# Expressivism and Moral Epistemology

José Rodrigo Valencia Pacheco

Submitted in accordance with the requirements for

the degree of Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

September 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

To Rafael Pacheco Jimenez, in loving memory. Your evident moral wisdom but complete lack of any expressive character may well be a counterexample to this whole project.

# Acknowledgements

The moments when I felt getting a PhD was such an unreachable achievement, the one thing that kept it reasonable to believe it was possible was to think of all the incredible help I have received.

First, I feel lucky and thankful to have had three outstanding supervisors in Leeds. Pekka Väyrynen, my main supervisor, was incredibly patient and encouraging, took every part of my unreasonably ambitious project seriously, and held my work to a high standard that improved every aspect of it —and every aspect of my approach to philosophy at the same time. Thank you very much for everything, Pekka. Rach Cosker-Rowland joined the project later, but her interest, encouragement, and less abstract approach to questions in moral epistemology were just what I needed to carry it to the finish line. Thank you, Rach. Lastly, Daniel Elstein was my co-supervisor for the first couple of years, and I have an incredible debt to him. He was excited about my project when even I was not, and although I knew a few things about expressivism before coming to Leeds, he always insisted on teaching me the most important one: how to think like an expressivist. Thank you, Daniel.

When I applied for my PhD in Leeds, I knew it would be a great place to carry out my project. Yet, I had no idea how rich the discussion on the topics I was working on would be there or the opportunities I would get to talk about my project with so many excellent philosophers working on the same or adjacent topics. Thanks to Jack Woods and Neil Sinclair for agreeing to be my examiners and for a great *viva*. And for very helpful conversations, thanks to Graham Bex-Priestly, Matthew Chrisman, John Divers, Will Gamester, Camil Golub, Sebastian Köhler, Michael Ridge, Mark Schroeder, Christine Tiefensee and Teemu Toppinen. Beyond direct interactions, I must say that your work has always made expressivism such an interesting and rich research program, and getting to discuss my ideas from this thesis with you was such a privilege; thank you very much for that.

Five years ago, it was obvious that if I were going to do a PhD, it would be on metaethics. But I would have never believed them if somebody told me it would be on expressivism and moral epistemology. I thought that expressivism, as a metaethical theory, was hopeless, and I, as an epistemologist, even more. Thanks to Matt Bedke and Jonathan Ichikawa for showing how mistaken I was when I was an MA student at UBC. And thanks also for their excellent suggestions that shaped

many aspects of the project when it was only a vague idea. Gustavo Ortiz Millán and Moises Vaca Paniagua were my first metaethics teachers, and this PhD would not have been possible without their constant support from when I was an undergrad to the point when I was applying to Leeds. But most importantly, I hope the many valuable lessons they taught me about philosophy are still somewhat recognizable here —they are, for me!

One of the best parts of working on philosophy has been meeting great colleagues and friends in other philosophy students. I owe many thanks to the Leeds PGR community. Of special mention are Simon Graf for his friendship, for reading parts of this dissertation and the many conversations we had about some of these topics; Jakob Donskov for his incredible encouragement and reassuring philosophical agreement; Sam Mason for being such a great company when the Leeds PGR metaethicists turned into a dying breed; Jyl Schacher for being an awesome friend; and Alex Brehier-Stamatis, Ludivica Adamo, Maddy Page and Matt Clark for their friendship and putting up with my constant talk of expressivism and anti-realism in different seminars and reading groups —rest assured many of the ideas you had to listen to did end up here. Thanks also to Keiran Bateman, Miriam Bowen, Bumbăcea, Pei-Lung Cheng, David Herring, Olof Leffner, and Andreas Bruns. My Mexican philosophy friends have been a constant source of support throughout my academic journey, and my work in this dissertation was no exception. Thanks to Daniel "el inmortal" Garibay Garcia and Jose Navarro Talavera for being my go-to philosopher of language and epistemologist, respectively, and for reading parts of this dissertation with interest and care. Thanks to Afra M. Montero Ríos, Raul Ibarra Herrera, Alicia Flores Morales, and Aliosha Barranco Lopez for all your encouragement, support, and inspiration. And thanks to Z for the constant supply of music and means to listen to it. Doing philosophy is sometimes (or often!) a bleak, disheartening and difficult task, but doing philosophy *with you* all has always been fun and exciting. Thank you very much for that.

I have no way to capture here the massive amount of gratitude I feel for my wife, Caroluna. Thank you for following me to where my academic life took me, for letting me give you a space to find your own vocations, and for helping me find a life that we could love no matter the place and circumstances. But most importantly, I often find it hard to believe that one can do academic philosophy without a Carolina that makes them feel appreciated for their work at the right times but also does not take philosophical problems too seriously, so can give you a less suffocating perspective

on them; thank you so much for being my Carolina throughout all these wonderful years.

