beliefs, it is unclear how they could be justified through perceptual experience or testimony. The question of what

---

hemlock for a right kind of reason (his reason surely wasn't that he found hemlock tasty), though his reasons for wanting to drink it might have been good reasons.

[24]Dorr does not problematize it not because he thinks it is unproblematic, but because he thinks P2 already problematic enough to raise his objection. Arguably, P1 is in fact even more problematic than P2, and arguably noncognitivists need to say much more about it in order to solve the wishful thinking problem. I will have to say a bit more on this issu at the end of this chapter.

is a reason for a moral belief is pressing quite independently of cognitivism and noncognitivism. But intuitively, the fact that murder causes great harm provides a reason for believing that murder is wrong. What kind of reason is this?

Prima facie, the reason for believing that murder is wrong which is provided by the fact that murder is harmful looks like a TRR because it seems to bear on the question of whether or not murder is wrong. But this raises a problem because if the reasons for moral beliefs intuitively look like TRRs, then they would be RKRs for cognitive states. This would, of course, nicely fit the cognitivits. But for the noncognitivists this means that they have to explain how something that looks like a TRR, and would thus be a RKR for a cognitive state, can be a RKR for a noncognitive state.

What if the RKR is a TRR?

Can noncognitivists explain that what looks like a TRR is in fact not a RKR for a cognitive state, but a RKR for a noncognitive state? I suggest that noncognitivists give the following answer: with respect to accepting P2, and with respect to moral claims more generally, what looks like a TRR, and what would thus be a RKR for a cognitive state, can also be understood as an action-related (ARR) or value-related reason (VRR), and thus can also be understood as a RKR for a noncognitive state. An action-related reason, I have said above, is a reason for a mental state that is provided by some fact that also provides a reason for action. A value-related reason is a reason for holding a mental state that is provided by some feature of some object that makes it valuable (or disvaluable). If what looks like a TRR for accepting P2 can be understood as a ARR or VRR, then what looks like a TRR and would thus be a RKR for a cognitive state might in fact be a RKR for a noncognitive state.

This might sound surprising, but it is in fact not that implausible that sometimes it is not so clear for what kind of mental state some fact is actually a reason for. Take, for instance, the fact that the edge of the cliff is slippery. We might take this to constitute a reason for *believing* that it is dangerous to walk on the edge of the cliff. But we might also take it to constitute a reason for *fearing* to fall off of it which is a reason for a noncognitive state. Likewise in the moral case: the fact that murder is harmful can be regarded as providing a truth-related reason for accepting 'murder is wrong', which if cognitivism is correct would be a cognitive state. But it can also be regarded as an action- or value-related reason. If murder is harmful, then something about the action or value of murdering, namely its harmfulness, might be regarded as providing a reason for a certain pro-attitudes (such as for instance, disapproval of murdering). And likewise the fact that lying is harmful also plausibly provides an action-related reason, that is a reason for a certain action, namely refraining from murdering. On this view the harmfulness of

murdering provides a reason for disvaluing lying or for intending not to lie, which are both noncognitive attitudes.[25] Noncognitivism can be understood as the claim that in the case of beliefs about moral matters we mistake noncognitive pro-attitudes with non-moral contents with cognitive states with moral contents. So it would not be surprising, the noncognitivists can claim, if, for that reason, we also mistake what are in fact action-related reasons for noncognitive states for truth-related reasons for cognitive states.

It is an interesting observation that a potential reason for accepting P2 can plausibly be interpreted as TRR or ARR/VRR, as this is compatible with cognitivism as well as noncognitivism. What intuitively looks like a TRR for accepting P2, and would thus be a RKR if accepting P2 is a cognitive state, can also be regarded as a RKR for accepting P2 if it is a noncognitive state.

If this argumentation is sound, noncognitivists can make sense of what it means to have what looks like a TRR for accepting P2: it is to have a ARR or VRR for a noncognitive state. Hence, if what Edgar learnt by reading his philosophy books and reflecting on his moral commitments is that lying is harmful or violating others autonomiy, say, then noncognitivists can claim that this in fact provides him with a RKR for accepting P2. Since Dorr seems to grant the noncognitivists that it is possible that Edgar in addition also has a RKR for accepting P1, they can conclude (via RKR-Transmission) that this gives Edgar a RKR for accepting C.

RKR in, RKR out?

In the previous section I have argued that noncognitivists can make sense of how what intuitively appears to be a TRR, and would thus be a RKR for a cognitive state, can also be understood as being a RKR for a noncognitive state. Yet, it might be objected that even if noncognitivists can make sense of how something can be a RKR for accepting P2, the reason is still *only* an ARR or VRR, and this might be insufficient for generating a RKR, that is TRR, for accepting C.

The worry is this. In all the cases I discussed above (concerning the theoretical reasoning for resurrection and the practical reasoning for crime) the following seemed to hold: the 'strongest' reason that a piece of rational reasoning generates is at most as 'strong' as the 'weakest' reason for a premise state. What I mean by this is that we have seen that if we put in an ARR in to a piece of reasoning, the most we get out is a ARR, and this holds even if we have TRR for the other relevant premises. For instance, in the crime case when we put in an ARR for intending P6, as we did in Sokrates' case, we received a ARR for (C3). The same held for the cases of Anselm and Pascal:

---

[25]Note how well this fits with Schwitzgebel's view which we confronted in the section on dispositionalism about belief that the boarders between morally *believing* that x is better than y and *valuing* x over y are not perfectly sharp.

if we put in a Non-WKR (in our case an ARR) for P4 also only leads to a Non-WKR (again an ARR) for C2.

This raises a worry about the noncognitivists' strategy: if accepting P2 is a noncognitive state the RKR for which is a ARR or VRR, then how can this result in a RKR for C which must be a TRR, since it is a cognitive belief state? In other words, can a piece of reasoning really generate a RKR (i.e. TRR) for a cognitive conclusion-state, when one of the RKR we put in is an ARR or VRR for a noncognitive state? Is putting in RKRs really sufficient to get RKR out, no matter what kind of state the conclusion-state is? Is this not still wishful thinking?

Lets have a look at what happens in a case of, what I will call, *paradigmatic* wishful thinking: reasoning from a desire, and a desire alone, to a belief. In such a case it seems clear that we do not get a RKR for a cognitive conclusion-state, even if we have a RKR for the noncognitive premise-state from which we started. Consider the following paradigmatic piece of wishful thinking:

Paradigmatic wishful thinking

(P7) Desire that I will not be infected with the corona virus.

(C4) Belief that I will not be infected with the corona virus.

Arguably, this piece of reasoning will not give you a RKR for the conclusion-belief, even if you hold the premise-desire for a RKR. Arguably even if you desire not to be infected by the virus because the virus is potentially lethal, that is you hold your desire for a RKR, this does not give you a TRR, and thus RKR for believing you will not be infected by it. Does this show that having a RKR for the premise-states of a piece of reasoning is insufficient for getting a RKR for the conclusion-state?

I do not think so. There is a crucial difference between Edgar's case and paradigmatic wishful thinking: Edgar's modus ponens reasoning is governed by a requirement of rationality, whereas paradigmatic wishful thinking is not. In other words, in cases of paradigmatic wishful thinking the rationality condition (i) of RKR-Transmission is violated.

So, I suggest that noncognitivists should insist that the explanation of why paradigmatic wishful thinking does not give agents a RKR for the conclusion belief is not that the premise state is a noncognitive state. Rather the explanation why this form of reasoning fails to give agents a RKR for the conclusion belief is that paradigmatic wishful thinking is not rational. So I suggest that noncognitivists answer the question of how it is possible to put in a ARR/VRR and get a TRR out, by stressing the fact that it must be possible that the reasoning in question can in principle be rational, that is it is subject to a requirement of rationality. So, cases of paradimatic wishful thinking do not provide counterexamples to RKR-Transmission. The reason

why paradigmatic cases of wishful thinking do not provide us a RKR for a belief even if we have a RKR for the desire on which we base our belief, is not that the our premise-state is a noncognitive state, but rather in such cases condition (i) of RKR-Transmission is violated.

7.3.6 A Counterexample to RKR-Transmission?

My proposal might be attacked by attempting to provide a counterexample to RKR-Transmission. It might be claimed that there are cases in which a piece of reasoning is subject to a requirement of rationality and an agent is in all the premise-states and for a RKR, yet the agent does not plausibly get a RKR for the conclusion state, in particular not a TRR for a cognitive state. One might think that such a case can be given by a piece of backwards practical reasoning. I shall now consider such a case.

Backwards Reasoning

RKR-Transmission states that putting RKR in into a piece of rational reasoning is sufficient for getting RKR out, no matter whether the states involved are cognitive or noncognitive. It might, however, be objected that this is still insufficient. Consider the following group of states for instance:

P8 Belief that if I get killed in a car accident, my wife will be set for life financially.

P9 Intention that my wife will be set for life financially.

C5 Intention that I get killed in a car accident.

Let us assume that this form of reasoning is governed by the instrumental requirement of rationality, which we can formulate as follows:[26]

Instrumental Requirement: Rationality requires that if you believe if p, then q, and you intend q, then you intend p.

According to the instrumental requirement of rationality it is rationally required of you that if you hold the belief P8 and the first intention P9, then you also hold the second intention C5. There is, however, a widely known

---

[26]Actually, it is not clear that this is the correct formulation of the requirement. Most philosophers of rationality prefer not to formulate the instrumental requirement in terms of a conditional, but rather in terms of 'necessary' or 'sufficient means' or 'means implied' (Broome, 2013, 160ff). Schroeder, in an exercise in his book on noncognitivism (M. Schroeder, 2010a, 101) uses a conditional formulation, but he chooses to formulate it in parallel to modus ponens reasoning, that is he states that one is required to intend the consequent of a conditional one believes if one intends its antecedent. However, Schroeder surely uses this simple conditional formulation only for the sake of exposition, not because he takes it to be the correct formulation.

problem. Requirements of rationality prescribe only that particular patterns of coherence hold among your mental states. But requirements themselves do not offer guidance on what the correct ways are to bring it about that those patterns hold given your current epistemic situation. In general, with respect to the instrumental requirement there are three ways to be rationally coherent: intend C5, not intend P9 or not believe P8. But even though all three ways are ways of becoming instrumentally rational, not every way is a correct way to respond to every possible epistemic situation. To see this consider the following case.

Suppose that Allan believes P8 and he does so for a RKR. As a loving husband he also intends P9 and also for a RKR. Quite unsprisingly, however, Allan does not intend to get killed in a car accident C5, and he has strong independent RKR for not intending this (e.g. he loves his life more than anything else). Currently he is rationally incoherent, since he holds a means-end-belief, intends the end, but does not intend the means. One way to "abstractly"[27] satisfy the requirement would be if he did not believe P8. So, may he therefore not be justified in dropping this belief? Does his intending P9 for a RKR and his not intending C5 for a RKR give him a reason to reject his belief in P8? The answer seems to be a clear No. doing so seems to be a case of wishful thinking. At most it can rationally bring him to *reconsider* the original reasons for his belief.[28] But it does not give Allan a direct reason for dropping the belief even though all my conditions are satisfied, in particular it does not give Allan a RKR for dropping his belief in P8. So it seems that we have a counterexample to RKR-Transmission.

Not Every Requirement Entails a Basing Permission

I think that this objection can be answered. A full answer would take me too far afield, however, and so I will only roughly sketch what I think noncognitivists should say in defense. According to John Broome, every genuine non-contraposed requirement of rationality implies what he calls a 'basing permission of rationality' (Broome, 2013, 189/258). Call this the 'basing thesis'. Roughly, a basing permission tells you not only how your mental states should be related in order to be rationally coherent, as requirements of rationality do. Rather they tell you which states you are rationally allowed to base on others, that is which changes you are rationally allowed to make given your current epistemic situation.

Most contraposed requirements of rationality do not imply a basing

---

[27]I take this notion from (Wedgwood, 2017, 18).

[28]This is something Dorr also explicitly grants. According to Dorr, the right way for desires and feelings to influence beliefs is "by prompting you to seek out new evidence, rethink your beliefs about what the old evidence supports or engage in a priori reflection, rather than by causing you simply to rule out some possibility without regard of the evidence" (Dorr, 2002, 99).

permission of rationality. The instrumental requirement above is a well-known example. One conforms with the requirement in the above-mentioned ways, but not every way to bring it about that one conforms with it is rationally permissible. For instance, it is rationally permissible to base an intention that p, on your belief that if p, then q, and your intention that q. But, and this is the point, it seems rationally impermissible to drop your belief that if p, then q, on the basis of your intending q, and your not intending p even though doing so would make your mental states rationally coherent.[29] This phenomenon is discussed in the literature on rationality under the heading of the 'asymmetry' of wide-scope requirements of rationality.[30] The crucial point in our current context is that noncognitivists can accept that RKR-Transmission fails for contraposed requirements, while accepting that it holds for all non-contraposed requirements. Hence, at most what the above objection shows is that we need to refine RKR-Transmission to account for the impermissible cases. To do this we simply modify condition (i) by explicitly restricting it to non-contraposed requirements:

refined RKR-Transmission An agent has a RKR for the conclusion-state of a piece of reasoning (theoretical, practical-instrumental...) if and only if

(i2) the reasoning is governed by a non-contraposed rationality requirement,

(ii) the agent holds the premise-states, and

(iii3) the agent holds all the premise-states for right kinds of reasons.

Since, as I have argued, it is possible for Edgar to satisfy all those conditions even if noncognitivism is true, this solves the wishful thinking problem.

Let me quickly summarize what I have argued in the previous paragraphs. I have argued that in order for a piece of reasoning to provide a RKR for the conclusion-state agents must also hold the premise-states for the RKR and the reasoning must be governed by a genuine non-contraposed requirement of rationality. The idea has been that noncognitivists can utilize this principle to argue that even if the RKR for some premise-state is not a TRR for a cognitive state, but a ARR/VRR for a noncognitive state, this is nevertheless sufficient for providing reasoners with a RKR for a cognitive conclusion state, and so with a TRR. I argued that the crucial difference between moral modus ponens reasoning and paradigmatic wishful thinking is that wishful thinking cannot be rational whereas it is part of the noncognitivist's program to defend that moral modus ponens reasoning can be rational (see chapter 3). So I suggested that the fact that Edgar's reasoning can be rational,

---

[29]Cf. Broome, 256.

[30]See for instance, (M. Schroeder, 2004).

whereas paradigmatic wishful thinking cannot, is the reason why putting in a RKR delivers a RKR as output. I presented a potential objection levelled against the joint sufficiency of conditions (i), (ii), and (iii3) for providing an agent with a RKR for the conclusion state of a piece of reasoning. I argued that this objection can be answered by restricting RKR-Transmission to non-contraposed requirements. Since Edgar's case is a case of modus ponens reasoning is it covered by the non-contraposed modus ponens requirement, and so RKR-Transmission allows noncognitivists to conclude that if Edgar has a RKR for all premise-states, this gives him RKR for the conclusion-state, even if some premise-state might be a noncognitive state.

Much of the strength of Dorr's objection stems from the fact that paradigmatic wishful thinking cannot justify our beliefs, and that if noncognitivism is true, then Edgar's case resembles paradigmatic wishful thinking in that in both cases we base a belief one a desire-like state. Dorr suggests that what is bad about wishful thinking is the fact that it bases a belief on a noncognitive state. From this he seems to conclude that we can *never* be justified in basing a belief on a noncognitive state. Much of my reply to his objection is to argue that this is an overgeneralisation: not every form of basing a belief on a noncognitive state is so obviously bad as paradigmatic wishful thinking. It is true that in the case of paradigmatic wishful thinking it is bad to base a belief on a noncognitive state, but, I suggest, this is not due to the premise-states' being noncognitive, but rather due to the reasoning's not being governed by a requirement of rationality. What makes paradigmatic wishful thinking bad is not the fact that one bases a belief exclusively on a desire, but that doing is irrational.

7.3.7 A Case of Basing a Belief on a Noncognitive State

I suspect that, despite all my efforts, not everyone will be convinced of the truth of RKR-Transmission even in its refined version. What would really be nice to have is a positive example where it is intuitively convincing that basing a factual belief on a noncognitive state is justified. One might think that giving such a case is impossible because, as I have just suggested, this would be a case of paradigmatic wishful thinking and thus bad thinking. But I think that such a case can actually be given.

Suppose, at least for the sake of argument, that the following is a genuine requirement of rationality:

Intention-Belief: Rationality requires that if you intend that p, then you believe you (probably, possibly) will p.[31]

---

[31]It is controversial that there really is such a norm. Michael Bratman, for instance, is cautious. He holds that "there is, other things being equal, an important kind of irrationality involved in intending to act in ways inconsistent with one's beliefs. ... But there need be no irrationality in intending to *A* and yet still not believe that one will.

Now consider the following pair of states:

P10 Intention to stop at the red light.

C6 Belief that I will (possibly, probably...) stop at the red light.

Suppose then you intend to stop at the red light, and you have good reasons, in fact good reasons of the right kind for your intention (e. g. you know that crossing a red light can be very dangerous). Would this give you a reason, in fact a truth-related reason, for believing that you (probably, possibly) will stop at the red light? I am inclined to say it does. If someone asks you why you believe that you will stop at the red light and you answer because you intend to stop there, it is not absurd to regard this as a justification for your belief. Compare this with a case where someone asks you why you believe you will not be infected by the corona virus and you answer that you believe this because you do not want to be infected by it. It is absurd to regard this as a justification of your belief. Hence, the case of P10 to C6 is a case that even though it even structurally resembles paradigmatic wishful thinking, here holding a desire-like state gives one a reason for a belief, and it seems that part of the explanation of why this works is again that this kind of reasoning is governed by a requirement of rationality.

Of course, the plausibility of this case depends on a number of controversial issues like, most importantly, the question of whether the alleged requirement really is a requirement, but also on the question of whether intending p implies the belief that one will p.

But this should also make you careful when criticizing the case. When you criticize this argument you have to bear in mind that perhaps, contrary to what I have been assuming for the sake of argument, Intention-Belief is not really a requirement of rationality. Part of my thought experiment is the assumption that it *is* a requirement, but that does not mean that it actually is a requirement. Perhaps there are no requirements like this.[32] Moreover, when criticizing this case we must also bear in mind that perhaps

---

In contrast, there will normally be irrationality in intending to A and believing that one will not A; for there is a defeasable demand that one's intentions be consistent with one's beliefs" (Bratman, 1987, 38). Bratman distinguishes both cases. He calls cases of intending to A without believing that one will cases of 'intention-belief incompleteness', and cases in which one intends A and beliefs that one will not A cases of 'intention-belief inconsistency'. He thinks "intention-belief inconsistency is closer to criticizable irrationality than intention-belief incompleteness" (p.38.Cf. also 133).

[32]It might also be suggested that such requirements are not requirements of rationality, but are rather rooted in what one could call evidential norms. E.g. it might be claimed that rationality requires that if one desires that p, then one believes that one desires that p. This, it might be argued, is not a requirement of rationality governing the coherence of mental states, however. Instead, it might be said this 'requirement' derives from the norm that one must belief what is true. Hence, if it is true that you desire that p, you must belief that you desire that p, simply in virtue of the fact that you ought to believe

Intention-Belief might not appear to be a requirement because of the way it is currently formulated. Most philosophers of rationality offer much more complicated formulations of what they take to be requirements in order to deal with counterexamples. For instance, many of the requirements Broome discusses include an ability condition.[33] It is highly likely that we must add to the above Intention-Belief requirement the condition that not only do you intend p, but also you believe that it is up to you whether you will p, that is you believe you are able to do p or that it is within your power that one does p.