None of this thesis, my PhD, or my academic career so far would have been possible without the enormous support from my family. My PhD was the first time some of us would even hear something like this was possible and a relatively viable career option, yet you always trusted, encouraged, and supported me. The biggest thanks are, of course, to my mom, Ana. She made every accomplishment writing this dissertation feel more meaningful, and every obstacle feel more manageable. Like Gustavo Cerati said, “Mamá sabe bien. Tarda en llegar, y al final, al final hay recompensa.” Almost as significant is the thanks to all the rest of my family in Las Flores: my aunts, Alma, Paty, and Lupe; to Sergio and Ceci, to my grandparents Rafael and Josefina; and all the cousins, Vuzi, Jatziri, Mariana, Diana, and Oscar. A massive thanks also to my mother-in-law, Carolina, and my brothers-in-law, the cuñados, Beto and Bastian. It feels unpardonably unfair that for all your support over these years, you deserve so much from me, and all I keep giving you is mentions in philosophy dissertations, but here is one more!

Lastly, many thanks to the Mexican National Council for the Humanities, Sciences, and Technologies (CONAHCyT) for the financial support during all my PhD studies.

# Abstract

Expressivism is the view that the primary function of moral thought and talk is practical as opposed to representational. In this thesis, I offer the first systematic study of the prospects and problems of an expressivist moral epistemology. The thesis is divided into two parts. The first one is about the challenge of epistemic evaluability: explaining how moral judgments can be both desire-like in their practical function and belief-like in their epistemic evaluability. In Chapter 1, I offer a new way to understand the challenge based on how the epistemic evaluability of states like beliefs is explained by their constitutive norm of aiming at the truth. Chapters 2 and 3 are about possible solutions. In Chapter 2, I explore the prospects of indirect explanations of epistemic evaluability. While I do not think these views work, they teach us valuable lessons about how expressivists should think of epistemic evaluability. In Chapter 3, I offer a different direct response.

The second part of the dissertation is about epistemic evaluations. In moral practice, we evaluate moral judgments as justified, rational, warranted, counting as knowledge, etc. Instead of explaining each of those, I focus on some of the central concepts we use or assume when making these evaluations: evidence, reliability and a prioricity. In Chapters 2 and 3, I argue that expressivists can explain how evidence and reliability work as applied to moral judgments by securing a non-accidental connection with truth. In Chapter 4, I change the focus to concepts central to the sources of those epistemic evaluations. Using resources developed in other parts of the thesis, I argue that expressivists do not need to deny and can explain how some of the sources that ground our epistemic evaluations for moral judgments are a priori.

# Contents

"The deepest thing in our nature in this Binnenleben (as a German doctor lately has called it), this dumb region of the heart in which we dwell alone with our willingnesses and unwillingnesses, our faiths and fears. As through the cracks and crannies of subterranean caverns the earth's bosom exudes its waters, which then form the fountain-heads of springs, so in these crepuscular depths of personality the sources of all our outer deeds and decisions take their rise. Here is our deepest organ of communication with the nature of things."

—Willam James, *Is life worth living?*

# General Introduction

In his last discourse, Mexican revolutionary Ricardo Flores Magón (1918) protested that one of his friends, Raúl Palma, was arrested because he *understood* it is a principle of social and human justice that every human being has the right to satisfy their basic needs just in virtue of coming to life, and devoted his life to helping other proletariats understand it too. It would not be a stretch to say that 'understanding' here is meant *epistemically*: that there was something about Palma's views that got right what is morally right or wrong. Like Flores Magón, we often think and talk about our moral views in epistemic terms. We talk about *learning* what is the wrong thing to do, not *knowing* or being *unsure* what the morally best decision is or worry about the blame we can place on people in situations of moral *ignorance*. The epistemic status of our views plays a very important role in our practices. For example, it is vital for us to get the best responses to our moral questions of what to do or how to live. The best way we can approach this task is to come up with responses with good epistemic standing, those that we can be certain of, that are epistemically rational, warranted, justified, or even count as knowledge.

The epistemic status of our moral views is the subject of moral epistemology. We can narrowly define moral epistemology as the area of ethics and metaethics that concern the possibility and conditions for moral knowledge, moral inquiry, moral reasoning, moral evidence, the sources of moral knowledge like understanding or testimony, etc.[1] However, when we do moral epistemology, we may be asking two different kinds of questions. On the one hand, one could worry about what is the best way to form our moral views. For example, how to be sure that what we are doing is morally good, the best way to proceed in the face of uncertainty, whether it is permissible to form moral views based on testimony, and the like. We can call the domain of these questions *normative* moral epistemology. That will not be the topic of this dissertation. Instead, think about a different question we may ask about the nature of moral knowledge, justification, warrant, etc. What is it, for example, for a mental state with moral content to be epistemically justified or supported by evidence or what is the meaning of knowledge attributions. We may call this moral *meta*-epistemology. The questions we ask when we do moral meta-epistemology assume responses to complex questions in metaethics: the study of the nature of our

[1] For an overview see McPherson (2020).

moral thought, language and practice. Expressivism, one of the most prominent views in metaethics, intersects with moral meta-epistemology in this way.