7.3.8   Exploiting P1

Even if we do not find that the red light example is an example of a case where it is rationally permissible to base a belief on a noncognitive state, there is a further crucial difference between Dorr's case of Edgar and the red light case, which might explain why noncognitivists do not have to treat Edgar's case as a case of wishful thinking. Unlike in the red light case, Edgar's case does not only involve a single premise the acceptance of which is a noncognitve state. Instead it involves two premises, of which the first is a mixed moral-non-moral sentence. In Edgar's case it is important that Edgar accepts P2 and for a RKR. But it is equally important that he accepts the mixed P1 and has a RKR for it. Noncognitivists could and should argue that the interaction of P1 and P2 as well as the fact that one needs reasons for accepting them are both essential to explaining why it gives Edgar a RKR for accepting C. This has also been noted by Schroeder who says that even though "Edgar starts by accepting P1 and comes to accept P2, his acceptance of P1 is not, itself, irrelevant to the justification for C" (M. Schroeder, 2010b, 182f.).

Of course, noncognitivists owe us an account of what will be a RKR for accepting P1, and an answer will depend on their account of the nature of this state. I will not try to offer such an account here. But I want to mention what I take to be the most promising direction towards an answer. P1 is a mixed sentence. It has a moral antecedent, and a non-moral consequent. For noncognitivists the state of accepting P1 must somehow be a function of the states expressed by the antecedent and the consequent.[34] There will be RKR for accepting the antecedent, as well as RKR for accepting the consequent. Hence, noncognitivists must argue that the reason for accepting P1 is somehow a function of the reasons for accepting the antecedent and the consequent. This is unlikely to be a trivial matter, because what can be

---

only what is true, and not because there holds some rational relation between the desire and your belief to desire.

[33]Cf. (Broome, 2013, 162).

[34]This is part of the Frege-Geach problem, more precisely, of what I have called the constructablity problem.

a reason for the antecedent will normally not be a reason for the accepting a complete conditional.[35]

Since the acceptance of P1 stands in a construction-relationship to the descriptive belief that would be expressed by its consequent, it is not totally implausible that the combination of P1 and P2 can deliver a TRR for accepting C. In other words, in contrast to paradigmatic wishful thinking and my positive case (red light), Edgar's case does not start from purely noncognitive premise-states. This might explain how it is possible that his reasoning is capable of giving him a TRR for believing C, though neither of the reasons for the premise-states is strictly a TRR.

I am well aware that is not the end of the story. Much of my discussion in this chapter was fairly abstract, since I did not want to presuppose any specific noncognitivist theory. My goal was to investigate an unexplored path to solve the perhaps most notorious problem for noncognitivism in epistemology. Noncognitivists who want to pursue this particular path still have to show how exactly the details can be filled out in order to receive a plausible theory. The details might differ tremendously depending on what kind of noncognitivist theory one defends. In this chapter, I have assumed that we are concerned with local forms of noncognitivism according to which purely non-normative sentences (e. g. 'liars will be punished in the afterlife') express cognitive states (the belief that liars will be punished in the afterlife), and I have ignored what a global noncognitivist of a Schroederian kind could say in reply to the wishful thinking objection. But even within the camp of localists there are so many different conceptions regarding the structure and nature of the states expressed moral and mixed sentences that I cannot investigate this here further.

7.4   Conclusion

In this chapter, I investigated a new direction for solving Cian Dorr's famous wishful thinking problem. Dorr's major claim is that if noncognitivism is true, then accepting the premises of a valid argument containing moral sentences cannot give one a reason of the right kind for accepting its conclusion, especially when the conclusion is factual. In his view basing a belief on desire-like states can never be justified—doing so is wishful thinking. I tried to make it plausible that reasoning (partly) from desire-like states to beliefs need not be wishful thinking, but can in fact give one a reason of the right kind for a factual belief, namely when the reasoning is governed by a requirement of rationality, and one accepts the premises also for right kinds

---

[35]Explaining how the reason for a conditional is a function of the reasons of its elements is further complicated by the fact that intuitively being told that P1 (i.e. testimony) can provide a RKR for accepting P1, whereas being told that P2 is not obviously a way of providing a RKR for accepting P2.

of reasons.

Dorr's discussion suggests that if noncognitivism is true, then moral reasoning can bootstrap reasons for factual beliefs into existence merely by adopting a noncognitive state. If what I have argued in this chapter is correct, noncognitivism does not suggest that it is that easy. In order for a piece of reasoning to generate a reason for a factual belief the reasoning must be governed by a genuine requirement of rationality and one needs to have the right kinds of reason for all the premise-states. These are demanding conditions to be satisfied, but they can be satisfied even if noncognitivism is true.

If what I said in this chapter is correct, then moral reasoning is crucially different from paradigmatic wishful thinking even if noncognitivism is true. Moreover, it turns out that reasoning with moral beliefs differs from paradigmatic theoretical reasoning (i.e. from factual beliefs to factual beliefs) as well as from paradigmatic practical reasoning (i.e. from factual beliefs and intentions to intentions), though there are similarities between them as well. Just as theoretical reasoning, moral reasoning can lead to new beliefs, and just as practical reasoning it can start from combinations of cognitive and noncognitive states. So moral reasoning might be considered to be a special form of reasoning, located in between paradigmatic theoretical and paradigmatic practical reasoning. This fits noncognitivism because noncognitivists have always emphazised the similarities as well as the differences between moral and factual beliefs, and they also tried to respect the practical dimension of holding moral beliefs. Noncognitivist theories of moral reasoning account for the similarities as well as the differences and explain why it is a special kind of reasoning. Other than the cognitivists they do not treat reasoning with moral beliefs as pure theoretical reasoning, even though sometimes moral claims can be used to justify purely factual claims.

Part III

Moral Motivation:
Internalism & Humeanism

Chapter 8

Two Problems with Motivation

8.1 Introduction

In this chapter, I am concerned with answering our final question: *how can moral beliefs motivate to act?* It is almost a dogma of metaethics that noncognitivism is better at explaining moral motivation than cognitivism.[1] In this chapter, I want to challenge the dogma. Though I am not the first to question whether noncognitivism is better at explaining *all* phenomena regarding moral motivation, most prominently amoralism and akrasia, my claim is more radical: noncognitivism is just as bad as or even worse than cognitivism when it comes to explaining moral motivation.

In support of this claim this chapter presents two unnoted problems for noncognitivism with explaining moral motivation. The *first problem* is that Humeanism plus internalism not only entails the falsity of cognitivism (the well-known *motivation argument*[2]), but also entails the falsity of noncognitivism. The problem is so simple that it is surprising that it has been unnoticed so far: if motivation requires a belief as well as a desire (Humeanism), and moral judgments motivate by themselves (internalism), then moral judgments cannot be desires only.

---

[1]As it is often with dogmas, one rarely finds people stating them explicitly. One of the best places to find them is by looking into introductory textbooks. For instance, Andrew Fisher in answering what makes noncognitivism attractive says that it "helps to explain the practical nature of morality" (Fisher, 2011, 32). Similarly, Richard Joyce writes that "one of the major attractions of noncognitivism is that it ... does a good job of accounting for the motivational efficacy of moral judgment" (Joyce, 2021, 3.1). Alexander Miller in his widely read introduction to metaethics states "that non-cognitivism sits better with the best account of moral motivation, the Humean theory of motivation" (Miller, 2003, 56). For similar claims see (van Roojen, 2015, 159).

[2]For an excellent discussion of the motivation argument see (Eggers, 2014) who in his first footnote also lists many of the relevant papers on it which, for this reason, I do not want to recite here.

The *second problem* is that Humeanism plus externalism also strongly suggest the falsity of noncognitivism. Again, the argument is simple: moral judgments when combined with suitable external desires motivate to act (externalism), but if motivation requires belief as well as desire (Humeanism), then moral judgments cannot be desires only.

This chapter is structured as follows. I start my discussion of the first problem by clarifying the central positions involved in it, that is cognitivism and noncognitivism, motivational internalism, and the Humean theory of motivation (section 8.2). Next I deal with what I take to be the major objection to my first problem for noncognitivism, namely that I am misconstruing the theses of Humeanism or internalism and that on the correct understanding the problem dissolves (section 8.3). To this I reply by making two points. First, if we look into the literature on Humeanism and internalism, we see that I do not misconstrue those theses. Second, I argue that if we modify those theses in ways that would actually allow the problem to be solved, then those modified theses stop being plausible. I conclude that although those modification may allow noncognitivists to solve the first problem with motivation, from the fact that those substantial modifications are needed in order to solve the problem it follows that noncognitivism is not obviously better at explaining moral motivation than cognitivism.

In section 8.4 I then turn to the second problem for noncognitivists with explaining moral motivation. The second problem, I argue, forces noncognitivists to give up Humeanism, and demands an anti-Humean solution. The major challenge for such a solution is to explain, in a systematic way, how pairs of noncognitive desire-like states can interact to generate the relevant motivation. Even though I think that my own dispositional form of noncognitivism is capable of solving the problem, I leave it open how good this solution is, and whether other forms of noncognitivism will be able to provide similar solutions. Since the second problem is a problem Humean cognitivists do not have, this suggests that noncognitivism is not only not better, but actually worse at explaining moral motivation than cognitivism which contradicts the above-mentioned dogma.

8.2   First Motivation Problem for Noncognitivism

A central claim of metaethics is that there is an internal or necessary connection between moral judgment and motivation. This claim is so-called 'motivational judgment internalism', short: *internalism*. It is a major challenge in metaethics to provide an explanation of the truth of internalism. It is generally thought that noncognitivists have an easier time explaining internalism than cognitivists, especially when we assume the Humean theory of motivation. The first problem I present in the following section is supposed to show that this is false: noncognitivism is just as bad at explaining

internalism than cognitivism. The second problem, which I will introduce further below, is then supposed to show that noncognitivism is even worse at explaining moral motivation more generally than cognitivism.

In this chapter three positions take center-stage, namely noncognitivism, internalism, and Humeanism. I start by saying what those positions are.

8.2.1 Cognitivism and Noncognitivism

As in the previous chapters I will understand cognitivism and noncognitivism as claims about the nature of the mental states expressed by moral sentences (i.e. moral judgments). In this chapter, I will understand cognitivism and noncognitivism as the following claims:

Cognitivism: moral judgments are beliefs (and nothing else, in particular not desires).

Noncognitivism: moral judgments are desires (and nothing else, in particular not beliefs).

I will in the following not always mention what is written inside the brackets, but leave it implicit. Moreover, please note that although noncognitivism is incompatible with cognitivism, noncognitivism is not the negation of cognitivism. Furthermore, note that here I have chosen to define cognitivism and noncognitivism not in terms of cognitive and noncognitive states (as I have done in the previous section), but in terms of belief and desire. I will explain my choice of terminology after I have introduced internalism and Humeanism in the next sections.

8.2.2 Motivational Internalism

What is the thesis of so-called 'motivational judgment internalism', or short: *internalism*? Unfortunately, even though internalism is regarded as one of the most central theses in metaethics, there is no consensus about the precise formulation of it. Most often when metaethicists speak about motivational judgment internalism they introduce it as the claim that there is an *internal* connection between moral belief and motivation, and this is where internalism has its name from.

We will see in a moment that internalism can be formulated in a variety of ways, but the following is one of the most widely accepted formulations, and this is also how I will understand the position here:

Internalism: It is possible that if a person holds a moral judgment, then she is motivated by that judgment itself to act in accordance with it.

In short, internalism says that moral judgments can motivate by themselves.

Some comments. The 'by itself' in this formulation of internalism is supposed to capture what is meant by saying that moral judgment has an 'internal' or 'intrinsic' connection with motivation. The idea that the motivation comes from the moral judgments themselves and not from something 'besides', 'additional' or 'external' to the moral judgment is what distinguishes it from its opposing view, namely externalism. Below I shall defend this understanding of internalism against the objection that I am misconstruing what internalists mean by 'internal'.

'To act in accordance with the moral judgment' is normally understood as capturing some relation between the content of the moral judgment and the object of one's motivation. Many internalists make this explicit by formulating internalism in something like following way: if someone believes (or judges) that it is right to ϕ, then she is motivated to ϕ; or similarly, if someone believes it is wrong to ϕ, then she is motivated to refrain from doing ϕ.[3] However, for the argument I will make below it is not important that a person who holds a moral judgment is motivated to act in accordance with it. For my argument to work I only need the weaker claim that if a person holds a moral judgment, then she can be motivated by the judgment itself.

The 'is motivated' in formulations of internalism is often qualified by an 'at least somewhat'. This is done in order to capture the nowadays widely shared and plausible assumption that, pace Plato, the motivation moral judgments generate is not overriding, in the sense that it cannot be outweighed by motivation coming from other sources besides morality, such as, for instance, prudence. In the following I will implicitly take 'is motivated' to be equivalent with 'a least somewhat motivated'.[4]

Internalism, as it is formulated here, is what has been called 'unconditional' internalism because it does not include any additional conditions, such as a *ceteris paribus* clause, or the condition that the person must be rational, strong-willed, or psychologically normal etc.[5] As far as I can tell, these conditions are largely irrelevant to my argument.[6]

Finally, internalism is often formulated as a necessity claim. I, however, use the weaker formulation as a possibility claim. It has been emphasized by Swartzer (2015), Sinhababu (2017), Eggers (2014) and others that the

---

[3]Daniel Eggers, however, has argued that 'to act in accordance with' can be understood a broader fashion (Eggers, 2014, 2457). He does so mainly to make room for a noncognitivist treatment of akrasia, but this broader understanding is not what internalists normally have in mind.

[4]You can drop the implicit qualification, if you think that the motivation provided by moral judgments is overriding. Anyways, nothing in my argument hinges on this qualification.

[5]For more on the distinction between conditional and unconditional internalism see (Björnsson et al., 2015, 1.4) and in the same volume (Eggers, 2015).

[6]Though below, I will shortly consider the rationality condition in an attempt to solve the noncognitivists problems with motivation. But there I shall argue that this solution is unsuccessful anyways.

internalists do not have to make the stronger claim that moral judgments always motivate by themselves, but it is enough to claim that they *can* motivate by themselves because it is precisely this possibility that the externalist denies. Moreover, the weaker claim is in fact preferable to the stronger, because so internalists might allow for the possibility of amoralism and akrasia, that is cases in which someone holds a moral judgment, but is completely unmotivated by it.

8.2.3 The Humean Theory of Motivation

The *Humean theory of motivation* is the combination of the following two claims:

Humeanism: a person S is motivated to do M iff S has a means-end-belief that M is a means to end E and S has a desire to do E.

Distinctness: Belief and desire are 'distinct existences' in the sense of being modally separable: for any pair of belief and desire it is always possible to imagine that someone holds the belief, but not the desire, and *vice versa*.[7]

According to the Humean theory of motivation, having a suitable pair of the distinct attitude-types of desire and belief is necessary and sufficient for having a motivation.

I have formulated the Humean theory of motivation, as it is traditionally done, in terms of beliefs and desires. I could formulate it in terms of cognitive and noncognitive states, or in terms of representational or motivational states. Likewise, I could also formulate it in terms of different directions of fit, or in terms of different functional roles. Nothing hangs on this terminological choice. The only thing that is important to the Humean is that whatever way we describe those states, states of both types are needed for motivation, and they are assumed to be categorically distinct.

Why do Humeans think that belief and desire are categorically distinct types of mental states? Michael Smith offered an argument that no belief can be a desire and vice versa based on the different directions of fit of beliefs and desires.[8] He argues that states with a mind-to-world direction, e. g. beliefs, tend to cease to exist, when we find out that the world is not how the belief represents it to be, whereas states with world-to-mind direction, e. g. desires, do not tend to cease to exist, when we find out that the world is not how we want it to be, rather they tend to persist then. Since no state can tend to cease to exist and not tend to cease to exist at the same time, the

---

[7]I separate the Humean theory of motivation into two claims because I take it to be possible, in principle at least, to accept Humeanism without holding distinctness. This will be relevant in section 8.4.2 below.

[8]See (Smith, 1994b, 115).

idea of a state that is a desire and a belief at the same time (i.e. a besire) is, according to Smith, incoherent. Hence, beliefs and desires are categorically distinct mental states.

Humeanism entails the following:

Humean Corollary 1: If S believes that M is a means to E, but does not desire E, then S is not motivated to M.

This is often formulated as the claim that belief taken by itself cannot motivate, in particular that belief cannot motivate in the absence of a desire. Humeanism also entails the following:

Humean Corollary 2: If S desires E, but does not believe that E is a means to M, then S is not motivated to M.

The second corollary is often ignored in the debate, and we shall see shortly it is what causes the noncognitivists' first problem with motivation.

8.2.4 Terminology

I want to make some comments on my choice of terminology in this chapter. In the previous chapters I used the neutral term 'moral judgments' as a shorthand for the rather cumbersome 'the states, whatever they are, expressed by moral sentences'. I explained that contemporary quasi-realist noncognitivists think they can legitimately regard moral judgments as moral *beliefs*. So it would be possible to formulate noncognitivism not as a claim about moral judgment, but about moral belief. In the first half of this chapter, however, I will stick to the term 'judgment' because I want to reserve the term 'belief' for what I previously called 'cognitive state'. The alternative would be to distinguish two different notions of belief, one for the folk-psychological one (perhaps written with a small 'b'), and another one for a philosophically more restricted one (perhaps written with a capital 'B'). Noncognitivism would then be the view that moral beliefs are not Beliefs. This would easily become confusing. Moreover, motivational *judgment* internalism, as the name signals, is also normally formulated in terms of judgment and not in terms of belief. If I we would formulate internalism as a claim about the motivational force of moral beliefs, this would also lead to confusion given the understanding of belief I use in this chapter.

This brings me to my other choice. In this chapter, I use 'belief' and 'desire' not to denote our folk psychological concepts of belief and desire (the ones which the quasi-realist wants to earn the right to). Rather I use those terms in a more restricted philosophical sense. For instance, I use the term 'belief' in the sense in which all beliefs are cognitive, representational states, have a mind-to-world direction of fit, track features of the world, aim at the truth and so on. Likewise, I use 'desire' here to refer to noncognitive, conative

states, that have a world-to-mind direction of fit, aim at their realization or satisfaction and so on. I do so because Humeanism is traditionally formulated in terms of beliefs and desires, and not in terms of cognitive and noncognitive states or directions of fit or in terms of 'distinct' functional roles, though modern Humeans clearly have a philosophically narrower understanding of those states in mind when they speak of beliefs and desires. If we wanted, we could formulate Humeanism not in terms beliefs and desire, but in those other terms, in terms of cognitive and noncognitive states, for instance. In any case: as long as we use the terminology consistently this will not make a difference to my argument. I think that my prefered way of speaking is less confusing than others, and also seamlessly fits with the citations from the internalism and Humeanism literature. That is the reason why I have made these terminological choices.

8.2.5   Motivation Argument Against Noncognitivism

With all those definitions and clarifications in place, let us return to our question from the beginning of this chapter: How can we explain moral motivation, in particular, how can we explain internalism?

Noncognitivists are standardly taken to have a ready answer. For instance, Harman writes the following: "To think that you ought to do something is to be motivated to do it... Emotivism can explain this" (Harman, 1977, 33). The explanation seems trivial: if moral judgments are desires, and desires are motivational states, then it is clear why moral judgments are intrinically or necessarily connected with motivation: they just are themselves motivational states. Arguments along those lines can be found at various places. I shall list only three here as examples:[9]

> The conventional wisdom is that the natural dialectical beneficiary of the truth of internalism is expressivism [i. e. noncognitivism]. This is because the expressivist holds that the relevant judgments just *are* motivational states. It is therefore very easy for the expressivist to accommodate even very robust forms of internalism. As it is sometimes put, the expressivist can capture internalism "for free," whereas cognitivists must tell some special story either to accommodate internalism or explain it away. (Ridge, 2015, 135)

> It should be fairly obvious that the internalist constraint prima facie favors noncognitivism: if moral judgments manifest conative states of mind, they will necessarily have some motivational

---

[9]Further examples can be found for instance in (Björnsson et al., 2015, 3), (Svavarsdottir, 1999, 167f.), (Swartzer, 2015, 6), (Rosati, 2016, 3.1), (Miller, 2003, 6), (Smith, 1994b, 7ff), and (Fisher, 2011, 139).

force, whereas if they manifest beliefs, it needs to be explained why they would—unlike most beliefs—be necessarily motivating. (Svavarsdottir, 1999, 167)

[I]nternalism is thought to lend credence to noncognitivist views of moral judgment. According to this line of argument, since moral judgments can be motivational in the way that ordinary beliefs are not, these judgments are more likely constituted by some noncognitive state. (Swartzer, 2015, 6)

Since noncognitivism appears to offer so simple an explanation of internalism, this provides abductive evidence for noncognitivism. So we may conclude that, probably, noncognitivism is true. Call this argument the *inconclusive motivation argument for noncognitivism.*

Cognitivists, in contrast, cannot offer an equally simple explanation of internalism. It is hard to see why moral judgments are internally or necessarily connected with motivation, that is why they can motivate by themselves, if moral judgments are beliefs, and beliefs are, as they are often described, representational, but motivationally 'inert' states. Since cognitivists do not have a simple explanation of internalism, this provides inductive or abductive evidence against cognitivism. Call this the *inconclusive motivation argument against cognitivism.*

It is well-known that the inconclusive argument against cognitivism can be transformed into a real argument against cognitivism if we supplement it with Humeanism. This is because cognitivism, internalism, and Humeanism are regarded as jointly inconsistent. No matter how we look at the "inconsistent triad"[10], we are forced to give up at least one of them. The problem of deciding which one to give up, and how to live with the consequences of this decision, is the "big issue of metaethics" (Pigden, 2009, 80) or the "central organizing problem in contemporary meta-ethics" (Smith, 1994b, 11): the so-called *Moral Problem.*

As a result of this inconsistency, we can make a number of arguments. For instance, we can argue that if moral judgments are beliefs (cognitivism), and if moral judgments can motivate by themselves (internalism), then beliefs can motivate by themselves (falsity of Humeanism). We can also argue that if moral judgments are beliefs (cognitivism), and if beliefs cannot motivate by themselves (Humeanism), then moral judgments cannot motivate by themselves (falsity of internalism).

The argument that interests us here, however, is that Humeanism together with internalism entails the falsity of cognitivism. This is known as the 'motivation argument', though it would be more accurate to call it the *motivation argument against cognitivism* which can be stated as follows:

---

[10]See (McNaughton, 1988, 23) and (Smith, 1994b, 12).

The Motivation Argument against Cognitivism

Humeanism: Motivation requires belief and desire. (Corollary: Beliefs cannot motivate without desire; beliefs cannot motivate by themselves.)

Internalism: Moral judgments can motivate by themselves.

Falsity of Cognitivism: Hence, moral judgments are not beliefs only.

Russ Shafer-Landau describes this argument as the "perennial problem for moral realists [i. e. cognitivists]" (Shafer-Landau, 2003, 191). In more colloquial terms he describes the argument as saying that "necessarily, moral judgements motivate. Beliefs don't. Therefore moral judgements aren't beliefs" (Shafer-Landau, 2003, 121).

As I have remarked above, however, the negation of cognitivism is not the same as noncognitivism. This implies that even though this argument is an argument against cognitivism, it is *not* an argument *for* noncognitivism. This is so despite the fact that Shafer-Landau misleadingly refers to it as the "Non-cognitivist Argument" (Shafer-Landau, 2003, 120).

Even though this argument is not strictly an argument for noncognitivism, the inconclusive argument for noncognitivism from above and the motivation argument against cognitivism together are standardly taken to provide strong reasons for believing in noncognitivism in metaethics. The motivation argument against cognitivism establishes at least the negative half of noncognitivism: that moral judgments are not beliefs. And the inconclusive argument for noncognitivism at least suggest the truth of its positive half: that, probably, moral judgments are desires. Overall this is taken to make a good case for noncognitivism.

There is, however, a simple problem that surprisingly has been ignored in the debate. The problem is that Humeanism together with internalism not only entails the falsity of cognitivism—it also entails the falsity of noncognitivism. The argument is straightforward and parallels the motivation argument against cognitivism. Let us call this argument the *motivation argument against noncognitivism* and formulate it as follows:

Motivation Argument against Noncognitivism:

Humeanism: Motivation requires belief and desire. (Corollary: Desires cannot motivate without belief; Desires cannot motivate by themselves.)

Internalism: Moral judgments can motivate by themselves. (Moral judgments can motivate without belief.)

Falsity of Noncognitivism: Moral judgments are not desires only.

Hence, the famous motivation argument in metaethics entails the falsity of cognitivism as well as the falsity of noncognitivism.[11]

8.3   Solutions?

I just made a motivation argument against noncognitivism. This argument can also be stated as follows:

Motivation Argument against Noncognitivism:

Internalism: If one holds a moral judgment, one is motivated.

Humeanism: If one is motivated, one holds (a desire and) a belief.

Falsity of Noncognitivism: If one holds a moral judgment, then one holds (a desire and) a belief.

How should noncognitivists react to this argument? Obviously, they can either reject Humeanism or reject internalism. They could, for instance, reject internalism completely, and become externalists. But this would amount a complete capitulation.[12] Giving up Humeanism would also be a hard bullet to bite, since giving up Humeanism leaves the noncognitivist without a general philosophical theory of motivation. So rejecting either of these views has serious drawbacks.

In the following sections I will discuss whether there are some less radical moves noncognitivists can make. More precisely, I discuss a number of ways of modifying either internalism, Humeanism, or even noncognitivism, without completely giving up the original spirit of those positions.

---

[11]Does this show that Humeanism is incompatible with internalism? Not necessarily. Since noncognitivism is not the negation of cognitivism, the motivation argument does not entail a contradiction. Moreover, depending on whether we take distinctness to be an essential part of the Humean theory of motivation, and depending on how we understand the distinctness claim, the motivation argument might not only entail the falsity of cognitivism and noncognitivism, but also the falsity of besire theory and hybrid theory. Without going into the debate about how to best understand or even make sense of the notion of 'besire', I suspect that on most understandings of 'besire', Humeanism cum distinctness paired with internalism rules out besire theories. But I suspect that there is a notion of 'hybrid state' that allows it not to be treated as a single unified state of besire, and therefore hybrid theory might be compatible with Humeanism, distinctness and internalism. So I think that there remains the possibility that the 'natural dialectical beneficiary' of the motivation argument is neither cognitivism nor noncognitivism, but hybrid theory. This claim has also been defended in a paper by Michael Ridge in (Ridge, 2015). But Ridge's reasons for thinking that hybrid theory best explains moral motivation are less general than the reason given here.

[12]Moreover, even if noncognitivists give up internalism, they will not only have failed to explain this central thesis of metaethics, they will still be confronted the second problem for noncognitivism which I will introduce below.

8.3.1 Modifying Internalism

Before I take a look at how one might want to modify internalism, I want to reply an objection against my argument. The objection is that I am misconstruing the thesis of internalism, and that the argument is therefore unsound. More precisely, it might be objected that I am misinterpreting what internalists mean by saying that motivation is *internal* to moral judgment. Internalists, it might be said, never claimed that moral judgments motivate *all by themselves*, that is without the presence of *any other states* beside the moral judgment itself.

However, as an empirical claim about what internalists actually say, this is false. As Eggers notes, most authors in the debate take 'intrinsically motivating' to mean the same as "motivate without the help of other mental states ... or on their own" (Eggers, 2015, 2454). Likewise, according to Russ Shafer-Landau a mental state is 'intrinsically motivating' iff it motivates "in virtue of its own nature and content" (Shafer-Landau, 2003, 147), and he is not the only one who thinks of internalism in this way as the following small collection of prominent examples demonstrates:

> The term 'internal' is metaphorical, but I think it can be cashed out as an *eo ipso* claim... Moral judgments *eo ipso* place those who make them in a motivational state to act in accordance with those judgments. The *eo ipso* here captures the idea that the motivation to act morally does not require something other than the moral judgments, such as an accompanying desire that is not itself part of the judgment. (Prinz, 2015, 62)

> Necessarily, moral judgements are potentially motivational, i.e., they can motivate the person making them to act in accordance with the judgement ... by themselves. (Eggers, 2015, 2458)

> Sometimes, when the agent engages in some motivated action, ϕ, the complete explanation of her action will appeal to her moral outlook [judgment] *itself* as the ultimate source of motivation to ϕ. (Swartzer, 2015, 5, italics added)

> in order for moral judgments to issue in action or motivation, no psychological state *or* mental act combines with the moral judgments. (Zangwill, 2015, 48)

> attractive and widely accepted, form of motivational internalism... Necessarily, if one judges that ϕ-ing would be desirable, then, if one is rational, one is *thereby* also motivated to ϕ. (Toppinen, 2015, 150, italics added)

> motivational internalism states that moral judgments are "intrinsically" motivating; in other words, they motivate *on their*

> *own* rather than in collaboration with a distinct conative state. (Svavarsdottir, 1999, 163, emph. added)

> we should understand the core claim of motivational internalism as the [...] claim that a moral judgment made by an agent about *what she herself (currently) ought to do* will entail motivations on her part to act in accordance with this judgment. (Manne, 2015, 261)

> If someone judges that it is right that she ϕs, then, *ceteris paribus*, she is motivated to ϕ. (Smith, 1994b, 12)

> it is an internal and necessary fact about an agent that, if she sincerely judges that X is good, she is motivated to pursue the course of action X. (Miller, 2003, 6)

> Necessarily, if one sincerely judges an action right, then one is motivated to some extent to act in accordance with that judgment. (Shafer-Landau, 2003, 121)

> If someone judges that it is right that she ψs then, ceteris paribus, (and without the aid of any pre-existing desire) she is motivated to ψ. (Pigden, 2009, 81)

The list of examples could easily be continued. As these examples show, internalism as I understand it here, is widely accepted in the literature. I take this to show that it is not very plausible that I am totally misconstruing the thesis of internalism. Even if not all internalists understand internalism in the way I do above, there is at least a great number of internalists who understands it in the way I do.

Additional State Internalism

It might still be objected that even if internalists often sound *as if* they claim that moral judgments motivate by themselves, they do not actually mean what they say, but have a slightly different claim in mind.

For instance, it might be claimed that when internalists, like Jesse Prinz or Charles Pigden above, say that moral judgments motivate without accompanying desires, they implicitly want to say that moral judgments motivate at least when combined with an additional belief. So it might be argued that internalism really is the view that moral judgments motivate when combined with an additional belief. Call this view 'additional-belief-internalism':

Additional-Belief-Internalism (ABI): A moral judgment, when combined with other beliefs and *only beliefs* of the agent, can motivate to act.

Even though this is not how internalism is standardly stated, and it is quite surprising why internalists normally forget to mention this important condition, there are actually some internalists who in fact do mention the extra condition. For instance, David McNaughton in his discussion of the "inconsistent triad" of Humeanism, internalism and cognitivism stated it thus:

> A moral opinion, when combined with other beliefs of the agent, can motivate him by providing him with reason to act. (McNaughton, 1988, 23)

Another philosopher who also mentions the belief-condition is Neil Sinhababu. Even though Sinhababu mostly refers to internalism as the rather vague claim that "the internalist view [is] that moral judgments alone can motivate us" or "moral judgments can produce their own motivational force"(Sinhababu, 2017, 66/5), he on a single occasion also mentions the belief-condition. He says:

> All internalists need is that moral judgments under normal psychological conditions could themselves produce motivation to act, *when combined with means–end beliefs.* So if I judge that eating meat is wrong, believe that eating this taco will be eating meat, and am under normal psychological conditions, those mental states will provide some motivation not to eat this taco. (Sinhababu, 2017, 7, emphasis added)[13]

Why should we think internalism must include a belief-condition? It must be noted that McNaughton as well as Sinhababu come from a strongly Humean background and so are clearly biased towards Humeanism. Internalism, however, is a self-standing philosophical claim and most internalists, even Humean internalists, seem to feel no pressure to add a belief-condition.

---

[13]I want to note two strange things about Sinhababu's quote. First, it sounds weird to say that moral judgments motivate "by themselves" and immediately add "when combined". This is like saying "Look mom, I can tie my shoes completely by myself, at least when you help me with it". Second, Sinhababu is a declared Humean, a cognitivist and an externalist. So, in his view, moral judgments are beliefs only, and beliefs can motivate only in combination with desires. So for a moral judgment to motivate it must be combined with an external *desire*. His example, however, suggests that moral judgments, which according to his cognitivism are beliefs, when combined with a suitable means-end belief can also provide motivation. How is this possible, given that he is a Humean who claims that motivation requires a desire? Where in Sinhababu's example does the necessary desire come from? I want to mention a further subtle difference between what typical internalists say and what Sinhababu says. Typical internalists claim that when someone judges that x is wrong, that person is motivated to refrain from doing x. Sinhababu's example is slightly different, however. Due to the fact that there are many different suitable means-end-beliefs, combining them with a moral judgment might motivate to all kinds of actions where it is not obvious that agents holding the judgment are motivated to act 'according to the judgment'. So it might be that Sinhababu's example points to a different phenomenon and not merely an instantiation instance of internal motivation.

The belief-condition is not normally present in contemporary discussions of motivational internalism, and in its adding a belief-condition to internalism philosophers like McNaughton and Sinhababu seem to explicitly take care of the Humean claim that motivation requires a desire as well as a *belief*. By adding a belief-condition they rule out the possibility that a missing belief could get in the way of a Humean explanation of internalism. I mention this implicit bias because I suspect that many noncognitivists after getting notice of the motivation problem which I raise here for them, will claim that the belief-enriched form of internalism (i. e. ABI), is what they always really had in mind, and that they simply forgot to mention this all to obvious condition.

They must, however, acknowledge the fact that the belief-condition is almost never explicitly mentioned in discussions of motivational internalism where the focus is on the truth of internalism alone as an interesting philosophical thesis, and not on whether it allows to tip the metaethical balance in favor or against cognitivism in combination with Humeanism. The belief-condition is mentioned only in some very rare cases where internalism is explicitly used as part of the motivation argument against cognitivism, such as in McNaughton and Sinhababu. When we normally confront people, students, friends, even other philosophers, with the thesis of motivational internalism for the first time, we do not point out to them that a very important extra condition that must be satisfied is that we must not only hold the moral judgment, but also certain other beliefs as well. Moreover, if we take a look at the recently published collection of papers on motivational internalism (Björnsson et al., 2015) we also nowhere find the belief-condition.

All this strongly suggests, that motivational internalism is not normally understood as containing such a belief condition. If internalism really must contain such a condition, then it has to be explained, how it could have been overlooked by so many philosophers defending as well as criticizing this important philosophical claim.

However, even if internalism is not normally understood as ABI, noncognitivists might still want to embrace it because it would help them to escape the motivation argument against noncognitivism. When we combine Humeanism with ABI, we can no longer derive that if one holds a moral judgment, one holds (a desire as well as) a belief which would be inconsistent with noncognitivism:

Additional-Belief-Internalism (ABI): If one holds a moral judgment and combines it with a suitable additional belief and only that belief, one is motivated.

Humeanism: If one is motivated, one holds (a desire and) a belief.

These two claims are compatible with noncognitivism, that is with the claim that moral judgments are desires only, because the belief necessary for motivation need not come from within the moral judgment itself, but can be identified with the additional belief.

Two further points are worth mentioning about ABI: First, noncognitivism together with Humeanism offers a good explanation of ABI: if moral judgments are desires, pairing them with a suitable additional belief is sufficient, according to Humeanism, to motivate to act. This explains the truth of ABI. Second, ABI together with Humeanism still entails the falsity of cognitivism, because we can derive that if someone holds a moral judgment with a suitable additional belief, then he will hold (a belief as well as) a desire. Since we have combined the moral judgment with an additional belief and nothing else, the desire must come from within the judgment itself. Both these points lend support to noncognitivism.

However, if we allow noncognitivists to escape the motivation argument against noncognitivism by modifying internalism, it seems that we must also allow cognitivists escape the motivation argument against cognitivism by making an analogous move by modifying internalism in similar way. For instance, cognitivists could propose to modify the view such that moral judgments motivate when *combined* with suitable additional desires besides the judgments. Call this form of internalism 'additional-desire-internalism' (short: ADI):

Additional-Desire-Internalism: If one holds a moral judgment and a suitable desire and only that desire, one is motivated,

and combine it with:

Humeanism: If one is motivated, one holds (a desire and) a belief.

From this, again, we cannot derive that if one holds a moral judgment, one holds a desire (as well as a belief) which would be as inconsistent with cognitivism as it would be with noncognitivism. In other words, Humeanism plus ADI is compatible with cognitivism, the claim that moral judgments are beliefs only. And again two further points are worth mentioning: First, cognitivism together with Humeanism offers a good explanation of ADI: if moral judgments are beliefs, pairing them with a suitable additional desire is sufficient, according to Humeanism, to motivate to act. This explains the truth of ADI. Second, ADI together with Humeanism still entails the falsity of noncognitivism, because we can derive that if someone holds a moral judgment with a suitable additional desire, then he will hold a belief (as well as a desire). Since we only combined a desire with the moral judgment, the belief must come from within the moral judgment itself. Both these points lend support to cognitivism.

So, we seem to have a dialectical stalemate. The situation between the noncognitivist who accepts ABI and the cognitivist who accepts ADI seems to be perfectly symetrical. Both ABI and ADI require an additional state besides the moral judgment in order to motivate agents. But why, then, should noncognitivists think that they are allowed to accept ABI (and thus avoid their motivation problem), whereas cognitivists are not allowed to accept ADI (and thus confront their motivation problem)?[14]

The crucial question is: is there any difference between ABI and ADI that could make it more legitimate to call one of them internalism, but not the other? Formulated differently: Is there a sense in which motivation can be said to be internal to moral judgment in ABI, but not be internal in ADI?

It might be claimed that motivation can be said to be internal to moral judgment when what comes from within the moral judgment is some *crucial* or *special* ingredient for motivation. What could this special bit be? An answer which I think many internalists would be inclined to give is that it provides a 'motivational state', has 'motivational force' or 'impetus', or is the "ultimate source of motivation".[15]

This seems to allow us to break the dialectical stalemate between the cognitivist who embraces ADI and the noncognitivist who embraces ABI. How? Consider ADI first. If the crucial bit for motivation coming from within moral judgment is supposed to be that it provides a motivational state or force, this cannot be what is special about moral judgment if we assume ADI. This is because ADI also provides a motivational state, namely an additional desire. Moreover, if we assume cognitivism, then the motivational force cannot come from within the moral judgment because it is a belief, and beliefs are not normally regarded a motivational states.[16]

What about ABI? If we assume that the motivational force comes from within the moral judgment, then ABI allows us to say that this is something special about moral judgment, because ABI only provides an additional belief, and beliefs are not motivational states. Moreover, if we assume noncognitivism and ABI, then it is clear how the crucial bit for motivation comes from within moral judgments. It comes from within the moral judgment because they are desire-like states, which are normally regarded as motivational states.