Expressivism is difficult to characterize, but most views that fall under that banner share a commitment to a *practical* function for our moral thought and language.[2] More precisely, expressivism, as I will be using the term, refers to a family of three interconnected views. First, *expressivism*, properly speaking, is the view that the function of moral thought and talk is not descriptive but practical. It does not aim to represent the world but to produce and guide action. Most, but not all, expressivists are also *non-cognitivists*; they explain this practical function in terms of language expressing conventionally associated mental states that are practical in how they are appropriately connected with intentional action.[3] We usually conceive of desires as the paradigmatic case of a mental state connected with action and motivation, so non-cognitivist expressivists claim that moral judgments are just desire-like states. Lastly, most non-cognitivist expressivists are also *quasi-realists*;[4] they take as a central goal of the expressivist project to explain how our desire-like moral thought and talk have some surface belief-like features. For example, we talk about *believing* that equality is morally valuable or claims like 'using stereotypes to depict minorities is wrong' as *true* or *false*. Quasi-realist expressivists agree that these are central to moral practice. However, they just appear to be incompatible with their views, they are just surface features of our fundamentally desire-like moral thought and talk, so they can be explained and vindicated but using expressivist-friendly tools. Following the common contemporary practice, I will use 'expressivism' to refer to the view that comprises these three projects and mark the difference when relevant.

[2] Classic examples of expressivism can be found in Ayer (1936), Stevenson (1944), Blackburn (1993, 1998a) and Gibbard (1990, 2003).

[3] There is a controversy of around what the appropriate level of linguistic explanation to fit this claim about expression: semantic (Schroeder, 2008a), meta-semantic (Chrisman, 2011; Ridge, 2014) or pragmatic (Yalcin, 2019)..

[4] These views are closely interconnected but largely independent and can be (and have been) defended separately. For example: Horgan & Timmons (2006) defend what they take to be a cognitivist version of expressivism, Beddor (Forthcoming) articulates a non-expressivist version of non-cognitivism, Yalcin (2019), Bedke (2018) and Frápolli & Villanueva (2012) all defend language centred versions of expressivism that do not need commitments at the level of moral judgments (and so are not non-cognitivists), and García-Ramirez (2018) defends a cognitivist version of quasi-realism. Here I will focus almost exclusively on non-cognitivist quasi-realist versions of expressivism. The reason for not including non-non-cognitivist versions is that expressivist views that have no commitments at the level of moral judgment are less likely to have any substantial connections with moral epistemology, and cognitivist versions of expressivism or quasi-realism do not face the distinctive problems that I will be discussing.

There are simple but substantial ways to connect expressivism with moral epistemology. We use or at least assume a variety of epistemic concepts when engaging in moral practice. From that point of view, we care about whether we *know* what the right thing to do is, how *sure* we are about our moral convictions, whether we can *trust* the moral views of our peers, etc. Expressivists, given their quasi-realist commitments, want to explain and vindicate these elements of moral practice. Thus, when it comes to these areas of moral practice, expressivism and moral epistemology have a common subject matter. It makes sense that expressivists want a moral epistemology because, in this sense, that is just part of their project. However, while expressivism is a thriving research programme, expressivists have provided little in the way of moral epistemology. Although expressivists have offered interesting and substantial proposals, they are unsystematic and often an afterthought to their more developed views on moral thought and language. My thesis remedies this; it develops two central elements of the foundations of an expressivist moral epistemology. First, I offer a novel interpretation of and solution to the challenge of epistemic evaluability: explaining how moral thought can be desire-like yet apt for epistemic evaluations. Second, I explain how an expressivist moral epistemology can account for notions central to epistemic evaluations: evidence, reliability, and a prioricity.

The first part of the dissertation will be about the challenge of epistemic evaluability. Like most other challenges to expressivism, it is about the way moral judgments, while desire-like, have distinctive belief-like features. The usual example here is that moral judgments are truth-apt, like beliefs. For instance, I think of my judgment that depicting minorities with stereotypes is wrong as true, just as I talk about my belief that Leon Trotsky died in Coyoacán as true. But desires are not like this. Desires may be accurately or truly ascribed to us. For example, it may be true that I want to learn German, but that is different from my desire being true. We can think about it this way: a belief is true if its content is a true proposition, but a desire with a true proposition as its content does not seem like the kind of thing that can be true or false. Other propositional functions of beliefs, like being embeddable in logically complex constructions, raise similar challenges. The challenge of epistemic evaluability takes a similar path.[5] We generally consider beliefs apt for attributing epistemic statuses: they can be epistemically justified, warranted, rational and even count as knowledge. Moral judgments are like this, too. Thinking that I know that

[5] The challenge was first introduced explicitly by Bob Beddor (2019). Although the idea that non-cognitivism is incompatible with epistemic evaluations has been a common assumption in moral epistemology for a while now. I will explain this in more detail in Chapter 1.

racism is wrong or that I am epistemically warranted in thinking so does not seem to be a conceptual confusion. But desires are very different in this respect. We can talk about desires being justified or rational, but in *practical* rather than *epistemic* terms, and it seems to be just a category mistake to think of desires as counting as backed by evidence or counting as knowledge.

The first chapter is about how to articulate the challenge. A proper challenge is based on a property of beliefs that both explains their epistemic evaluability and separates them from desire-like states. If moral judgments, being desire-like states, lack this property, then expressivism would be in trouble. I propose that the best candidate for this property is the constitutive norm that beliefs aim at the truth. This makes beliefs epistemically evaluable because epistemic evaluations plausibly correspond to means to get true beliefs. I call this picture of epistemic evaluability *veritism*. Veritism can give us a compelling challenge for expressivists since it is unclear how a similar norm of aiming at truth can regulate moral judgments given that they are desire-like states. Ordinary desires, or even desire-like attitudes like intentions, do not aim at truth; thus, it is difficult to see how moral judgments could have that aim either. Expressivists owe us an explanation. Veritism relies on a picture of beliefs and epistemic evaluations that some expressivists may want to resist. Thus, after establishing this version of the challenge, I deal with possible objections. The following two chapters are about ways to respond to it.