---

[14]I want to mention that even though ADI would help cognitivists avoid the motivation argument, cognitivists normally do not take internalism to be ADI. This is made clear by the fact that many cognitivists feel themselves forced to become Anti-Humeans because they believe that cognitivism plus internalism entails the falsity of Humeanism. Since, however, cognitivism paired with ADI instead of internalism does not entail the falsity of Humeanism, these philosopher must take internalism to be something other than ADI.

[15]Compare for instance (Sinhababu, 2017) and (Swartzer, 2015), but the terms 'motivational state' or 'force' are also used by many other philosophers in the debate.

[16]The latter point, of course, only holds against a Humean background. Anti-Humeans typically think that some beliefs, namely those with normative content, are also motivational states.

So, by appealing to motivational states or motivational force it seems we can break the symmetry between ADI and ABI in favor of ABI. ABI allows us, whereas ADI does not, to say that what is special about moral judgment is that what comes from within it, and is not provided by some additional state, is motivational force. This suggests that noncognitivists are more justified in appealing to ABI than cognitivist are in appealing to ADI.

There is a problem with this way of breaking the symmetry, however, crucially relies on whether we can make sense of what it is to be a motivational state or have motivational force. Can we make sense of this?

Why are desires regarded as motivational states? What is it for a state to be motivational or have motivational force? It cannot be that the state is necessary for motivation, since according to Humeanism, beliefs as well as desires, are both necessary for motivation. It can also not be that the state is sufficient for motivation, since neither desire nor belief by themselves is sufficient for motivation according to Humeanism. So, a desire's being motivational cannot consist in its being necessary or its being sufficient for motivation. And the same is true about beliefs.

Sinhababu is aware of this problem. Nevertheless he thinks that it makes sense to classify desires, but not beliefs, as motivational states. He writes:

> Like other Humeans, I describe desire and not belief as a motivational state. This isn't a matter of the sufficiency of desire for motivation, since desire must combine with a means–end belief. It's a matter of desire's *unidirectional* and *intention-constituting motivational effects*. All motivation is directed toward achieving a desired state, making it an intended state. But motivation to achieve or avoid believed means or ends, and how they figure in intention, depends on the *direction of desire*. If Olivia's desire to live combines with her belief that taking medicine makes survival more likely, it'll motivate her to take medicine, intending to live. If Ophelia's desire to die combines with the same belief, it'll motivate her to avoid taking medicine, intending to die. Believed means and ends can be objects of positive or negative motivation which we intend to achieve or avoid, while desire only makes its content the object of positive motivation and intention. This is why desires are often thought of as providing motivational force, which beliefs transfer from ends to means. (Sinhababu, 2017, 23f., emphasis added)

Sinhababu's explanation for why we are justified in calling desires motivational or having motivational force, but not beliefs, is unconvincing. His idea is this: combining a fixed belief with the desire to live or with the desire to not live (i.e. to die), leads to different motivations, positive or negative. Whether or not Ophelia will be motivated to take the medicine or will be

motivated to avoid taking it, depends on the 'direction' of her desire, when we hold fixed her means-end-belief. However, there would be an asymmetry in the 'motivational effects' between desire and belief, only if belief could not also change the 'polarity', so to speak, of motivation in a similar way, namely while also holding fixed some desire. But this is obviously possible. Suppose we hold fixed Ophelia's desire to live and combine it with the belief that taking medicine makes survival more likely. This will motivate her to take the medicine, intending to live. If we again combine the same desire to live with her belief that taking medicine makes survival not more likely, or even less likely, this will motivate her to avoid taking the medicine, intending to die. In this case the different directions of motivation obviously are not due to the different directions of desire, since we held the desire, and so its direction, fixed. Hence, the asymmetry between the motivational forces or effects between desire and belief which Sinhababu thinks to have spotted, does not exist. Consequently, the unidirectionality of desire cannot justify calling desires motivational, but not beliefs.

I take this to show that 'motivational' is a misleading term in the debate and should be avoided. Consequently, distinguishing between motivational states and non-motivational states can also not be used to break the symmetry between ABI and ADI, and so it can also not be used to break the dialectical stalemate between cognitivism and noncognitivism with respect to the relevant motivation arguments they face.

I conclude that there does not seem to be an interesting sense in which desires can be said to be 'motivational' or have 'motivational force' that does not have a parallel for beliefs. Hence, neither ABI nor ADI is better at capturing what it means for motivation to be internal to moral judgment. For this reason internalists are justified in sticking with my original formulation of internalism as the idea that moral judgments are intrinsically motivating in the sense that they can motivate *by themselves*, that is without the presence of any additional states besides the moral judgment itself. The most natural understanding of what it means for motivation to be 'internal' to moral judgment seems to be that holding a moral judgment is itself sufficient for motivation. This together with Humeanism, however, is precisely the claim which is incompatible with cognitivism as well as with noncognitivism. Hence, replacing internalism with some form of additional state internalism is not a promising way for noncognitivists to escape the motivation argument against them.

Motivational State Internalism

Another way in which noncognitivists might try to escape the motivation argument against noncognitivism by modifying internalism is the following. Sometimes philosophers formulate internalism as stating an internal connection not between moral judgment and *motivation*, but between moral

judgment and some *state* which is described as being 'motivational', like desire, intention or some other so-called pro-attitude. Jesse Steinberg, for instance, takes internalism to state a necessary connection between moral judgment and *desire*:

> Necessarily, if an agent makes a moral judgement, then she has some desire that favors, *inter alia*, any course of action that the judgement entails. (Steinberg, 2009, 2)

Similarly, Michael Ridge's favored version of internalism, which he affectionately calls 'Goldilocks internalism', states a necessary connection between moral judgment and the 'motivational' state of *intention*:

> Necessarily, for any agent A, If [sic] A is practically rational and makes a first personal judgment ... that A must ϕ in C, all things considered, then A's judgment is necessarily accompanied by an intention to ϕ in C. (Ridge, 2015, 141)

These formulations of internalism paired with Humeanism are compatible with noncognitivism: if moral judgments do not imply motivation, but imply only being in a state like desire or intention, then we cannot use Humeanism to infer, that if someone holds a moral judgment, he must hold a belief. Hence, noncognitivists who embrace a version of internalism along Steinberg's or Ridge's lines escape the motivation argument. Moreover, these forms of internalism are not only compatible with noncognitivism, rather each form noncognitivism implies some such form of internalism: if moral judgments are 'motivational' states like desires or intentions, one cannot hold a moral judgment, and lack the desire or intention.

Some authors, such as Caj Strandberg, think that this leads to another problem for the noncognitivists, namely that they are unable to explain akrasia (Strandberg, 2012). These authors think that such forms of internalism are problematic for noncognitivists because they make the connection between moral judgment and motivation 'too tight', in that it rules out the possibility that agents *fail* to be motivated, for instance in cases of severe depression, weakness of will etc. The alleged problem for the noncognitivist is that no matter how irrational or abnormal an agent is, she cannot fail to be motivated if she holds a moral judgment, because if holding a moral judgment is to hold a desire or motivational state, and desires motivate, then it is simply *impossible* to hold a moral judgment and fail to be motivated. In other words, these authors think that such forms of noncognitivism imply or get 'for free' motivational internalism. Teemu Toppinen explains the reasoning behind this nicely. He says that, according to noncognitivism, "if one judges that ϕ-ing would be desirable, one desires to ϕ. It is also often suggested that to desire to ϕ is to be motivated to ϕ... If this is so, then [noncognitivists] must accept *unconditional practicality* [i. e. what I here

call motivational internalism]" (Toppinen, 2015, 153, italics original). The standard response to this problem, as Toppinen notes, "is to distinguish between *desire* and *motivation*, and to suggest that normative judgments are desire-like states which are *normally* motivational but may fail to motivate... when the conditions are abnormal—as they may be in the case of severe depression" (Toppinen, 2015, 154). It should be clear that Humeans think this reasoning contains a mistaken assumption, namely that desires alone can motivate to do something. So, I think that the 'too tightness'-worry misses its mark.

But motivational-state-formulations of internalism faces at least two problems. First, if internalism is the thesis that moral judgments imply intentions, say, then how can a noncognitivist explain this internalism if her theory is in terms of being for for instance? The second problem is this that internalists have to decide *which* type of state they think is implied by moral judgment. For instance, Steinberg thinks it is desire, whereas Ridge thinks it is intention. It is natural to expect that for Blackburn it will be some sort of 'hoorying'-approval state, for Gibbard a state of norm acceptance, and for Schroeder a state of being for, and so on. This has a strong flavor of arbitrariness to it and it is not clear why *non*-noncognitivist internalists should accept such specific forms of internalism.

To avoid this arbitrariness noncognitivists might suggest that we should think of moral judgment not as being connected to some *specific* motivational state, but only to being in *some* motivational state. So they might hold the following version of internalism:

Motivational State Internalism: Necessarily, if one holds a moral judgment, then in virtue of doing so one is in some motivational state.

Just like internalism in terms of desire, motivational state internalism is compatible with Humeanism and noncognitivism; indeed, every form of noncognitivism implies motivational state internalism.

But there is still a problem. The problem with motivational state internalism is that, as I have already explained, there is no interesting sense in which desires can be said to be 'motivational states' that does not equally apply to beliefs. If it makes sense to speak of some beliefs (e. g. moral beliefs) as motivational states, as many cognitivists think, then this form of internalism at least does not favor noncognitivism over cognitivism when it comes to explaining motivational state internalism.

Moreover, originally, motivational internalism is called 'motivational' because it is concerned with a relation between moral judgment and motivation. This relation, however, is lost in motivational state internalism. For these reasons I am skeptical that internalists would really be willing to replace what I take to be their original understanding on internalism with motivational state internalism.

Motivational Necessitism

Another idea for noncognitivists to deal with the motivation argument is to get away from the idea that motivation is in some sense 'internal' to moral judgment, and instead adopt the weaker claim that motivation is necessitated by moral judgment, or as Ridge said above, that moral judgment is "necessarily accompanied" by motivation. This proposal has already been made by several others.[17] For instance, Sigrún Svavarsdóttir proposes that even though the label 'internalism' suggests that moral judgments motivate intrinsically or on their own, we should adopt the *weaker* thesis that moral judgments are necessarily motivating:

> Although the label 'internalism' suggests this strong version, it is more in line with the existing literature to formulate the thesis as claiming that moral judgments are necessarily connected to motivation... [this claim is weaker because] the motivation imported by a moral judgment has its roots in a distinctive conative state that is, however, necessarily connected to the moral judgment. (Svavarsdottir, 1999, 163)

We can reformulate this proposal as follows:

Motivational Necessitism: necessarily, if someone holds a moral judgment, then she is motivated to act.

Motivational necessitism is weaker than the above discussed forms of motivational internalism since it does not require that the motivation comes from 'within' the moral judgment itself, that moral judgments by themselves have 'motivational force' or are the ultimate 'source' of moral motivation. It does not state an internal, but only a necessary connection between moral judgment and being motivated.[18]

If noncognitivism and Humeanism were true, how could motivational necessitism be true? For this to be the case it would have to be true that whenever someone holds a moral judgment he additionally holds a suitable means-end-belief.

How could it be that a suitable belief is necessarily accompanied with every moral judgment? The simplest case would be that everyone necessarily holds the suitable beliefs irrespective of whether he holds the moral judgment

---

[17]See, for instance, Connie Rosati who says: "Traditionally, judgment internalism has been characterized as claiming either that motivation is internal to moral judgment, in the sense that moral judgment *itself* motivates without need of an accompanying desire ('strong internalism') or that there is a necessary connection between moral judgment and motivation 'weak internalism')" (Rosati, 2016, Ch. 3.2, italics added).

[18]I share Svavarsdóttir's view that the 'intrinsically'-reading of internalism is stronger than the 'necessarily'-reading, though others, such as Russ Shafer-Landau and Daniel Eggers, seem to think the opposite. Cf. (Shafer-Landau, 2003, 147)(Eggers, 2015, 2454).

or not. How could this be? One idea could be that the additional belief always is a trivial belief, and that everyone necessarily holds every trivial belief of the form that ϕ-ing is a means to one's ϕ-ing.[19] But how plausible is it that everyone always holds all trivial beliefs? I think it is just as unrealistic as the view that everyone always believes every tautology.

This issue could, however, perhaps be solved by conditionalizing motivational necessitism to rationality. If we assume that every agent that holds a moral judgment is motivated to act, if he is (ideally) rational, then we might also assume that one is rational only if one holds every trivial belief. Since the rational agent holds every trivial belief, he also holds the trivial belief which is suitable to motivate the agent when combined with the desire which according to noncognitivism is his moral judgment. This guarantees that for every moral judgment a rational agent holds, she also holds a suitable trivial belief in virtue of the fact that she is rational. This could explain why, necessarily, an agent who holds a moral judgment is motivated to act *if* he is rational.

It is, however, a controversial open question in the philosophy of rationality whether it is really required of rational agents that they hold all trivial beliefs.[20] So I will not pursue this option further, since we have already move quite a bit away from the usual formulation of motivational internalism.

8.3.2   Modifying Humeanism

In the previous section I critically discussed several ways to modify internalism so that noncognitivists can escape the motivation argument. In this section I want to discuss ways to modify Humeanism to escape the argument.

It has become clear from the previous discussion that one claim is particularly troublesome for noncognitivism. This is the often ignored Humean claim that desires without belief do not motivate (see Corollary-2 above). So suppose we drop this claim and replace Humeanism with the following milder form:

Mild Humeanism: One is motivated iff one holds (a desire and a belief) or a desire.

If we combine Mild Humeanism with

Internalism: If one holds a moral judgment, one is motivated.

we can derive that if one holds a moral judgment, then one holds (a desire and a belief) or a desire. Hence, Mild Humeanism paired with internalism

---

[19]For the notion of 'trivial' beliefs see (Sinhababu, 2017).

[20]It has been argued, for instance by Harman, that rationality cannot plausibly require of agents that they belief all tautologies since this would result in 'cluttering ones mind' which is to be avoided (Harman, 1986). Similarly it might be demanding to much of rational agents that they hold all trivial beliefs.

is compatible with noncognitivism (as well as with hybridism). Moreover, noncognitivism and Mild Humeanism together entail internalism. From:

Noncognitivism: If one holds a moral judgment, then one holds a desire,

Mild Humeanism: If one holds (a desire and a belief) or a desire, one is motivated.

we can derive:

Internalism: If one holds a moral judgment, one is motivated.

This speaks in favor of noncognitivism. So mild Humeanism seems to be an attractive option for noncognitivists.

It is well-know that cognitivists make a parallel move to escape the motivation argument against cognitivism by moving from Humeanism to Anti-Humeanism. They pair the following two:

Anti-Humeanism: One is motivated iff one holds (a desire and a belief) or a suitable belief.

Internalism: If one holds a moral judgment, one is motivated.

This is compatible with cognitivism. Moreover from:

Cognitivism: If one holds moral judgment, one holds a belief (and no desire), and

Anti-Humeanism: If one is motivated, one holds (a desire and a belief) or a suitable belief,

we can derive

Internalism: If one holds a moral judgment, one is motivated.

This speaks in favor of cognitivism. So Anti-Humeanism seems to be an attractive option for cognitivists.

The crucial question is: How plausible is Mild-Humeanism? Some philosophers seem to accept the idea that desires can motivate by themselves. For instance, Jesse Steinberg thinks that "desires are necessarily motivating" (Steinberg, 2009, 4). Steven Swartzer says that the assumption that "even *desires* could not be motivational states [is] an assumption that is incompatible with the truth of the Humean theory" (Swartzer, 2015, 2, italics original).

Most Humeans, however, normally reject the idea that desires can motivate by themselves. For instance, Jonathan Dancy says that the following:

if you want an orange but have no relevant beliefs - beliefs about the probable whereabouts of oranges, for instance - your desire is *impotent*; it is like a *blind urge seeking a guide*, an external agent to give it a direction. So each state needs the contribution of the other. A complete motivating state - a state which is sufficient for action - must be a combination of belief and desire. (Dancy, 1993, 1f., emphasis added)

Similarly, David McNaughton takes Humeanism to entail that:

desires without beliefs are blind, beliefs without desire are inert. (McNaughton, 1988, 21)

Likewise, for Alexander Miller, according to which Humeans hold that:

beliefs cannot on their own motivate someone to act. And neither can desires. (Miller, 2003, 219)

Even though this is also the view of one of the leading Humeans, Neil Sinhababu, he is willing to allow that at least for some *specific* desires Humeans might perhaps make an exception:

Beliefs might not seem necessary for motivation in cases where one desires to produce an *immediate bodily movement* like moving one's hand. Does this require the trivial belief that by moving one's hand, one will make it more likely that one moves one's hand? The Humean Theory requires the trivial belief. If I desired to move my hand and denied the trivial truth that by moving my hand I'd make it more likely that I moved my hand, I probably wouldn't move my hand. Perhaps I'd feel disappointed or resigned to my perceived inability. This is why the Desire–Belief Theory of Action requires belief. Still, the view that beliefs aren't necessary in these trivial cases remains a plausible Humean position. (Sinhababu, 2017, 3)

So according to Sinhababu, Humeans might allow that at least desires about immediate bodily movements do not need an additional belief in order to motivate to move one's hand. But we must be aware that the desires which, according to noncognitivists, constitute moral judgments arguably are not like the desire to move one's hand, but are normally more complicated and not about immediate bodily movements. Hence, Sinhababu's extension of Humeanism will arguably not cover the motivational force of moral judgments.

The crucial question for this way of escaping the motivation argument depends on the overall plausibility of Mild-Humeanism. I personally do not find it totally unappealing to think of every desire as giving rise to motivation,

but I must admit that for some more complex desires it seems that having them demands a belief as 'an external agent to give it a direction' to motivate at all. I cannot investigate Mild-Humeanism further here, and so, I have to leave it as an open question how plausible this way of solving the motivation problem is.

8.3.3 Modifying Noncognitivism

A final way I want to discuss to escape the motivation problem would be to modify noncognitivism. Noncognitivists might say that the belief required for being motivated by a moral judgment is already part of the moral judgment itself. This is incompatible with pure noncognitivism, of course, according to which moral judgments are desires. But perhaps noncognitivists could go 'mildly hybrid' without losing the original spirit of noncognitivism. How could this go?

The idea would be that moral judgments are *pairs* of desires and *trivial* beliefs. For instance, noncognitivists could claim that my judgment that murder is wrong is the pair consisting of my desire not to murder, and the trivial belief that my not murdering is a means to my not murdering.[21] This would obviously guarantee that whenever someone holds a moral judgment he will necessarily hold the belief necessary for motivation. This mildly hybrid form of noncognitivism would also allow noncognitivists to easily make sense of how motivation can be internal to moral judgment because according to it everything necessary and sufficient for motivation would be provided solely and completely by the moral judgment itself.

On the face of it going mildly hybrid seems to be a possible option for the noncognitivist in order to deal with the motivation argument. To my knowledge, no one has ever suggested that noncognitivism must be understood along the lines of what I have called mildly hybrid noncognitivism. Anyone who would want to propose such a form of noncognitivism in order so escape the motivation argument would at least have to explain away the seeming adhocness of the assumption that to hold moral judgments are in part constituted by trivial beliefs. Due to reasons of space, I shall, however, not investigate what kinds of issues such a revision of noncognitivism might give

---

[21]The view would only be 'mildly' hybrid, but not full-blown hybrid, because the beliefs that partly constitute the mildly hybrid state do not provide the truth-value for this state as they do in hybrid theory. Trivial beliefs are tautological or analytically true, and therefore always true. But moral judgments are not always true. So the truth-value of a moral judgment cannot, on the noncognitivist's mildly hybrid view, be inherited from the truth-value of the trivial belief that partly constitutes the moral judgment. It is crucial for the hybrid theories developed by Ridge (his 'ecumenical expressivism') and others that the truth-value of the hybrid state is inherited from the belief that partly constitutes it, in order to solve the Frege-Geach problem. Unlike on the hybrid theorist's account of logicality, the mildly hybrid noncognitivist cannot explain the logical properties of sentences by piggybacking on the truth-value of the belief that partly constitutes the moral judgment.

rise to. So again I leave it as an open question whether this option of dealing with the motivation argument against noncognitivism is more plausible that the cognitivist Anti-Humeanism that deals with the motivation argument against cognitivism by claiming that that some beliefs can motivate without the assistance of an external desire.

It is time for a quick summary. So far in this chapter, I have presented a problem for noncognitivists having to do with motivation. I argued that internalism together with Humeanism entail the falsity of noncognitivism. In a nutshell the argument is that motivation requires belief as well as desire, that moral judgments motivate by themselves, and so moral judgments cannot be desires only (just as they cannot be beliefs only). I then presented and discussed several modifications of internalism, Humeanism and noncognitivism as ways to escape the motivation argument against noncognitivism. Even though some of the proposed modifications might perhaps be viable options for the noncognitivists, it is clear that there are open questions to adress for noncognitivists and that they cannot simply assume that they get internalism 'for free.' The noncognitivists' problem with explaining internalism might not be unsolvable, but the discussion shows that, contrary to what is generally assumed, it is far from obvious that noncognitivists really have an easier time explaining how moral judgments can motivate agents to act accordingly.

8.4   Second Motivation Problem for Noncognitivism

In the rest of this chapter, I present a further unnoticed problem for noncognitivism. The problem is simple: intuitively, pairing a moral means-end belief with a suitable desire motivates agents to act. According to noncognitivism, however, moral beliefs are themselves noncognitive desire-like states. This raises the question: how can noncognitivists explain that a noncognitive desire-like state paired with a desire motivates agents to act? I call the challenge of answering this question the 'motivation problem'.

A Humean solution to the motivation problem is unlikely to be successful because Humeans think that motivation always requires the presence of a noncognitive desire state as well as a *cognitive* belief state. So noncognitivists will have to give an Anti-Humean solution and explain, in a systematic way, how combing two noncognitive desire states can motivate agents to act. I explain why this is a difficult, but perhaps not impossible task. I sketch how my own dispositionalist version of noncognitivism might be able to solve the problem.

In the previous half of this chapter, I presented an unnoticed problem for noncognitivists having to do with explaining moral motivation, especially so-called motivational judgment internalism. The discussion of the problem suggested that, contrary to what is generally assumed, noncognitivism is

not better than, but at most equally good (or bad) at explaining moral motivation as cognitivism.

In the remaining half of this chapter, my aim is to go one step further than this and suggest that noncognitivism is actually *worse* at explaining moral motivation than cognitivism. I do so by showing that noncognitivism runs into a further motivation problem which cognitivists do not have. To my knowledge the problem has so far been unnoticed in the metaethical literature. The problem is to explain how combining a moral means-end belief with a suitable desire can motivate agents to act. I call this problem the *motivation problem*.

For cognitivists who hold that moral beliefs are ordinary cognitive belief states the solution to this problem is pretty straight forward, especially when they assume the Humean theory of motivation (short: Humeanism). According to Humeanism having a suitably related pair of belief and desire is necessary as well as sufficient for motivation. Cognitivism paired with Humeanism's sufficiency claim provide a simple solution to the motivation problem: according to cognitivism, moral means-end beliefs just are cognitive belief states. It then follows from Humeanism that pairing such a state with a suitable desire generates a motivation.

For noncognitivists the motivation problem is much more difficult since they hold that moral beliefs are themselves desire-like noncognitive mental states, and not cognitive belief states. This means that noncognitivists must hold than when agents combine a moral means-end belief with a suitable desire, they actually combine a desire with another desire-like noncognitive state. It therefore seems that noncognitivists cannot use Humeanism in order to explain how combining those states motivates agents to act. On the contrary, if noncognitivists accept that moral beliefs combined with suitable desires motivate agents to act, noncognitivists seem forced to reject Humeanism because according to Humeanism motivation always requires the presence of a desire-like noncognitive state as well as a cognitive belief state. Though I shall investigate several options for noncognitivists to remain Humeans in a moment, it turns out that they are all unpromising, and so noncognitivists should really reject Humeanism.

The problem with rejecting Humeanism is that it leaves the noncognitivist without a theory of motivation. Noncognitivists are thus confronted with the task of giving a genuinely noncognitivist solution to the motivation problem which I understand as the challenge of explaining how a combination of a desire with a desire-like noncognitive state that constitutes a moral means-end belief can motivate agents to act.

Even though my general claim in this chapter is that noncognitivism overall is worse at explaining moral motivation than cognitivism, my claim is not that noncognitivism is completely unable to explain it. In fact, I shall argue, that my own dispositional form of noncognitivism, as developed before in section 5.4, is capable of providing a solution to the motivation

problem. But since, as will become apparent, a solution to the motivation problem is more difficult for noncognitivists than it is for cognitivists, it can plausibly be inferred that noncognitivism is generally worse at explaining moral motivation than cognitivism[22]

A comment on terminology in this section. In the previous section I used 'moral judgment' instead of 'moral belief' because this was more in line with general use in debates about motivational judgment internalism. Moreover, I wanted to reserve the terms 'belief' and 'desire' for the states that figure in Humeanism. I made this terminological choice in order to avoid confusion, and I explained that nothing substantial hangs on this choice. On the other hand, I also said that contemporary noncognitivists are quasi-realist in that they think they can 'earn the right' to regard moral judgments (i.e. the states, whatever they are, expressed by moral sentences) as moral beliefs. Since the following sections are not about motivational judgment internalism, but about noncognitivism and Humeanism I will in this section switch from 'moral judgment' to 'moral belief'.

8.4.1 The Motivation Problem

Consider a simple case. If I want my wife to always love me, and I believe that she'll do, if I'm a good man, then this motivates me to be a good man. At least, it motivates so normally and to a certain degree. We are all familiar with similar examples from our everyday life. Having desires and having beliefs about how to get what we desire, motivates us to do what we believe will help us to get it.

Intuitively, in this case the agent is motivated to be a good man. But it is not only intuitively so, it also is predicted by Humeanism which tells us that an agent is motivated to act if he has a desire combined with a suitable means-end-belief. According to Humeanism, this belief and that desire motivate an agent to be a good man, because desires and beliefs play different roles in the generation of motivation: our desires set goals, and our beliefs tell us what we need to do in order to realize our goals. This simple explanation can be stated by using other terminology than 'belief' and 'desire', for instance by invoking different directions of fit, speaking of representational and conative states, cognitive and noncognitive states, or different functional roles etc. But no matter how we formulate Humeanism the basic idea remains the same. For instance, instead of formulating it in terms of belief and desire, Humeans may say that there are states that represent ways the world is to be, and states that represent ways of how the world is. By representing ways the world might be the latter states provide us information about what needs to be done in order for the world to become the way states of the former type represent how it is to be. No matter which

---

[22]Especially Humean externalist versions of cognitivism seem to fare better in this respect.

terminology we use to explain motivation, what is crucial to Humeanism is that two kinds of mental states are needed to generate a motivation.

Cognitivists can and do accept Humean explanations along those lines because they hold that all beliefs, descriptive as well as moral ones, are in the business of providing information about how the world is. But noncognitivists have a problem here. The problem is that according to noncognitivists not all beliefs represent ways the world is, have a mind-to-world direction of fit, or are cognitive states etc. Rather some beliefs resemble desires in that they represent ways the world is to be, have a world-to-mind direction of fit, are noncognitive states etc. According to noncognitivism, the means-end belief in my good man-example is such a belief. This leads to the following question: how, if noncognitivism about moral belief is true, that is if moral beliefs are noncognitive desire-like states, can it be that combining such beliefs with desires can motivate us to do something? Answering this question is the *motivation problem* for noncognitivism.

Let us organize the above example as follows:

Belief: that if I'm a good man, my wife will always love me.

Desire: that my wife will always love me

Motivation: to be a good man.

There are a couple features to note about this example which I want to record as data. An agent can believe that if he's a good man, his wife will always love him, but at the same time *lack* the desire that his wife will always love him. More generally even, an agent can hold that belief but lack any desire whatsoever and *vice versa*. Many philosophers accept something even stronger, namely that for any belief whatsoever it is possible to hold it and lack any desire whatsoever. And *mutatis mutandis* for desire. Humeans, as we already know, often make this claim by saying that belief and desire are 'distinct existences'.

Another thing to note is that an agent can hold the belief and be *not motivated* to be a good man. More generally even, an agent can hold this belief and not be motivated to do anything at all.[23]

Moreover, an agent can hold the belief and be motivated *not* to be a good man. It is likely that he will be so motivated, for instance, if he desires to not to be always loved by his wife or if he is independently motivated not to be a good man (by some other belief and desire pair for instance).

---

[23]Note that if internalists are right, we cannot make the more general claim that for *any* belief it is possible to hold it and not be motivated at all. I do not want to rule out the possibility of internalism here, so I shall not assume that belief and motivation are 'distinct existences'. Nevertheless, even the internalists will agree, I think, that the above means-end-belief does not imply any motivation by itself, in fact, no means-end-belief does imply any motivation by itself.

Since an agent can hold the above belief, in fact any means-end-belief and either be motivated to do something, motivated to not do it or not be motivated at all, this suggests that what an agent is motivated to do depends not only on his beliefs, but also on his desires, and on how the belief and the desire are related. It is something about the *combination* of the belief and the desire that motivates an agent to be a good man. Of course, this need not be the only combination of belief and desire that motivates agents to be good men, but it shows that an explanation of why an agent with this particular combination of belief and desire is motivated to be a good man must mention their specific *interaction*. That an agent holding this belief and that desire is motivated to be a good man, is not because he holds this belief *or* because he holds that desire, but because he holds the belief in combination with the desire.

That those two states together motivate to be a good man is surely partly due to general features of beliefs (e. g. that they represent, aim at truth etc), and surely also partly due to general features of desires (e. g. that they set goals, aim at satisfaction etc). But it cannot be only due to the *general* features of beliefs and desires that an agent who holds them is motivated to be a good man, because not every combination of a belief and a desire motivate agents to be a good men. That an agent who holds the above pair of belief and desire has this specific motivation is also partly due to their being this *specific* belief and that *specific* desire. So, that an agent who holds this specific pair of belief and desire is motivated to be a good man, is because it is a pair of belief and desire, and because of the 'content' of this particular belief and the 'content' of this particular desire and the way they are related in virtue of their having this particular 'contents'.[24]

The main points sofar are that neither the above belief nor the above desire taken by themselves motivate an agent to be a good man. It is the combination of this specific belief and that specific desire and the interplay of these specific states, that is the interplay of the type of states as well as their 'contents' or specific roles, that generates the motivation.

All of the forgoing points can be regarded as constraints on an adequate explanation of how holding a belief-desire-pair motivates agents. If an explanation violates one of the constraints, in the sense of being unfaithful to the data, it cannot be regarded as fully adequate. For instance, if an explanation of why an agent who holds some belief and some desire has a specific motivation would be explained by appeal only to general features of belief, or features of some specific belief, would not be fully adequate because it disrespects the data that both states are relevant for generating motivation. Likewise, an explanation that does not appeal to the interaction between the belief and desire and their contents is also not fully adequate.

---

[24]I am careful with the notion of content here because much of the difficulty for noncognitivists to explain moral motivation is precisely that they reject that there are moral propositions (at least in a non-minimal sense) that could figure as the objects of moral beliefs.

8.4.2 Remain Humean?

As explained above, an important aspect of the motivation problem is that if noncognitivism is true, then cases like the good man case conflict with Humeanism. According to noncognitivism the belief that if I'm a good man, then my wife will always love me, must be analyzed as not being a cognitive state. Given that this noncognitive state combined with my desire to always be loved by my wife motivates me to be a good man, this raises the following question: How can it be that in this case we think that I am motivated to be a good man if noncognitivism is true?

From our previous discussion we already know, if the noncognitive state and the desire are really all that is involved in cases like this one, then Humeans will answer: *not at all.* According to Humeanism, a pair of a noncognitive state and a desire cannot motivate, because motivation requires belief. In other words, Humeans will treat the good man-case as a refutation of noncognitivism. But is there really no way for noncognitivists to accept that in the good man case the agent has the relevant motivation and yet remain Humean? I think no, but let us consider some possible options.

Perhaps the noncognitivist can stay true to his Humean convictions by claiming that the noncognitive state and the desire can motivate, if the belief which according to Humeanism is required for motivation is nevertheless somehow present. So it might be claimed that the required belief is either (1) involved in or implied by the noncognitive state or the desire, or (2) might simply be added as an extra to the noncognitive state and the desire. However, neither option seems plausible. I explain why now.

One obvious problem with the first option is that it contradicts pure forms of noncognitivism according to which moral beliefs are desire-like states and nothing else, in particular not beliefs. Quasi-realist noncognitivists may reply that they do not hold that moral beliefs are not beliefs, because, as I have explained in chapter 5, they take pains to argue that moral beliefs can be regarded as beliefs, despite the fact that they are at bottom noncognitive desire-like states. However, even if noncognitivists are justified in calling the noncognitive states that constitute the moral beliefs 'beliefs', this by itself does not solve the motivation problem because quasi-beliefs cannot bear explanatory weight. I will say more on this in my discussion of Beddor's case below where I explain why the mere fact that some state can be regarded as a quasi-belief is not enough to satisfy the Humean conditions for motivation.[25]

Confronted with this problem noncognitivists might want to stop being pure noncognitivists and go mildly hybrid. Recall that by going mildly hybrid I above meant that noncognitivists may allow that moral beliefs are not exclusively noncognitive states, but *pairs* of noncognitive states and

---

[25]One way to conceive of the issue is that noncognitivists will not be able to legitiamtely claim to have 'earned the right' to moral belief unless they have previously shown that they can solve the motivation problem.

'trivial' beliefs, where a trivial belief is a belief of the form that doing p is a means of doing p, or that doing p does not lower the probability of doing p, or that if one does p, then one does p.[26] The idea should be clear. If we accept a mildly hybrid form of noncognitivism, then the belief Humeans require for motivation, could be the trivial one involved in holding the moral means-end-belief, since these are pairs of noncognitive states and trivial beliefs. However, even if the noncognitivist treats the noncognitive state constituting the moral means-end belief that if I am a good man, then my wife will always love me as being partly constituted by some trivial belief, then this will still not help with the problem. The reason why this does not help is that, arguably, no trivial belief we can imagine paired with the desire to always be loved by one's wife can generate the moral motivation to be a good man.

Let me try to illustrate this at the example of a Blackburnian form of noncognitivism.[27] Let us suppose that a mildly hybrid Blackburnian form of noncognitivism analyses the belief that if I am a good man, then my wife will always love me as the pair of consisting of (i) the higher-order state of disapproving of combining approving of being some way as a man with not believing that my wife will always love me[28] and (ii) the trivial belief that if my wife will always love me, then my wife will always love me. Even though this would guarantee that an agent who holds the moral means-end belief *ipso facto* also is in some belief state, it remains totally unclear how combining this hybrid state with the desire to always be loved by my wife could motivate agents to be good men. For this reason I am highly skeptical that going mildly hybrid would really help the noncognitivist with solving the motivation problem.

What if the noncognitivist assumes that the belief required by Humeanism is not involved in (in the sense of being a part of), but only *implied* by the noncognitive state or desire? It must be noted that any such proposal contradicts the Humean dictum that beliefs and desires are 'distinct': no desire involves or implies a belief. However, it seems logically possible to drop the distinctness claim and still accept Humeanism. Humeanism only says that belief and desire are necessary and sufficient for motivation, not that beliefs and desires must be categorically distinct states.

How might one drop the distinctness claim? Noncognitivists might take inspiration from three kinds of philosophers who all are anti-Humean in the sense that they reject the distinctness claim. Those kinds of philosophers are (a) those who treat desires *as* beliefs about what is desirable or good, and (b) those who hold that one cannot desire what one believes (or knows)

---

[26]I take the concept of a trivial belief from (Sinhababu, 2017).

[27]Blackburn's major views on noncognitivism can be found in his (Blackburn, 1984b), (Blackburn, 1993), and (Blackburn, 1998).

[28]For a nice presentation of Blackburnian noncognitivism along those lines see (Miller, 2003, 60f.).

to be impossible or actual, and (c) those who hold that desires (or at least intentions) imply beliefs about what one will do.[29] These three views are highly controversial. Let us nonetheless consider them briefly in order to see if they could help the noncognitivists' with remaining Humean and solving the motivation problem.

The first idea (a) is an Anti-Humean theory of desires which David Lewis has dubbed the 'desire-as-belief' theory.[30] He himself argued that this theory "collapses into contradiction" (Lewis, 1996, 308) because he takes it to be inconsistent with the Bayesian requirement that an agent's desire for a proposition should not change upon learning the proposition is true.[31] But even if we ignore Lewis' arguments against the desire-as-belief theory, the idea that desire is belief about what is desirable cannot help the noncognitivists with the motivation problem because it is incompatible with noncognitivism.

Why are they incompatible? A central motivation for noncognitivism is that we can avoid having to postulate metaphysically 'queer' moral entities in our ontology and that we can provide a naturalistic account of morality if we can analyse belief with moral contents as desire-like states with non-moral contents. The desire-as-belief view, however, has the opposite direction of explanation: it explains desire in terms of beliefs with moral contents. For this reason, noncognitivism and this form of Anti-Humeanism do not fit together. Hence, embracing a desire-as-belief theory is not an option for noncognitivists to remain Humeans and yet deal with the motivation problem.

The second kind of philosopher holds that (b) one cannot desire what one believes to be impossible or actual. Noncognitivists might want to embrace this sort of view with the hope that the belief implied by the desire could help explaining the relevant moral motivation. But there is a problem. According to (b), desires do not imply certain beliefs, rather they imply the *lack* of

---

[29]All those mentioned options contradict the view that beliefs and desires are distinct. A weaker option that is compatible with distinctness would be to say that desires do not involve or imply certain beliefs, but only rationally require certain beliefs. This might allow noncognitivists to explain how the good man-case motivates. If the noncognitive state that constitutes the means-end-belief rationally requires a certain cognitive belief state, then it might be claimed that at least if an agent *is rational*, this belief state plus his noncognitive state, plus the desire state somehow generate the relevant motivation. I shall not investigate this rationality-option here, because the motivation problem, as I understand it here, is not a problem concerning rational motivation or the rationality of instrumental reasoning, but concerns motivation generally. Moreover, in order to evaluate the prospects of such a rationality-proposal we would still have to know what the relevant noncognitive state is that makes up the means-end-belief, which ordinary cognitive belief states it requires agents to be in, and most importantly, how all those states interact to generate the motivation to be a good man.

[30]Cf. (Lewis, 1988) and (Lewis, 1996).

[31]For objections to Lewis' rather technical 'updating argument' against this anti-Humean theory, see (Bradley & List, 2009) and (Campbell, 2017).

certain beliefs. In other words: if one desires something, one does *not* believe that it is already actual. So option (b) does not provide noncognitivists with a belief which according to Humeanism is required for motivation.