Truth is central to the challenge of epistemic evaluability. And when accounting for truth, expressivists often take an indirect route. Instead of explaining what it is for moral judgments to be true or false, they explain what it is to *think* or *talk* about them as true or false. Expressivists argue that in this way, we can realize that truth is not the kind of substantial property that they need to explain, and so there is nothing incompatible between their picture of moral judgments as desire-like and their truth-aptness. This strategy is sometimes called *side-stepping*. There are three ways to articulate a similar side-stepping account for epistemic evaluability. In Chapter 2, I explore their prospects and problems. Ultimately, I do not think that they work. But there is a lot they can teach us about how expressivists think about the issues surrounding epistemic evaluability, and they will inform what I take as the best response.

Drawing lessons from the three side-stepping strategies, I proceed in Chapter 3 to develop a direct, substantial expressivist account of epistemic evaluability. The proposal has three parts that correspond to the three main elements of veritism: truth,

regulation and evaluation. The idea is that correct moral judgments, those we want to get out of moral inquiry, have the crucial truth-like property of being stable through improvements of coherence, sensitivity to finer details, imagination, maturity, etc. This stability property will work as our proxy for truth. Moral judgments are epistemically evaluable because they aim at the truth by aiming at stability. However, this cannot be the whole story. Desire-like states, in general, do not aim at stability; they have their own form of regulation. But expressivists still have good reasons to think the norms of desire-like states regulate moral judgments. Thus, I argue that the aim at stability for moral judgments is just a specific version of a general norm of desire-like states of aiming at the good or reasons. Lastly, combining these two elements can establish a close parallel to veritism and help expressivists solve their problems with epistemic evaluability.

Having discussed matters of epistemic *evaluability* and how expressivists should address them, in the second part of the dissertation, I will explore how expressivists should make sense of the central notions involved in the epistemic *evaluations* of moral judgments. This is a complex task given the variety of epistemic evaluations we make in moral practice —as shown by the different examples presented so far. The way I will approach this task, then, is by picking examples of concepts central to our epistemic evaluations: evidence, reliability and a prioricity. These will give us a comprehensive and representative case that expressivists can have adequate accounts not just of epistemic evaluability but of epistemic evaluations. In the introduction to part II, I will offer a complete overview of the reasons for these examples and the general strategy to tackle them.

This is the first systematic study of moral epistemology for expressivism, so the project is ambitious. Thus, it is important that I am clear on its limitations. First, I approach some of these issues, like epistemic evaluations, by exploring how expressivists can make sense of some central examples. But given the space constraints, I will leave out many others, like justification, knowledge or epistemic rationality. Some of these are already (at least partially) addressed in the existing literature.[6] We could think of my proposals here as broadening the case for an expressivist view of epistemic evaluations. But more substantially, I hope the examples I picked will help offer a deeper foundation for expressivists to think about other evaluations more systematically. Evidence and reliability, for example, are

[6] See Blackburn (1996), Gibbard (2003) and Ridge (2007) for knowledge, Lenman (2007) for justification, and Hayward (2017) for rationality.

helpful in fixing facts about justification, knowledge, rationality, certainty, etc. In this same vein, there are interesting and relevant questions about how expressivist moral epistemology interacts with views on normative epistemology. For example, whether expressivism pairs better with some form of virtue epistemology or whether it is compatible with the possibility of moral testimony. I will have very little to say about that further issue.

Second, the most central focus of contemporary expressivists has been on language-oriented issues. While I will have nothing to say about those, I often assume that expressivists have plausible responses to them. This is because, while most of these issues are outside the scope of the thesis, they have implications for what I say here. The most notable example here is the Frege-Geach problem —the problem of showing how expressivists can account for moral contents or propositions. This means that expressivists would be able to make sense of complex truth-functional conjunctions with moral contents like negations, conditionals, conjunctions, etc.[7] It is common to think of propositions not only as the meaning of claims in contexts and the contents of propositional mental states (like moral judgments presumably are[8]) but also as the primary objects of epistemic evaluations. That expressivists can respond to this challenge may seem like a substantial and problematic assumption. However, I believe that part of what involves remedying the little attention moral epistemology has received from expressivists is to set these issues aside.