In reply to this noncognitivists might assume that the following also holds: if one desires something, one *believes* that this something is *not* yet actual. This tells us that desire not merely implies the lack of a belief, but also a belief.[32] Suppose this assumption is true, and we apply it to the desire in the good man case from above. Then the above desire that my wife will always love me implies my belief that it is not already the case that she will always love me. Does this help the noncognitivist with the motivation problem?

I do not think so. The belief that it is not already the case that my wife will always love me paired with my original desire that she will always love me by themselves do not motivate me to be a good man. This is not surprising, of course, because the noncognitive state making up the means-end-belief also must play a role in the generation of that motivation. However, in order to find out whether the noncognitive state that makes up the means-end-belief together with the desire and the belief implied by it is able to generate the relevant motivation, noncognitivists would have to tell us at least two things: namely, first, what the noncognitive state is, and, second, what the 'mechanics' of the interaction of this noncognitive state with the desire and the belief is.[33] Even if we would know the nature of the noncognitive state I do not see how we could explain the mechanics of the interaction of all the states involved here to generate the relevant motivation.

In contrast, the mechanics cognitivist Humeans can offer is simply this: belief that if p, then q, combined with desire that q, motivates that p. For the purposes of the cognitivist Humean this is detailed enough. For the noncognitivist who takes the belief that if p, then q, to be a noncognitive state with a possibly completely different structure, it is not clear how the mechanics works even if desires imply beliefs about what is not actual. For this reason, it is at least far from obvious that noncognitivists are in a position to answer whether combining a noncognitive state with a desire that implies a belief allows noncognitivists to explain how those states motivates an agent to be a good man.

The same holds for the above suggestion (c) that desires or intentions imply beliefs about what the agent will do.[34] Suppose that the desire or

---

[32]Noncognitivists might claim that this follows from (b) together with certain doxastic assumptions about believing, such as that believing that p is equivalent to not believing that not p.

[33]The picture can be complicated even further if we assume that the noncognitive state constituting the means-end belief also implies some (possibly rather complex) belief about what is not already actual.

[34]Whether this so-called 'cognitivism about intrumantal rationality' is true is highly controversial. For discussion see (Bratman, 1987, 30-41), (M. Schroeder, 2010b, 92-95), and (Broome, 2013, 164f.).

intention that my wife will always love me implies the belief that I will make it the case that my wife will always love me. Again, the same questions arise as before: what is the noncognitive state, and how does combining it with the desire and the belief implied by it (this time a belief about what one will do) generate the motivation to be a good man? I think that noncognitivists do not have a satisfying answer to this.

For the reasons just discussed it is implausible that the belief which according to Humeanism is necessary for motivation can somehow be involved in or implied by the noncognitive state or the desire. So, option (1) is not promising.

What about the second option (2) that the required belief in not involved or implied in the noncognitive state or the desire, but that we simply add it to them? Here, of course, we too have the problem that it is not clear what this belief would have to be in order to be of any help, and how it would interact with the non-cognitive state and the desire to motivate, especially since the motivation in which it results is to acquire a *moral* property. In addition to this problem, option (2) also has the further problem that it runs against the above data that the moral belief and the desire together are *all* that is needed to generate the motivation to be a good man.

To summarize: In this section I argued that it is unpromising for noncognitivists to assume that the belief which according to Humeanism is necessary for being motivated comes from the noncognitive state making up moral belief or the desire, and that adding an external belief conflicts with the intuitive data that nothing more is need to motivate the agent than his means-end-belief and the desire. This strongly suggests that the motivation problem forces noncognitivists to give up Humeanism.

Where does this leave us? It should be clear that giving up Humeanism does not bring noncognitivists closer to a solution to the motivation problem. On the contrary, it makes the problem harder, since noncognitivists now are in need of a plausible theory of motivation and have to come up with their own. Such a theory must enable noncognitivists to explain how the interplay or 'mechanics' of the state constituting the moral means-end belief and the desire can motivate an agent to be a good man. In the next section I will have a look at a proposal to give such an explanation which Bob Beddor, though in a different context, has recently suggested. Even though I think that Beddor's specific proposal is not really adequate, I think something like it is a promising direction for noncognitivists. After presenting Beddor's proposal and explaining why it is incomplete, I will show how my own version of noncognitivism, dispositional noncognitivism, might be able to explain how coupling two noncognitive states may generate the relevant motivation.

8.4.3 Desire as a Disposition to Act

In a recent paper Bob Beddor makes some remarks that might help the noncognitivists with what I call the motivation problem (Beddor, 2020). Beddor's focus is not on explaining motivation, but on providing a solution to the previously discussed 'many attitudes problem' which is the problem for noncognitivists of extending their noncognitivist account of moral belief to other attitudes such as moral desires, moral hopes, moral fears etc. Part of the many attitudes problem is, as he says, to 'make sense' of normative or moral desires, that is desires with a moral content. His suggestion, which he bases on a proposal made by Sebastian Köhler in (Köhler, 2017), is that all attitudes can be defined by making reference to the mental state of belief. More precisely, he thinks that every non-belief attitude, such as desire, hope, fear etc., can be defined in terms of the functional role it plays in combination with belief. In this way, Beddor thinks, noncognitivists can deal with the many attitudes problem: once noncognitivists have a noncognitivist account of moral belief, they can define all other moral attitudes, such as moral desire, by telling us its functional relationship to moral belief. Even though the many attitudes problem is not my topic here, parts of Beddor's discussion are clearly helpful for the motivation problem.

Beddor considers the following case of an agent whom he calls Michelle. He asks us to imagine that Michelle desires the following:

No Wrong: I do nothing wrong today.

and also believes:

Promise: The only way to ensure that I do nothing wrong today is to keep my promise. (Beddor, 2020, 2797)

Arguably, these two states motivate the agent to keep her promise. Other than myself Beddor sees no problem here for the noncognitivist:

> Noncognitivists can happily accept all of this. Where they part ways with the cognitivist is in their account of the relevant means-ends belief (*Promise*). For the noncognitivist, Michelle's belief in promise is itself a nonrepresentational, desire-like state. For example, noncognitivists like Gibbard will analyze Michelle's belief in promise as a 'contingency plan': a *pro tanto* plan to blame herself for failing to keep her promise, in the contingency that she fails to do so. (Beddor, 2020, 2797, italics original)

How exactly can noncognitivists explain Michelle's motivation to keep her promise? According to Beddor, the noncognitivist's explanation is simple. He assumes that part of the functional role of desire is the following "motivational condition on desire" (Beddor, 2020, 2796):

Disposition-to-Act: If S desires A, then S is disposed to perform whatever
action(s) S believes has the best chance of satisfying A. (Beddor, 2020,
2796)[35]

Dispositions-to-Act implies that the agent who holds the desire that
she does nothing wrong today is disposed to perform whatever action she
believes has the best chance of making it the case that she does nothing
wrong today. Of what action does she believe this? In order to answer
this, Beddor suggests, we simply have to look at Michelle's above belief
(Promise) from which we can simply read off that *keeping her promise* is
the action Michelle believes to have the best chance of satisfying her moral
desire. Hence, according to Beddor, the motivational condition on desire
"predicts" that a person who believes that keeping her promise has the best
chance of doing nothing wrong today, and who also desires to do nothing
wrong today, is disposed to keep her promise which, according to Beddor, is
tantamount to be motivated to keep her promise.

Does this solve the motivation problem for noncognitivists? I think: no.
This is not a solution because the noncognitive state (e. g. the Gibbardian
nonrepresentational, desire-like 'contingency plan', or whatever other state
noncognitivists take it to be) constituting the moral belief does not play
any role in the explanation. If Beddor were right, then we would be able to
explain how a moral belief together with a moral desire can motivate without
knowing either what noncognitive state the belief is or what the desire is. But
we cannot explain it so. This track of explaining how noncognitive states can
interact to generate motivation is too fast. The noncognitive state underlying
the moral belief must be explanatorily and causally relevant, since the moral
belief by itself has no nature besides being some underlying noncognitive
state—it is a mere quasi-belief. In other words, if we were told only what
the relevant noncognitive state is and that desiring disposes agents to do
what they believe to have the best chance of satisfying some goal, we would
be unable to predict what the agent is motivated to do.

If an agent is in the noncognitive states underlying the quasi-beliefs and
desires, then this will cause her to be motivated independently of whether
or not she knows which quasi-belief it is. The causal work of actually
motivating agents to act is done exclusively by the underlying noncognitive
states, not by the quasi-belief. Agents do not need to know the quasi-realist
description or belief-name of the noncognitive state in order for these states
to motivate them to do certain things. Simply by being in whatever states
the noncognitivist tells us constitute the promise-belief and the desire, agents
find themselves to be motivated. As Beddor's explanation stands, it is totally
in the dark how the noncognitive state underlying the belief and the desire to
do no wrong interact with each other, so as to cause Michelle to be motivated

---

[35]I replaced Beddor's variable 'p' with 'A' to be consistent with my use below.

to keep her promise. Hence, Beddor's explanation does not tell us how the underlying mechanics of motivation is supposed to work if noncognitivism is true.

8.4.4   Solving the Problem

Can noncognitivists do better? What noncognitivists have to offer in order to solve the motivation problem is the following: they have to reformulate the motivational condition on desire so that it fits their favored noncognitive analysis of moral belief. This is not an easy task, since it has not only to explain Beddor's case and my good man case from above. Rather, noncognitivists have to explain the general 'mechanics' of motivation in terms that apply to *all* means-end-belief/desire pairs irrespective of whether they are purely descriptive, purely moral or mixed combinations of both.

Can noncognitivists reformulate the functional role of desires (Disposition-to-Act) in such a way that it applies to the state which noncognitivists take beliefs (moral or non-moral) to be and get the right predictions about motivation? How can this be acceved? A rough idea would be this:

Rough Idea: For an agent S to desire that A is for S to be disposed to perform whatever action X such that S is in the (potentially *noncognitive*) state M and X is somehow provided by M.

The challenge is to show, in a general way, how some action can be 'somehow provided' by some mental state. Doing so will of course depend on the noncognitivist's account of the nature of the relevant means-end-belief. I will here sketch a solution based on my own dispositional noncognitivism.

Let me first state the proposal and then unpack it. Here is my proposal:

Disposition-to-Act (reformulated): For S to desire that A is for S to be disposed to perform whatever action β^* such that S is disposed to (if β, then α), where the semantic content of A is α and β is the semantic content of B.

In order to understand this let me quickly recap the cornerstones of my dispositional noncognitivism. Recall that above in the section on Attitude Semantics I developed a formal expressivistic semantics in order so solve the expressivists' problems with meaning and logic (aka the Frege-Geach problem). The key idea there was that sentences express mental states with some specific structure. In the atomic case, first-order sentences of the form '$F(t)$', where F is some predicate and t some singular term, always express a mental state of the form $M(R_1(R_2(t)))$. I then gave this structure a dispositional interpretation. For instance, building on ideas from Mark Schroeder (M. Schroeder, 2008a), I suggested that the sentence 'murder is wrong' expresses the disposition to be for blaming for murdering. So I

replaced 'M' with 'disposition', 'R_1' with 'being for', 'R_2' with 'blaming', and the singular term 't' with 'murdering'. Likewise for descriptive sentences. For instance, I suggested that 'grass is green' expresses the disposition of proceeding-as-if grass is green (again I will abbreviate 'proceeding as if' as 'pai'). So slightly more formally, my proposal was the following:

$$\text{`murder is wrong'} \quad \text{expresses} \quad \overbrace{\text{DIS}}^{M}(\overbrace{for}^{R_1}(\overbrace{blaming}^{R_2}(\overbrace{murder}^{t}))),$$

$$\text{`grass is green'} \quad \text{expresses} \quad \overbrace{\text{DIS}}^{M}(\overbrace{pai}^{R_1}(\overbrace{green}^{R_2}(\overbrace{grass}^{t}))).$$

In the expressivistic semantics I associated every sentence A with a semantic value which in the metalanguage I denoted by α, where the semantic value includes everything followed by the mental state 'M'. So the semantic value of 'murder is wrong' is $for(blaming(murder))$, or more colloquially, it is the property of *being for blaming for murder* (that is without the prefixed disposition). From the way the expressivistic semantics is set up it follows that every semantic value α involves what I called 'outer' and 'inner relations', R_1 and R_2, respectively. This is important because I treat α^* as simply being α with every outer relation R_1 appearing in it being crossed out. For instance if A is 'grass is green', then α is *proceeding as if grass is green*, and so α^* is just *grass is green*. Similarly, if A is 'murder is wrong', and so α is *being for blaming for murdering*, then α^* is *blaming for murdering*.

Lets test this proposal. Start with a purely descriptive example. If I desire that I relax, and believe that if I take a warm bath, then I will relax, this intuitively motivates me to take a warm bath. How can the dispositionalist noncognitivist explain this? Let B be 'I relax' and let A be 'I take a warm bath'. According to dispositional noncognitivism the semantic contents of B and A, β and α respectively, are *proceed as if I take a warm bath* and *proceed-as-if I relax*. So, to believe that if I take a warm bath, then I will relax is to be disposed to make it the case that (if I proceed as if I take a warm bath, then I proceed as if I relax).[36] According to the reformulated Disposition-to-Act my desire to do A disposes me to perform whatever action β^* such that I am disposed to (if β, then α). Since β is *proceed as if I take a warm bath*, β^* is *I take a warm bath*. It follows that I am disposed to take a warm bath, which is tantamount to being motivated to take a warm bath.

Hence, my proposal makes the right prediction in this case. Crucially, the dispositional mental state (and not its quasi-realistic belief description) which is identified with the means-end-belief has played a real role in explaining how this belief and desire together generate this specific motivation. Moreover, both the state making up the belief as well as the desire played a role, and only those two states played a role—no additional state was required to generate the motivation, and neither the belief nor the desire by themselves

---

[36]I ignore here for the sake of simplicity that my semantics in chapter 4 handles conditionals slightly different.

would have been sufficient for generating the motivation. So, this account is genuinely noncognitivistic and it respects what I above called 'data'.

Now consider my moral example from the beginning of this section. Let B be 'I am a good man' and A be 'My wife will love me'. According to dispositional noncognitivism, to believe that if I am a good man, my wife will love me is the state of being disposed to make it the case that if I am being for praising the way I am as a man, then I proceed-as-if I am being loved by my wife.[37] In other words, the semantic contents of B and A, β and α respectively, are *being for praising the way I am as a man* and *proceeding as if I am being loved by my wife*. Since I desire that A, that is I desire being loved by my wife, the reformulated Disposition-to-Act predicts that I am disposed to be β^*, that is disposed to make it the case that the way I am as a man is one that is praised. Hence, my desiring that my wife will love me and believing that if I am a good man, my wife will love me disposes me to make it the case that the way I am as a man is praised.[38] Being so disposed, noncognitivists should claim, is tantamount to being motivated to be a good man. Like in the purely descriptive case, this account makes the correct prediction. Moreover, again, the noncognitive state making up the means-end-belief plays a real role in explaining how the moral means-end-belief together with the desire generates this specific moral motivation. Hence, the dispositional noncognitivist has a solution to the motivation problem.

I will not discuss how my dispositional noncognitivism deals with the case of combining a purely moral means-end-belief with a moral desire to yield a moral motivation, since this follows exactly the same pattern as the purely descriptive and the problematic mixed cases.[39]

8.5   Conclusion

In this section I have introduced an unnoticed problem for noncognitivism, which I called the motivation problem. The problem is to offer an explanation of how means-ends-beliefs paired with desires can motivate if some means-end-beliefs are themselves desire-like states as the noncognitivist claims.

---

[37] In this toy example I assume, for the sake of argument, that the semantic contribution of 'good' is the relation or attitude of 'praising' so that it aligns with Schroeder's suggestion that the semantic contribution of 'wrong' is 'blaming'.

[38] If you have strong objectivist inclinations about the status of moral facts, you will not accept that it is completely up to you to make it the case that no matter what way you are as a man, this way will be praised. Instead you will hold that the only thing that is within your power is to choose the way you are as a man and that you can choose some way which actually is praised.

[39] I want to mention that this solution is absolutely general not only in that it explains purely descriptive, purely moral, and mixed cases. It is general also in that it works for combinations of means-end beliefs and desires of arbitrary logical complexity. This is due the constructive solution to the Frege-Geach problem developed in chapter 4.

Even though the problem requires noncognitivists to abandon Humeanism because they lack the required cognitive belief state, and noncognitivists have to come up with their own non-Humean solution, it does not seem to be impossible to solve the problem. Yet it is a problem noncognitivists have to solve, and it is highly likely that not every form of noncognitivism will be able to solve it. I sketched how my own version of dispositional noncognitivism can deal with the motivation problem. Whether or not this solution is really viable and whether there are better ones, I want, however, to leave for others to judge, since my main intent here has been to bring attention to a novel and difficult problem for noncognitivism.

My overall argument in this chapter is not that noncognitivists cannot solve their problems with moral motivation, but only that they are problems, that noncognitivists have to solve them, and that solving them might not be trivial. Importantly, even if noncognitivists can solve those problems, especially the just discussed motivation problem, this suggests that cognitivists have an easier time explaining moral motivation. This is a rather surprising conclusion since it directly contradicts the widely accepted view that noncognitivism is superior to cognitivism when it comes to explaining moral motivation. If my arguments are correct, this has important consequences for the usual metaethical dialectic since noncognitivism's alleged ability to easily explain moral motivation is traditionally regarded as one of the major reasons why some philosophers become noncognitivists in the first place. If, however, noncognitivism is actually worse at explaining motivation than cognitivism, this removes a central reason to believe in the truth of noncognitivism.

Chapter 9

Conclusion

In this dissertation I developed and defended a form of expressivism in metaethics. In this conclusion I want to quickly recapitulate what this dissertation originally set out to do, what are its main claims and findings, as well as list some open questions and suggest directions for future research based on this study.

9.1 Where We Started

The central task of this dissertation has been to answer the question of how we can express moral belief, if we assume metaethical expressivism. In order to answer this question I broke it down into several sub-questions. In the order of appearance those questions have been the following: If metaethical expressivism is true,

1. *Expressing*: what is it to express a mental state?

2. *Moral Meaning*: how can moral sentences be meaningful and stand in logical relations?

3. *Moral Truth-Aptness*: how can moral sentences be true or false?

4. *Moral Belief*: what is it to hold a moral belief?

5. *Moral Reasoning*: how can moral beliefs stand in rational relations and is reasoning with them wishful thinking?

6. *Moral Motivation*: and finally, how do moral beliefs motivate us to act?

These questions set the structure of the dissertation and I organized them in three main parts. The first part was concerned with expressivism and dealt with the language-related questions of expressing a mental state

(chapter 2), compositionality and logicality (chapter 3), as well as the question of what it takes for a a formula of some language to be capable of being assessed in terms of truth and falsity (chapter 4). The second part was concerned with noncognitivism and the question of the nature of moral belief (chapter 5). Moreover, in this part I also dealt with two objections against noncognitivism which both targeted the rationality of moral belief (chapter 6 and 7 respectively). In the third and final part of this dissertation I switched gears and raised two unnoticed problems with how metaethical expressivists explain moral motivation (chapter 8).

9.2   Claims and Findings

The key claims and findings of this dissertation can be summarized as follows. In chapter 2 I investigated the relatively underexplored, yet fundamental notion of 'expressing'. As an answer to the first question from above I argued that a sentence expresses a mental state iff this sentence is conventionally used to perform a certain speech act where it is the sincerity condition of this speech act that speakers are in this mental state. This explication of 'expressing' satisfies all the constraints necessary in order to suit the needs of expressivists. For instance, it allows for insincerity, and offers a certain stability. These properties are needed if expressing is to function as the central semantic notion of expressivism. The sincerity account of expressing is a simple yet useful account and connects to the speech act theoretical origins of metaethical expressivism. In virtue of its reference to speech act theory, this account also nicely aligns with the theory of truth-aptness which I defended in chapter 4 according to which truth-aptness depends on the speech act of assertion. The sincerity account is a major contribution to a better understanding of the foundations of expressivism.

I then turned to the most notorious problem every expressivistic semantics faces: the Frege-Geach problem. In order to solve this problem, and thereby answer the second of the above questions, I developed, building on the work of Unwin and Schroeder, what I called Attitudes Semantics and Attitude Logic for a first-order language. This semantics and logic, I showed, solves the expressivists' problems with compositionality and logicality. The central idea of this chapter was that expressivists should enrich the structure of the mental state expressed by atomic moral sentences to match the structure of the mental state expressed by descriptive sentences. As a first pass I suggested that 'murder is wrong' expresses something like the following state: being for being for blaming for murder. I emphasized that I am not wedded to this particular interpretation, however. What is important for the solution to work primarily is that the states expressed have this kind of structure. In order to make this structural suggestion philosphically more plausible, and to avoid adhocness objections, I suggested a dispositionalist interpretation

of the structure of the states which attitude semantics assigns to sentences of the language. To accept 'murder is wrong' I suggested is to be disposed to be for blaming for murder. The dispositionalist proposal, as well as the interpretation in this particular case (in terms of 'being for' and 'blaming') is not carved in stone. But a strong reason in favor of dispositional expressivism is that it turned out to perfectly match the noncognitivists account of belief in the second part.

Before I turned to the second part and my discussion of moral belief, I dealt with a central goal on the quasi-realist expressivists' agenda, namely answering the third question about how moral sentences can be truth-apt. By carefully criticizing a number of proposals concerning what makes something truth-apt, I developed my own theory of truth-aptness. I argued that for something to be truth-apt there must be a conventional way to use it to make an assertion, or what is equivalent, a way to use it to express a belief. As a result we found that metaethical expressivists need to invert what I called the orthodox quasi-realist strategy. Instead of arguing from truth-aptness to belief, quasi-realist expressivists need to argue from belief to truth-aptness. We have seen that this puts great parts of the existing literature upside-down and sets old arguments into a new and interesting perspective. The important consequence of all this is that expressivists must explain how noncognitive states can be regarded as moral beliefs prior to their being able to conclude that sentences expressing those states are truth-apt. In other words, expressivists must answer the fourth question before they can answer the third question.

The second part of this dissertation was then concerned with 'earning the right' to moral belief, that is with answering question 4. My central aim in chapter 5 was therefore to sketch a general theory of the nature of belief that is compatible with metaethical expressivism. For such a theory to be compatible with expressivism, I explained that it would have to be able solve problems such as the one word problem, the many attitudes problem, the moral attitude problem and what I called the tightrope problem among others. The most promising option for expressivists, we saw, is within the functionalist tradition about mental states, of which dispositionalism is a species. After I presented Schwitzgebel's general dispositionalism about belief, and Köhler's functionalist 'conceptual role expressivism', it emerged that an attractive view is a mixture of both which I called 'dispositional noncognitivism'. Dispositional noncognitivism takes Schwitzgebel's attractive idea that beliefs are clusters of stereotypical dispositional properties, but it rejects the radical liberalism of Schwitzgebel's conception according to which no specific property is necessary for something to count as a belief. Instead, it agrees with Köhler's view according to which some properties are necessary for some state to count as a belief. According to dispositional noncognitivism it is essential to the mental state of belief that the set of dispositions constituting it involves dispositions that allow us to explain the

rational properties we expect beliefs to have, that is they have the right mind-to-mind conditions. In particular the dispositions must be able to explain why the state which we describe as the state of believing something is rationally incoherent with the state of believing the opposite. However, even though having this property is necessary or central to belief, it is not the only necessary condition and it is also not sufficient for belief. In the case of *descriptive* belief, it is also necessary that the set of dispositions constituting it includes dispositions to react to sensory input (i.e. mind-entry conditions). In the case of *moral* belief, it is also necessary that the set of dispositions includes dispositions to produce actions (i.e. mind-exit conditions). This dispositional understanding of belief also allows noncognitivists to draw a categorical distinction between descriptive and moral beliefs without collapsing the categorical distinction between beliefs and desires. Dispositional noncognitivism, however, not only solves the tightrope problem, it also offers solutions to the other above-mentioned problems. Moreover, this theory nicely matches and thus justifies dispositional expressivism. Most importantly, dispositional noncognitivism earns expressivists the right to claim that the states expressed by moral sentences are moral beliefs, despite the fact that they are ultimately constituted by noncognitive states. This allows the quasi-realist metaethical expressivist to conclude that moral sentences express moral beliefs. More precisely, it allows the dispositional expressivist from chapter 3 to claim that the dispositional states expressed by the sentences of attitude semantics can be regarded as moral beliefs, and that those sentences are therefore (via the truth-aptness theory from chapter 4) capable of being true or false. This completed my answers to the meaning and truth-aptness questions, that is questions 3 and 4.

Subsequently I turned to question 5 by considering two epistemological objections against metaethical expressivism, and especially its noncognitivism, in chapters 6 and 7 respectively. Though Baker's objection is directed at Schroeder's specific biforcated attitude semantics, discussing this objection was important because my own attitude semantics resembles it in crucial respects and one might have worried that Baker's objection therefore carries over. Fortunately, we have seen that Baker's objection rests on a number of mistaken assumptions and so does not apply to Schroeder's expressivism. It is therefore also highly unlikely, I argued, that a Baker-style objection against the here defended dispositional expressivism would be successful.

I then offered a new solution to what is considered the most difficult epistemological objection against metaethical expressivism: Cian Dorr's famous wishful thinking objection. Previous replies to the objection have all tried to show that even though accepting a moral sentence is holding a noncognitive desire-like state (if noncognitivism is true), coming to accept a moral sentence might nevertheless also be systematically accompanied by a change in cognitive states. I took a different path and directly questioned Dorr's major assumption underlying the objection, namely that it can never

be rationally justified to base a belief on a desire, since this would be wishful-thinking. I was able to show that this assumption is false and that rational reasoning with moral beliefs is not wishful thinking even if noncognitivism is true. In fact, noncognitivists are even able to explain how rationally reasoning with moral beliefs gives agents reasons of the right kind, that is epistemic reasons for factual beliefs. This is a major contribution to the research program of metaethical expressivism. Chapter 6 and 7 thus answered the reasoning question 5. This completed the second part of the dissertation and with it also my defense of metaethical expressivism.

In third and final part of this study I switched gears and presented two hereto unnoted problems for metaethical expressivism with answering question 6, that is with explaining moral motivation. My arguments were surprisingly simple. The first one was this: if motivation requires a belief as well as a desire (Humeanism), and moral judgments motivate by themselves (internalism), then moral judgments cannot be desires only. In order to block this argument, I discussed a number of modifications of the theses involved in it. Even though some of those modifications have the potential to escape the first problem, a more comprehensive assessment of those solutions must await another occasion.

My second argument was also quite simple: moral judgments when combined with suitable external desires motivate to act, but if motivation requires belief as well as desire (Humeanism), then moral judgments cannot be desires only. This argument, I argued, forces noncognitivists to step back from traditional Humeanism, and instead explain how pairs of noncognitive states can interact with each other to motivate agents to act. Building on the idea that to desire something is to be disposed to perform some action which is somehow contributed by making reference to some other mental state (i.e. a means-end-belief), I sketched how my dispositional form of metaethical expressivism is able to avoid the problem. However, here as with the first motivation problem, I did not attempt to give a fully worked-out reply. It is, for instance, an open question if other forms of metaethical expressivism are also able to provide a solution along the proposed lines. Hence, the above sixth question is not yet closed.

So, what about my original question: How can we express moral belief if metaethical expressivism is true? The short answer for which I have argued in this dissertation is this: by becoming a dispositional expressivist. For the dispositional expressivists to hold a belief is to have certain dispositions. Crucially, to hold a *moral* belief is to be disposed to be in a certain *noncognitive* state. Such states have a certain structure and nature which allows them to be systematically assigned to the formulas and explain how their meaning can be compositional and how they can stand in logical relations. Partly in virtue of their being able to explain those things, these states can be regarded as beliefs, and so the formulas expressing them are truth-apt.

9.3 Future Research, Open Questions, Directions

Even though this dissertation answered a number of important questions
about expressing moral belief, and it solved several problems and thereby
made some important contributions to the research program of expressivism
in metaethics, it also raises some open questions. In this final section I want
to say what I take to be interesting directions for future research based on
the claims and findings of this work.

Let me begin with my proposed solution to the Frege-Geach problem in
chapter 3. I have remarked at many places that the specific interpretations
of the mental states expressed by moral sentences like 'murder is wrong'
must be regarded as toy examples and I am not wedded to them in any
way. This holds in particular for my first pass proposal that 'murder is
wrong' expresses being for being for blaming for murder, but it even holds
for my dispositional interpretation. A fully worked-out expressivism along
the lines I have suggested must say much more about the precise nature
of the states and properties involved in the state expressed by sentences.
Fortunately, expressivists here do not have to completely reinvent the wheel.
Rather, they are well advised to look for inspiration into the literature on
the moral attitude problem, and more generally what other expressivists
have already said about the nature of moral judgments. For instance, it
would be an interesting and I suspect promising idea to see whether and how
Canberra-style approaches to the moral attitude problem such as the one
presented by (Köhler, 2013), can tell us something more specific about the
nature of the noncognitive state one is disposed to be in if one holds a moral
belief.

Another important point concerns the structure of the mental states.
For the sake of comparability I have closely followed Schroeder's way of
formalizing those state, e. g. by using lambda-abstractions. More colloquially,
Schroeder refers to the objects of his being for-states as 'properties'. This
suggests that for Schroeder, the state of being for is not a propositional
attitude. It is a question worth further investigation what the formal and
philosophical benefits would be if instead of saying that 'murder is wrong'
expresses the disposition to be for *blaming* for murder we would say, for
instance, that it expresses the disposition to be for the fact *that* murder is
blamed.

This is closely related to another important question, namely: are there
more elegant ways of formalizing attitude semantics? I am pretty sure that
there are, but I have to leave this to technically more advanced philosophers
than I am myself. Another important question is, of course, how the ideas of
attitude semantics can be applied to a propositional language, or a language
containing modal operators etc.

This dissertation defended what I called dispositional expressivism and dispositional noncognitivism. So dispositions play a central role in many parts of this study. Nevertheless, it is clear that a fully worked-out dispositional expressivism (and noncognitivism) would have much more to say about the nature of dispositions and the complicated issues which are discussed in the philosophy of dispositions. It is an interesting open question if every conception of dispositions will equally suit the needs of the dispositional expressivists, which suits it best, and which are perhaps completely unsuitable.[1]

Another interesting question that can be asked about dispositional expressivism is how it relates to the Stevensonian view that meanings are dispositions of words. This latter view is today known as *semantic* dispositionalism. Semantic dispositionalism is the view, roughly, that to speaker-mean something by a word is to be disposed to use it to do certain things.[2] For instance it might be said that to mean addition by '+' is to be disposed to give the sum of two numbers x and y, for any x and y, when asked for the sum of 'x + y'. It is well-known that Kripke thought semantic dispositionalism to be an inadequate response to Wittgenstein's rule-following considerations because dispositions cannot capture the fact that meaning is normative, that is that there are correct and incorrect ways to follow the semantic rules of language. The rule-following considerations also played an important role in metaethics between the cognitivists John McDowell and the noncognitivist Simon Blackburn. It would be interesting to revisit this debate against the background of Stevenson's dispositional theory of meaning and the dispositional expressivism which I have proposed in this thesis. I hope to return to this topic in future research.[3]

Finally, of course, there are many open questions regarding the two motivation problems I have raised in part three. Some of those question are: can desires motivate by themselves or at least some of them? Is going mildly hybrid a viable option for noncognitivists? Can other forms of noncognitivism, such as for instance Gibbard's norm expressivism, provide similar solutions to the second motivation problem by using the idea that desires are dispositions to act?

---

[1]I have taken myself the liberty to remain silent on the nature of dispositions in this dissertation because even the central figure of dispositionalism about belief, Eric Schwitzgebel, is surprisingly silent about this topic.

[2]Due to the reference to speaker-meaning, it would also be very interesting to see how semantic dispositionalism relates to what I have said in chapter 2 on expressing, especially Wayne Davis' Neo-Gricean approach.

[3]For an excellent exposition and contemporary defense of semantic dispositionalism see Arvid Båve (Båve, 2020). The central texts on rule-following are, of course, Wittgenstein's *Philosophical Investigations* (Wittgenstein, 1953, §243ff.) and Kripke's interpretation (Kripke, 1982/2010). Classic papers are also (P. A. Boghossian, 1989), (McDowell, 1984), (Blackburn, 1984a). The relevant texts in metaethics are (McDowell, 1981), (Blackburn, 1981), and (Miller, 2019a), (Miller, 2019b).

I conclude that this study makes some serious advances for metaethical expressivism and deserves to be investigate further. I am generally sympathetic to the expressivist research program, and I am cautiously optimistic about its prospects. My caution is due to the fact that even though metaethical expressivism has many things to offer for the naturalistically inclined moral philosopher, there remain several open questions. Nevertheless, I hope that the results of this work will help us to better understand what we do when we express moral beliefs.

References

Anscombe, E. (1963). *Intention.* Cambridge: Harvard University Press.

Armstrong, D. M. (1968). *A Materialist Theory of the Mind.* London: Routledge.

Arpaly, N. (2002). Moral Worth. *The Journal of Philosophy*, *99*(5), 223-245.

Artemov, S., & Fitting, M. (2021). Justification Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2021/entries/logic-justification/`.

Audi, R. (1973). The Concept of Believing. *The Personalist*, *53*, 43-62.

Audi, R. (1994). Dispositional Beliefs and Dispositions to Believe. *Noûs*, *28*(4), 419-434.

Austin, J. (1962). *How to Do Things With Words.* Cambridge: Harvard University Press.

Ayer, A. J. (1952). *Language, Truth, and Logic.* New York: Dover Press. (original 1936)

Baker, D. (2018). Expression and Guidance in Schroeder's Expressivist Semantics. *Erkenntnis*, *83*(4), 829-852.

Baker, D., & Woods, J. (2015). How Expressivists Can and Should Explain Inconsistency. *Ethics*, *125*(2), 391-424.

Barker, S. (2002). Is Value Content a Component of Conventional Implicature? *Analysis*, *60*(3), 268–279.

Beddor, B. (2019). Noncognitivism and Epistemic Evaluations. *Philosopher's Imprint*, *19*(10), 1-27.

Beddor, B. (2020). A Solution to the Many Attitudes Problem. *Philosophical Studies*, *177*, 2789–2813.

Belnap, N. (2001). *Facing the Future.* Oxford: Oxford University Press.

Bennett, J. (1976). *Linguistic Behaviour.* Cambridge: Cambridge University Press.

Bird, A. (2007). *Nature's Metaphysics. laws and properties.* Oxford: Oxford University Press.

Björnsson et al. (2015). *Motivational Internalism.* New York: Oxford University Press.

Blackburn, S. (1981). Rule-Following and Moral Realism. In S. Holzmann & C. Leich (Eds.), *Wittgenstein: To Follow a Rule* (p. 163-189). London:

Routledge.

Blackburn, S. (1984a). The Individual Strikes Back. *Synthese*, *58*, 281-301.

Blackburn, S. (1984b). *Spreading the Word*. New York: Oxford University Press.

Blackburn, S. (1988). Attitudes and Contents. *Ethics*, *98*(3), 501-517.

Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Clarendon Press.

Blackburn, S. (1998). *Ruling Passions*. Oxford University Press.

Boghossian, P. (1990). The Status of Content. *The Philosophical Review*, *99*(2), 157-184.

Boghossian, P. A. (1989). The Rule-Following Considerations. *Mind*, *98*, 507-549.

Boisvert, D. R. (2014). Expressivism, Nondeclaratives, and Success-Conditional Semantics. In G. Fletcher & M. Ridge (Eds.), *Having it Both Ways: Hybrid Theories and Modern Metaethics* (p. 22-50). Oxford: Oxford University Press.

Boisvert, D. R. (2016). Charles Leslie Stevenson. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/win2016/entries/stevenson/`.

Bradley, R., & List, C. (2009). Desire-as-Belief Revisited. *Analysis*, *69*(1), 31-37.

Braithwaite, R. B. (1932). The Nature of Believing. *Proceedings of the Aristotelian Society*, *33*, 129-149.

Brandom, R. (1994). *Making it Explicit. Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.

Bratman, M. E. (1999). *Faces of Intention*. Cambridge: Cambridge University Press.

Braun, D. (2011). Implicating Questions. *Mind and Language*, *26*, 574-595.

Broome, J. (2013). *Rationality Through Reasoning*. Oxford: Wiley Blackwell.

Budolfson, M. B. (2011). Non-Cognitivism and Rational Interence. *Philosophical Studies*, *153*(2), 243-259.

Burgess, A., & Sherman, B. (Eds.). (2014). *Metasemantics. New Essays on the Foundations of Meaning*. Oxford: Oxford University Press.

Båve, A. (2020). Semantic Dispositionalism Without Exceptions. *Philosophical Studies*, *177*, 1751–1771.

Campbell, D. I. (2017). Against Lewis on 'Desire as Belief'. *Polish Journal of Philosophy*, *11*(1), 17-28.

Cantwell, J. (2013). First Order Expressivist Logic. *Erkenntnis*, *78*(6), 1381-1403.

Cantwell, J. (2015). An Expressivist Bilateral Meaning-is-Use Analysis of Classical Propositional Logic. *Journal of Logic, Language, and Information*, *24*, 27-51.

Carnap, R. (1951). *Logical Foundations of Probability* (1962 2nd ed.). Chicago: University of Chicago Press.

Carroll, L. (1936/1871). Through the Looking-Glass. In *The Complete Works of Lewis Carrol* (p. 213-214). New York.

Carruthers, P. (2013). On Knowing Your Own Beliefs: A Representationalist Account. In *New Essay on Belief. Constitution, Content, Structure* (p. 145-163). New York: Palgrave MacMillan.

Chrisman, M. (2016). *The Meaning of 'Ought'. Beyond Descriptivism and Expressivism in Metaethics.* New York: Oxford University Press.

Clifford, W. K. (1877 [1999]). The Ethics of Belief. In T. Madigan (Ed.), *The Ethics of Belief and Other Essays* (p. 70–96). Prometheus.

Collins, J. (2002). Truth or Meaning? A Question of Priority. *Philosophy and Phenomenological Research*, *65*(3), 497-536.

Copp, D. (2001). Realist-Expressivism: A Neglected Option for Moral Realism. *Social Philosophy and Policy*, *18*(02), 1-43.

Dancy, J. (1993). *Moral Reasons.* Oxford: Blackwell.

Davidson, D. (1967). Truth and Meaning. *Synthese*, *17*(3), 304-323.

Davidson, D. (1984). *Inquiries into Truth and Interpretation.* Oxford: Clarendon.