A different but equally prominent aspect of contemporary expressivism that I will have little to say about is hybrid versions of expressivism. To overcome challenges like the Frege-Geach, some expressivists have expanded their accounts so that moral judgments are actually complex mental states composed of cognitive and non-cognitive elements.[9] Moral language then expresses these complex mental states. A crucial aspect of these views is that many belief-like features of moral thought and language have relatively straightforward explanations: moral thought is partly belief-like, and moral language partly expresses belief-like states. I will not explore the prospects of applying that same strategy to the present issues in expressivist moral epistemology. The challenges and possible solutions that I present in the thesis are not easily translatable into a hybrid framework. And so, trying to incorporate these views would have required its own separate discussion. Nevertheless, since many of the problems I discuss here target the desire-like nature of moral thought and how it

[7] For more on this see Schroeder (2008b) and Woods (2017).
[8] See Köhler (2017).
[9] See for example Bar-On & Chrisman (2009), Ridge (2014) and Fletcher & Ridge (2014).

seems incompatible with epistemic evaluability and evaluations, I hope the framework I offer will be useful for future work on how different accounts of that thought, including hybrid views, can help solve those problems.[10]

[10] Moreover, other tools developed by hybrid expressivists, like Michael Ridge's (2007) hybrid *epistemic* expressivism will be relevant at different points of the discussion.

# Chapter 1: The Challenge of Epistemic Evaluability

In our ordinary moral practices, we often make epistemic evaluations. We talk about *learning* what is the wrong thing to do, not *knowing* what the morally best decision is, or worry about the blame we can place on people in these situations of moral *ignorance*. These are the kinds of aspects of our moral practices that metaethical expressivists aim to explain and vindicate. However, given that most expressivists think that moral thought has a practical or desire-like function that contrasts with the function of beliefs and other epistemically evaluable states, it is a relevant and important question whether they can explain and vindicate these practices of epistemic evaluation. This is the challenge of *epistemic evaluability*. And as I will argue, it is the most fundamental obstacle to an expressivist moral epistemology. Because of what expressivists should want in a moral epistemology, how the challenge relates to existing expressivist proposals and the impact it has on the place expressivism can have in the contemporary landscape. Given this importance, the challenge will be the focus of the first half of the thesis. More specifically, I will devote this chapter to establishing what I take to be the best version of it, clarify its importance, and discuss its relation to existing expressivist projects. The following two chapters will be about possible ways to respond to it.

The structure of the chapter is as follows. Section I establishes the importance of the challenge for an expressivist moral epistemology. Section II addresses some issues with the initial formulation of the challenge to develop a novel and more compelling version based on a view I call *veritism*. Section III presents and responds to objections expressivists may have with this new formulation of the challenge. Veritism gives a central role to the notion of truth for epistemic evaluability. Given how much attention contemporary expressivists have devoted to truth, adapting the tools they developed should be helpful for addressing this challenge. Thus, in section IV, I map two general expressivist strategies to account for truth. These will provide a blueprint for our responses to the challenge of epistemic evaluability in the following chapters.

## I. Expressivism, Moral Epistemology and the Challenge

Expressivists, as non-cognitivists, claim that the function of moral thought is practical or desire-like as opposed to representational or belief-like. Thus, the most central challenge for an expressivist moral epistemology is to explain how moral thought can be both desire-like and epistemically evaluable. An epistemic evaluation can be broadly understood as the ascription of a positive or negative epistemic status,

like constituting knowledge, being justified, warranted, and so on. Epistemic evaluability consists of being apt to be evaluated in terms of one or more epistemic statuses in the way I just described. Beliefs are the paradigm example of an epistemically evaluable state. My belief that Leon Trotsky died in Coyoacán, for example, can be justified, warranted, rational or count as knowledge. If we want to account for epistemic evaluations in an expressivist framework, moral judgments must be epistemically evaluable in the way beliefs are. We need to be able to say, for example, that the judgment that racial discrimination is wrong is warranted, rational, can count as knowledge, etc. However, expressivists claim that moral judgments are different from beliefs and much closer to desires. And we generally do not think of desires as epistemically evaluable. Citing Bob Beddor —who offers the first explicit articulation of this challenge: "[Desires] can be evaluated as rational or irrational, as justified or unjustified. But if so, it seems to be *practical* rationality and justification that's at issue. Moreover, describing a desire as an item of knowledge seems like a category mistake." (2019: 3, his emphasis) Thus, expressivists need to explain how moral judgments can be at the same time desire-like in their connection with action but belief-like in being apt for an epistemic status.[11]

The challenge of epistemic evaluability is fundamental for an expressivist moral epistemology. As explained in the introduction, moral epistemologists have more than one way to approach our moral practices. First, consider the questions a moral epistemologist may ask as a participant of moral practice, like how we should epistemically justify our moral judgments, what conditions a state should meet to count as moral knowledge, etc. In the first instance, these would not be questions about the nature of our practices of justification and moral knowledge. Instead, these are questions in what we may call *normative* moral epistemology. This part of moral epistemology concerns how to conduct our ordinary practices that involve epistemic concepts and questions. Perhaps there are interesting connections between normative moral epistemology and expressivism, but expressivists are not primarily concerned with telling us how to conduct moral inquiry.[12] Instead, expressivists want to make sense of the tools participants of moral practice use to think and talk about those

[11] This focus on the expressivist conception of moral judgment is why Beddor presents it as a challenge specifically for non-cognitivism. As explained in the introduction, I talk about expressivism more broadly instead.