Davis, W. A. (1984). A Causal Theory of Intending. *American Philosophical Quarterly*, *21*(1), 43-54.

Davis, W. A. (2003). *Meaning, Expression, and Thought.* New York: Cambridge University Press.

Davis, W. A. (2005). *Nondescriptive Meaning and Reference: An Ideational Semantics.* Oxford: Clarendon Press.

Dennett, D. (1978). *Brainstorms.* Cambridge: MIT Press.

Dennett, D. (1987). *The Intentional Stance.* Cambridge: MIT Press.

de Rizzo, J. (2020). Grounding Grounds Necessity. *Analysis*, *80*(4), 639-647.

Divers, J., & Miller, A. (1994). Why Expressivists about Value Should Not Love Minimalism about Truth. *Analysis*, *54*(1), 12-19.

Donnellan, K. S. (1968). Putting Humpty Dumpty Together Again. *The Philosophical Review*, *77*(2), 203-215.

Dorr, C. (2002). Non-Cognitivism and Wishful Thinking. *Noûs*, *36*(1), 97-103.

Dreier, J. (1990). Internalism and Speaker Relativism. *Ethics*, *101*(1), 6–26.

Dreier, J. (1996). Expressivist Embeddings and Minimalist Truth. *Philosophical Studies*, *83*(1), 29-51.

Dreier, J. (2004a). Lockean and Logical Truth Conditions. *Analysis*, *64*(1), 84–91.

Dreier, J. (2004b). Meta-Ethics and the Problem of Creeping Minimalism. *Philosophical Perspectives*, *18*, 23-44.

Dummett, M. (1959). Truth. *Proceedings of the Aristotelian Society*, *59*, 141-162.

Eggers, D. (2014). The Motivation Argument and Motivational Internalism.

Philosophical Studies, *172*, 2445–2467.

Eggers, D. (2015). Unconditional Motivational Internalism and Hume's Lesson. In Björnsson et al. (Ed.), *Motivational Internalism* (p. 85-107). New York: Oxford University Press.

Engel, P. (2005). Belief as a Disposition to Act: Variations on a Pragmatist Theme. *Cognitio*, *6*(2), 167-185.

Enoch, D. (2003). How Noncognitivists can Avoid Wishful Thinking. *The Southern Journal of Philosophy*, *41*, 527-545.

Eriksson, J. (2014). Elaborating Expressivism: Moral Judgments, Desires and Motivation. *Ethical Theory and Moral Practice*, *17*(2), 253–267. Retrieved from http://www.jstor.org/stable/24478575

Fara, M. (2005). Dispositions and Habituals. *Noûs*, *39*(1), 43-82.

Fassio, D. (2013). A Blind-Spot Argument Dispositionalist Accounts of Belief. *Acta Analytica*, *29*, 71-81.

Field, H. (1986). The Deflationary Conception of Truth. In C. W. Graham Macdonald (Ed.), *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic* (p. 55-117). Oxford: Blackwell.

Fisher, A. (2011). *Metaethics. An Introduction.* New York: Routledge.

Fodor, J. (1975). *The Language of Thought.* New York: Cromwell.

Freitag, W. (2018). Wittgenstein on 'I believe'. *Grazer Philosophische Studien*, *95*(1), 54-69.

Gertken, J., & Kiesewetter, B. (2017). The Right and the Wrong Kind of Reasons. *Philosophy Compass*, *12*(5), 1-14.

Gibbard, A. (1990). *Wise Choises, Apt Feelings.* Cambridge: Harvard University Press.

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

Grice, P. (1968). Utterer's Meaning, Sentence Meaning, and Word Meaning. In *Studies in the Ways of Words* (p. 117–137). Cambridge: Harvard University Press.

Grice, P. (1989). *Studies in the Ways of Words.* Cambridge: Harvard University Press.

Hanna, R. (1991). How Ideas Became Meanings: Locke and the Foundations of Semantic Theory. *The Review of Metaphysics*, *44*(4), 775-805.

Hare, R. (1952). *The Language of Morals.* Oxford: Oxford University Press.

Harman, G. (1977). *The Nature of Morality.* New York: Oxford University Press.

Harman, G. (1982). Conceptual Role Semantics. *Notre Dame Journal of Formal Logic*, *23*(2), 242-256.

Harman, G. (1986). *Change in View.* MIT Press.

Hengst, S. (2020). Expressivism and Explaining Irrationality: Reply to Baker. *Erkenntnis*. Retrieved from https://doi.org/10.1007/s10670-020-00290-9

Hieronymi, P. (2006). Controlling Attitudes. *Pacific Philosophical Quarterly*,

87, 45-74.

Holton, R. (2000). Minimalism and Truth-Value Gaps. *Philosophical Studies*, *97*(2), 137-168.

Horgan, T., & Timmons, M. (2006). Cognitivist Expressivism. In *Metaethics After Moore* (p. 255-298). Oxford: Clarendon Press.

Horty, J. F. (2001). *Agency and Deontic Logic*. Oxford: Oxford University Press.

Horwich, P. (1990). *Truth*. Oxford: Blackwell.

Horwich, P. (1993). Gibbard's Theory of Norms. *Philosophy and Public Affairs*, *22*, 67-78.

Horwich, P. (1998). *Meaning*. Oxford: Oxford University Press.

Hyman, J. (2014). Desires, Dispositions and Deviant Causal Chains. *Philosophy*, *89*(1), 83–112.

Jackson, F., Oppy, G., & Smith, M. (1994). Minimalism and Truth-Aptness. *Mind*, *103*(411), 287-302.

Jackson, F., & Pettit, P. (1998). A Problem for Expressivism. *Analysis*, *58*(4), 239-251.

Jacobson, D. (2013). Wrong Kind of Reasons Problem. In H. LaFollette (Ed.), *The International Encyclopedia of Ethics* (p. 5553-5561). New York: Blackwell Publishing.

Jake Quilty-Dunn, E. M. (2018). Against Dispositionalism: Belief in Cognitive Science. *Philosophical Studies*, *175*, 2353–2372.

James, W. (1897). The Will to Believe. In *The Will to Believe and Other Essays in Popular Philosophy* (p. 1-31). New York: Dover Publications.

Joyce, R. (2021). Moral Anti-Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/sum2021/entries/moral-anti-realism/`.

Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.

Kirkham, R. L. (1995). *Theories of Truth. A Critical Introduction.* Cambridge: MIT Press.

Kivy, P. (1992). 'Oh Boy! You Too!' Aesthetic Emotivism Reexamined. In H. Hahn (Ed.), *The Philosophy of A. J. Ayer* (p. 309-328). La Salle: Open Court.

Kolodny, N. (2005). Why be Rational? *Mind*, *114*(455), 509-563.

Kripke, S. (1982/2010). On Rules and Private Language. In Martinich (Ed.), *The Philosophy of Language* (5th ed., p. 626-638). New York: Oxford University Press.

Köhler, S. (2012). The Frege-Geach Objection to Expressivism: Structurally Answered. *Journal of Ethics and Social Philosophy*, *6*(2), discussion note.

Köhler, S. (2013). Do Expressivists have an Attitude Problem? *Ethics*, *123*(3), 479-507.

Köhler, S. (2017). Expressivism, Belief, and All That. *The Journal of Philosophy*, *114*(4), 189-207.

Köhler, S. (2018). Expressivism, Meaning and All That. *Canadian Journal of Philosophy*, *48*(3-4), 337-356.

Kölbel, M. (2018). Perspectival Representation and Fallacies in Metaethics. *Canadian Journal of Philosophy*, *48*(3-4), 379-404.

Lee, W. (2018). Reasoning, Rational Requirements, and Occurrent Attitudes. *European Journal of Philosophy*, *26*, 1343-1357.

Leitgeb, H. (2017). *The Stability of Belief. How Rational Belief Coheres with Probability.* Oxford: Oxford University Presss.

Lenman, J. (2003a). Disciplined Syntacticism and Moral Expressivism. *2003*, *66*(1), 32-57.

Lenman, J. (2003b). Noncognitivism and Wishfulness. *Ethical Theory and Moral Practice*, *6*, 265–274.

Levinson, S. (1983). *Pragmatics.* Cambridge: Cambridge University Press.

Lewis, D. (1970). How to Define Theoretical Terms. *The Journal of Philosophy*, *67*(13), 427-446.

Lewis, D. (1988). Desire as Belief. *Mind*, *97*(387), 323-332.

Lewis, D. (1996). Desire as Belief II. *Mind*, *105*(418), 303-313.

Lewis, D., Smith, M., & Johnston, M. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *63*, 89–174. Retrieved from `http://www.jstor.org/stable/4106918`

Loar, B. (1981). *Mind and Meaning.* Cambridge: Cambridge University Press.

Locke, J. (1690). *An Essay Concerning Human Understanding* (P. Nidditch, Ed.). Oxford University Press.

Lynch, D. (2011). Expressivism and Plural Truth. *Philosophical Studies*, *163*(2), 385-401.

Lynch, M. P. (2009). *Truth as One and Many.* Oxford: Oxford University Press.

MacKay, A. F. (1968). Mr. Donnellan and Humpty Dumpty on Referring. *The Philosophical Review*, *77*(2), 197-202.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong.* Harmondsworth: Penguin.

Manne, K. (2015). Tempered Internalism and the Participatory Stance. In Björnsson et al. (Ed.), *Motivational Internalism* (p. 260-281). New York: Oxford University Press.

Marcus, R. B. (1990). Some Revisionary Proposals about Belief and Believing. *Philosophy and Phenomenological Research*, *50*, 133-153.

Martinich, A. P. (1980). Conversational Maxims and Some Philosophical Problems. *The Philosophical Quarterly*, *30*(120), 215-228.

McCormick, M. S. (2018). No Kind of Reason Is the Wrong Kind of Reason. In K. McCain (Ed.), *Believing in Accordance with the Evidence. New Essays on Evidentialism* (p. 261-276). Springer.

McDowell, J. (1981). Non-Cognitivism and Rule-Following. In H. et al. (Ed.), *Wittgenstein: To Follow a Rule* (p. 141-162). London: Routledge.

McDowell, J. (1984). Wittgenstein on Following a Rule. *Synthese*, *58*, 325-363.

McDowell, J. (1985). Values and Secondary Qualities. In *Morality and Objectivity* (p. 110-129). Routledge.

McHugh, C. (2012). The Truth Norm of Belief. *Pacific Philosophical Quarterly*, *93*(1), 8-30.

McNaughton, D. (1988). *Moral Vision. An Introduction to Ethics*. Oxford: Blackwell.

Miller, A. (2003). *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.

Miller, A. (2019a). Rule-Following, Meaning, and Primitive Normativity. *Mind*, *128*, 735-760.

Miller, A. (2019b). Rule-Following, Moral Realism and Non-Cognitivism Revisited. In D. M. et al. (Ed.), *Ethics in the Wake of Wittgenstein* (p. 103-125). London: Routledge.

Neale, S. (1992). Paul Grice and the Philosophy of Language. *Linguistics and Philosophy*, *15*, 509–559.

Ogden, C. K., & Richards, I. A. (1923). *The Meaning of Meaning*. New York: Harvest.

O'Leary-Hawthorne, J. (1994). Truth-Aptness and Belief. In *Philosophy in Mind. The Place of Philosophy in the Study of the Mind* (Vol. 60, p. 215-242). Dordrecht: Kluwer Academic Publishers.

Ortiz-Millán, G. (2018). Wrong Dirction? A Criticism of Direction of Fit. In G. Ortiz-Millán & J. C. Parcero (Eds.), *Mind, Language and Morality. Essays in Honor of Mark Platts* (p. 26–48). London: Routledge.

Paakkunainen, H. (2015, january). *Ethics Discussions at PEA Soup: Derek Baker and Jack Woods' "How Expressivists Can and Should Explain Inconsistency," with a critical précis by Mark Schroeder.* `http://peasoup.us/2015/01/ethics-discussions-at-pea-soup` `-baker-and-woods-how-expressivists-can-and-should-explain` `-inconsistency-with-precis-by-schroeder/`.

Pigden, C. R. (2009). If Not Non-Cognitivism, Then What? In C. R. Pigden (Ed.), *Hume on Motivation and Virtue. New Essays* (p. 80-104). New York: Palgrave MacMillan.

Price, H. (1988). *Facts and the Function of Truth*. Oxford: Blackwell.

Prinz, J. (2015). An Empirical Case for Motivational Internalism. In Björnsson et al. (Ed.), *Motivational Internalism* (p. 61-84). New York: Oxford University Press.

Richard, M. (2015). What Would an Expressivist Semantics Be? In *Meaning Without Representation* (p. 137-159). Oxford: Oxford University Press.

Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.

Ridge, M. (2015). Internalism: Qui Bono? In Björnsson et al. (Ed.),

Motivational Internalism (p. 135-149). New York: Oxford University Press.

Rosa, L. (2016). Rational Doxastic Dispositions and the Epistemic Regress Problem. *dialectica*, *70*(4), 589–607.

Rosati, C. S. (2016). Moral Motivation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/win2016/entries/moral-motivation/`.

Rosen, G. (1998). Blackburn's Essays in Quasi-Realism. *Nous*, *32*(3), 386-405.

Russell, B. (1903). *The Principles of Mathematics*. London: Norton.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson's University Library.

Schiffer, S. (1972). *Meaning*. New York: Oxford University Press.

Schroeder, M. (2004). The Scope of Instrumental Reason. *Philosophical Perspectives*, *18*(Ethics), 337-364.

Schroeder, M. (2008a). *Being For*. Oxford: Oxford University Press.

Schroeder, M. (2008b). Expression for Expressivists. *Philosophy and Phenomenological Research*, *LXXVI*(1), 86-116.

Schroeder, M. (2008c). How Expressivists Can and Should Solve their Problem with Negation. *Noûs*, *42*(4), 573-599.

Schroeder, M. (2010a). How to Be an Expressivist About Truth. In C. D. Wright & N. J. L. L. Pedersen (Eds.), *New Waves in Truth* (p. 282-298). New York: Palgrave Macmillan.

Schroeder, M. (2010b). *Noncognitivism in Ethics*. New York: Routledge.

Schroeder, M. (2011, unpublished). *Attitudes and Epistemics*. University of Southern California.

Schroeder, M. (2013). Two Roles for Propositions: Cause for Divorce? *Noûs*, *47*(3), 409-430.

Schroeder, M. (2015a). *Expressing our Attitudes*. Oxford: Oxford University Press.

Schroeder, M. (2015b). Higher-Order Attitudes, Frege's Abyss, and the Truth in Propositions. In *Expressing our Attitudes* (Vol. 2, p. 55-74). Oxford: Oxford University Press.

Schroeder, T. (2020). Desire. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/sum2020/entries/desire/`.

Schroeter, L., & Schroeter, F. (2017). Metasemantics and Metaethics. In T. McPherson & D. Plunkett (Eds.), *The Routledge Handbook of Metaethics* (p. 519-535). London: Routledge.

Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Nous*, *36*(2), 249-275.

Schwitzgebel, E. (2013). A Dispositional Approach to Attitudes: Thinking

Outside of the Belief Box. In *New Essays on Belief* (p. 75-99). New York: Palgrave MacMillian.

Schwitzgebel, E. (2019). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/fall2019/entries/belief/`.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Setiya, K. (2018). Intention. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/fall2018/entries/intention/`.

Shafer-Landau, R. (2003). *Moral Realism. A Defence*. Oxford: Clarendon.

Silk, A. (2015). How to be an Ethical Expressivist. *Philosophy and Phenomenological Research*, *XCI*(1), 47-81.

Simpson, M. (2018). Solving the Problem of Creeping Minimalism. *Canadian Journal of Philosophy*, *48*(3-4), 510–531.

Sinclair, N. (2006). The Moral Belief Problem. *Ratio*, *XIX*, 249-260.

Sinhababu, N. (2017). *Humean Nature*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (2006). *Moral Skepticisms*. Oxford: Oxford University Press.

Smith, M. (1994a). Minimalism, Truth-Aptitude and Belief. *Analysis*, *54*(1), 21-26.

Smith, M. (1994b). *The Moral Problem*. Malden: Blackwell Publishing.

Smith, M. (1994c). Why Expressivists about Value Should Love Minimalism about Truth. *Analysis*, *54*(1), 1-11.

Soames, S. (2013). Cognitive Propositions. *Philosophical Perspectives*, *27*(1), 479-501.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (Second edition 1995 ed.). Oxford: Blackwell.

Staffel, J. (2019). Expressivism, Normative Uncertainty, and Arguments for Probabilism. In *Oxford Studies in Epistemology* (Vol. 6, p. 161-189). Oxford: Oxford University Press.

Stalnaker, R. C. (1984). *Inquiry*. Cambridge: MIT Press.

Steinberg, J. (2009). Weak Motivational Internalism, Lite: Dispositions, Moral Judgments, and What We're Motivated to Do. *Canadian Journal of Philosophy*, *35*, 1-24.

Stevenson, C. L. (1937). The Emotive Meaning of Ethical Terms. *Mind*, *46*(181), 14-31.

Stevenson, C. L. (1945). *Ethics and Language*. New Haven: Yale University Press.

Stoljar, D. (1993). Emotivism and Truth Conditions. *Philosophical Studies*, *70*(1), 81-101.

Strandberg, C. (2012). Expressivism and Dispositional Desires. *American*

Philosophical Quarterly, *49*(1), 81–91.

Streumer, B. (2017). *Unbelievable Errors. An Error Theory about All Normative Judgements.* Oxford: Oxford University Press.

Svavarsdottir, S. (1999). Moral Cognitivism and Motivation. *The Philosophical Review*, *108*(2), 161-219.

Swartzer, S. (2015). Humean Externalism and the Argument from Depression. *Journal of Ethics and Social Philosophy*, *9*(2), 1-17.

Thompson, M. (2008). *Life and Action.* Cambridge: Harvard University Press.

Tiefensee, C. (2010). *Expressivism, Minimalism, and Moral Doctrines* (dissertation, University of Cambridge). Retrieved from `https://philarchive.org/archive/TIEEMA`

Toppinen, T. (2015). Pure Expressivism and Motivational Internalism. In Björnsson et al. (Ed.), *Motivational Internalism* (p. 150-166). New York: Oxford University Press.

Tuomela, R. (1977). *Human Action and its Explanation.* Dortrecht: Reidel Publishing Company.

Unwin, N. (1999). Quasi-Realism, Negation and the Frege-Geach Problem. *The Philosophical Quarterly*, *49*(196), 337-352.

Urmson, J. O. (1968). *The Emotive Theory of Ethics.* London: Hutchinson.

van Roojen, M. (2015). *Metaethics. A Contemporary Introduction.* New York: Routledge.

Wedgwood, R. (2010). Schroeder on Expressivism: For - or Against? *Analysis*, *70*(1), 117-129.

Wedgwood, R. (2017). *The Value of Rationality.* Oxford: Oxford University Press.

Whiting, D. (2010). Should I Believe the Truth? *dialectica*, *64*(2), 213-224.

Whiting, D. (2013). Truth: the Aim and Norm of Belief. *Teorema: International Journal of Philosophy*, *32*(3), 121-136.

Wittgenstein, L. (1953). Philosophical Investigations. In Anscombe et al. (Ed.), *Philosophical Investigations.* Malden: Wiley-Blackwell.

Wright, C. (1992). *Truth and Objectivity.* Cambridge: Harvard University Press.

Zangwill, N. (2015). Motivational Externalism: Formulation, Methodology, Rationality, and Indifference. In Björnsson et al. (Ed.), *Motivational Internalism* (p. 44-60). New York: Oxford University Press.