[12] Perhaps there are ways in which it does not make sense to conduct moral inquiry given that the main function of our thought and talk is not fundamentally representational. More generally, understanding moral practice should be important for how we conduct it. I do not want to deny this. My point here is that drawing these lessons can only make sense once we established what expressivist should say about moral metaepistemology, and so that should be our primary concern.

questions: what are the mental states involved in thinking of certain moral views as knowledge, what roles do these play in reasoning, what does it mean when we called our moral judgments justified, knowledge, rational, etc. This is part of what we can call moral meta-epistemology. Expressivists do not only share a subject matter with moral epistemology but have specific goals within moral meta-epistemology. A central task for moral meta-epistemology is to make sense of our thinking and talking involved in our first-order practices and theorizing. And a significant portion of those practices consist of evaluating our moral judgments epistemically. Thus, it is crucial even to start that project that our thinking and taking under expressivism is the right fit for those practices. And this is precisely what the challenge of epistemic evaluability puts into question.[13]

A second reason why the challenge is central is that, given their quasi-realist commitments, expressivists have had things to say about epistemic evaluations, but their existing views do not address the challenge. The usual expressivist approach to epistemic evaluations is *epistemic expressivism*. The idea is that epistemic evaluations are just a subset of our normative thought and talk and, thus, should be apt for the same kind of treatment expressivists give to normative moral thought and talk.[14] Mirroring its metaethical counterpart, according to epistemic expressivism epistemic normative language has the function of expressing certain non-representational or desire-like mental states. Allan Gibbard (2003), for example, argues that our knowledge talk expresses mental states like *plans* to rely on someone's judgment —and so epistemic though and talk helps us keep track of patterns of trust and reliance. Simon Blackburn (1996) argues that we should understand attributions of knowledge in terms of taking a judgment to be beyond any plausible revision or improvement, where the right interpretation of our thinking and talking of improvement is expressivist. This view should give expressivists some valuable tools to develop their moral epistemology. However, it is unclear whether epistemic expressivism can address the challenge of epistemic evaluability. If we understand the mental states we express with epistemic evaluations as desire-like, that by itself says nothing about their proper objects and whether they are desire-

[13] There are other aspects of our ordinary practices that can lead to similar worries. For example, many believe that assertions are governed by epistemic norms, like that one should only assert what they know or believe justifiably. And in moral practice we often make moral assertions. When I claim that racial discrimination is wrong, for example, I am clearly not presenting that as a question or an order. Thus, to make sense of moral assertions expressivists need to get clear on their commitments in moral epistemology. See Ridge (2009) for a somewhat similar suggestion.

[14] For a general overview of this approach to epistemic language, see Chrisman (2012).

like. The fact that the state I express when I say that "I know racial discrimination is wrong" is desire-like says nothing about what kind of state the judgment that racial discrimination is wrong could be. Epistemic expressivism might be the right view about the epistemic evaluation of beliefs, but more needs to be said to make it compatible with desire-like states. Thus, existing formulations of epistemic expressivism seem to be orthogonal to the challenge —although I will have more to say about the *potential* of epistemic expressivism for explaining epistemic evaluability in the next chapter.[15]

Third, the explicit formulation of this challenge is recent but has important connections with the existing literature. The challenge helps us put in order and vindicate a general worry about expressivism that has been pervasive in the literature of moral epistemology. Authors trying to sketch the landscape of contemporary moral epistemology often conceive of expressivism not only as not having a view of the matter but as a position that, in essence, cannot have or does not aim to have one. According to them, the expressivist view of moral thought and language based on non-cognitive mental states is supposed to question or at least revise our ordinary practices of associating moral judgments with anything like knowledge or epistemic justification. Authors like Richmond Campbell (2015) and Robert Audi (2011: 380-1) think expressivism is incompatible with the mere idea of moral knowledge. And authors like Aaron Zimmerman (2010: 180-1) and Richard Joyce (2018: 293-7) go as far as to claim that expressivists are committed to the denial of moral knowledge.[16, 17] There is an important point in these views that is obscured by some significant misconceptions about expressivism. Most of these authors seem primarily concerned with the expressivist commitment to non-cognitivism. Thus, we can connect what is important about their views with the challenge of epistemic evaluability: even if expressivists take ideas like moral knowledge seriously, moral judgments conceived by expressivists are still not the right fit for the evaluations we make while engaging in moral epistemology.[18] Expressivist moral epistemology then seems to be based on a category mistake.

---

[15] Beddor (2019: 5) offers a similar argument.

[16] To be clear, the difference between these two positions is that for the former scepticism is an unwelcome result of adopting expressivism while for the later it was intended. The difference is important since the latter is, for most cases, a mischaracterization of expressivism.

[17] Similar claims can be found in Machuca (2018: 4), Sayre-McCord (2013: 10), Pritchard (2018: 142) and McGrath (2019: 4).

[18] Here I am not claiming that every epistemic challenge for expressivism has this general form. Dorr (2002), Egan (2007) and Street (2011) are important examples of challenges that (at least *prima facie*) concede epistemic evaluability for expressivism but raise problems for

A response to the challenge of epistemic evaluability is thus crucial to make sense of the expressivist primary approach to moral epistemology, existing expressivist proposals do not address it, and the challenge unifies some of the main concerns moral epistemologists have with expressivism. Thus, this challenge is the right place to start our project here. But in order to know how to address it and assess our options, we need a more precise articulation.

## II. The Challenge: Veritism

The challenge of epistemic evaluability rests on a claim about the scope of epistemic evaluations: beliefs are paradigmatically apt for them, and desires, being very different from beliefs, are not. However, it is not clear what makes desires and beliefs different in *this* respect. Perhaps one could argue that the difference between these two states is enough to pose a challenge to expressivism since beliefs are the *primary* or the *exclusive* bearers of epistemic evaluations, and so characterizing moral judgments as anything else (desire-like states or otherwise) would make them not evaluable in this way.[19] However, there are other mental states apart from beliefs that seem to be apt for epistemic evaluations, like credences (Moss, 2016) or even things that are not mental states at all, like patterns of inference (Alston, 2005: 5), assertions (Williamson, 1996) or even complex scientific practices (Kitcher, 1990).[20] These are presumably also very different from beliefs, so claiming that desires are too will not be enough to establish the challenge. What a proponent of this challenge needs to pose a real problem for expressivists is a property that (1) beliefs have and desires cannot share, and (2) has a robust connection with epistemic evaluations.[21]

In his initial formulation of the challenge, Beddor (2019: 3-4) argues that the functional role of beliefs makes them epistemically evaluable and different from desires. After all, desires and beliefs are often characterised as having distinct and even contrasting functional roles.[22] There are, however, two problems with this

different parts of the execution: their accounts of reasoning, fallibility and reliability respectively.

[19] This is a view normally associated with Roderick Chisholm (1991).

[20] A different but more controversial example here is linguistic knowledge which some claim requires states other than beliefs to be epistemically evaluable. See Ezcurdia (2004) for a defence of this idea.

[21] Another possible example is Gibbard's (1990: 63) view that there is a sense of warrant we can ascribe to feelings. If we understand feelings as non-cognitive states that would amount to a counterexample to the challenge. As it will become clear later on I agree with many of Gibbard's points, but because of the way I will reformulate the challenge, my strategy to get to them will be somewhat different.

[22] For a classic defence of this distinction see Stalnaker (1984) and for an overview of recent developments see Schwitzgebel (2019).

proposal. First, as helpful as it is to meet condition (1), it is unclear how to connect functional role with epistemic evaluability, as (2) requires. Functionalist theories of beliefs are not primarily concerned with their epistemic status. Instead, they explain how to characterise them as a distinctive mental state, given their possible or actual connections with stimuli, resulting behaviour, and other mental states. Second, epistemic evaluations are normative. To say that a belief is not justified or does not count as knowledge is not merely to describe it, but to say (at least) that it falls short of a certain sort of standard and is a legitimate target of some kind of criticism on that basis. Functionalism has, historically, had a hard time making sense of this normativity.[23] Even expressivists like Gibbard (2005; 2012) and Toppinen (2015) rule out a purely functionalist view of beliefs partly because of this. Expressivists then have reasons to be suspicious of a challenge grounded in a *general sense* of functional role.[24]

I believe that *normativist* views of belief have a better chance to account for their epistemic evaluability. According to normativism, a complete picture of the nature of beliefs should include the norm or norms that regulate them or specify their correctness conditions.[25] The most prominent version of this view is the one that says that beliefs *aim at* truth. There are many different ways to articulate this view, but we can stick to a minimal characterisation according to which a belief is correct if and only if it is true. This characterisation is meant to be minimal in that it does not take a stand on substantial debates. However, it is still meant to be normative in that it offers correctness conditions that capture the constitutive norm of beliefs. Thus, the correct revision, adoption, or discarding of beliefs is, in principle, according to their truth and falsehood. We have then that:

B) A belief that *p* is correct iff *p* is true

Normativism may not be incompatible with functionalism, after all, B) can be part of the norms that specify the functional role of belief (e.g., Engel, 2005; 2007). My claim, however, is that it is this truth norm, however we choose to integrate it into an account of mental states, that can secure a connection with epistemic evaluability.

[23] For arguments in this vein see Velleman (2000) and Shah and Velleman (2005).

[24] As I will go on to explain, I do not rule out the possibility that the challenge relies on a more specific version of functionalism. My claim, however, is that functional role *by itself* is not enough to establish a proper challenge. See Glüer & Wikforss (2013) and Bykvist and Hattiangadi (2013) for discussion of these issues.

[25] For a general overview and discussion of this view see Chan (2013). For some of the main proponents of the view see Velleman (2000) Wedgwood (2002), Shah and Velleman (2005), Gibbard (2005).

A common strategy to explain what is distinctively epistemic about epistemic evaluations is to contrast them with other forms of evaluation; the main examples being pragmatic or practical. Suppose a person threatens to destroy everything I hold dear unless I believe that 2+2=5. Then, it may be rational in the prudential sense for me to do so, but not in the epistemic sense. A natural way to explain the difference is that from the epistemic point of view, we are concerned with truth.[26] This view explains why, for example, we evaluate beliefs in terms of the evidence in favour or against the truth of their content or why inconsistent beliefs are usually deemed epistemically problematic.

Being true, however, is not enough to specify the epistemic status of beliefs. That is why the aiming part of normativism is crucial. As Ralph Wedgwood explains, "even if irrational beliefs can be correct [in being true], the only way in which it makes sense to aim at having a correct belief is by means of having a rational belief." (2002: 276; my clarification). We can generalise his point about rationality to other epistemic statuses. Belief aims at truth, but truth is not transparent to us. Positive and negative epistemic evaluations of beliefs then concern whether or how well they are being regulated by truth: having beliefs formed by unreliable processes is a poor way to achieve truth; having beliefs that are justified by evidence, while not failsafe, is much better, and more so the better justified they are. Epistemic evaluations then *correspond* to our means to achieve the aim of truth, and they can be positive or negative according to how well or badly they fare in that respect.[27] Let us call this way of combining normativism about belief with this view of epistemic evaluations *veritism*.[28] Following condition (2), the property of beliefs of aiming at truth grounds their epistemic evaluability.

Having accounted for (2) the robust connection between beliefs and epistemic evaluations, we can start seeing how to support (1) that desire-like moral judgments differ from beliefs in this respect. Desires do not aim at the truth, so they are unfit for correctness conditions like B). However, veritism is not the view that epistemic evaluations are *necessarily* about beliefs. Truth helps us demarcate epistemic

[26] For a survey of how prevalent this idea has been in epistemology see David (2001, 2014).
[27] Notice that this formulation avoids the main objection against the view that justification aims at truth, by opening the possibility of justified false beliefs. They would just be the ones formed followed the right method and norms but because of other factors still ended up being false.
[28] I borrow this label from Cowie (2019: Ch. 7). Although he meant it as a view about epistemic norms in general whereas here, I use in a more restricted sense for the norm involved in epistemic evaluability. This view also owes much to Nolfi (2015) but differs in some crucial respects.

evaluations because from the epistemic point of view we want, so to speak, an accurate picture of the world. A belief that *p,* unlike other representational states like imaginings or suppositions, involves a commitment to the truth of *p*.[29] And we can plausibly think of this commitment as one about *p* being part of our accurate picture of the world. Beliefs are not just instrumental in getting to our epistemic goal of truth but an essential part of the goal itself; they make up our accurate picture of the world. Thus, this picture of epistemic evaluability gives beliefs a central role in our practices of epistemic evaluability but it does not follow that there are *no* other objects apt for them.

There are at least two ways to explain other proper objects of epistemic evaluability. By developing these two options we can make sense of the problem for expressivists since they are both, in principle, closed to them. My aim here then is not to defend or develop any of these views in detail, what we need from them is just a clear view of the options for expressivists to sharpen the challenge. On the one hand, some objects appear epistemically evaluable via the epistemic evaluability of beliefs. For example, some argue that assertions are connected with the doxastic states of the utterer via their own norms. They do not aim at the truth. But to correctly assert that *p,* the speaker must have a doxastic state that *p* with good epistemic standing. In asserting that 'Leon Trotsky died in Coyoacán', I present myself as, for example, knowing it or believing it justifiably. This can explain why assertions seem to be epistemically evaluable. For every correct assertion that *p,* there must be a positive epistemic evaluation for a belief that *p*.

Expressivists may want to argue that, like assertions, desire-like moral judgments only *appear* to be epistemically evaluable.[30] That option is problematic since expressivists want to vindicate our ordinary moral thought and talk, which seems to be committed to moral judgments being and not just appearing to be justified, rational, warranted, etc. Additionally, following the view just sketched, correct assertions that *p* depend on the epistemic status of a doxastic state that *p*. To get a complete parallel, moral judgements that *p* would then appear to be epistemically evaluable because they depend on a separate doxastic state that *p*. If this *p* is the same content, then moral thought is, at some more fundamental level, belief-like: it requires a belief with a normative content *p* to play an important explanatory role.

[29] See Shah and Velleman (2005: 497) for the role of this claim in normativism.
[30] Beddor's (2019) own proposal seems to follow a similar strategy. He argues that moral judgments are epistemically evaluable because they are *grounded* on beliefs. However, since he as a different understanding of epistemic evaluability, my objections here do not apply to his view.

For a judgment that racial discrimination is wrong to be epistemically evaluable, like assertions, there must be a different attitude with the content that racial discrimination is wrong, which *is* epistemically evaluable. This is a conclusion that expressivists must resist if their view is not going to collapse with a form of cognitivism. But if this is not the same content, then it is unclear why moral judgments can draw from the evaluations of those further attitudes. Either way, this alternative does not seem helpful for expressivists.

On the other hand, some other candidates for epistemic evaluation may be apt for an explanation similar enough to veritism to warrant epistemic evaluations of their own. Two possible examples are patterns of inference and credences. Patterns of inference often operate over representational contents. And even if they do not aim at the truth, they plausibly aim at truth-preservation (when they are deductive) or truth-ampliation (when they are inductive). We have then a form of regulation similar to that of beliefs. Non-truth-preserving patterns ought to be revised or discarded. Credences are an interesting example. It is controversial whether they aim at truth. But even other candidates for that aim, like accuracy, suggest a picture similar to veritism. If credences aim at accuracy, they plausibly also make up our accurate picture of the world. In this way, they can have correctness conditions similar to B) which warrant similar forms of epistemic evaluation.[31]

Expressivists could offer an account of moral judgments that follow veritism closely enough to warrant epistemic evaluations. However, credences and patterns of inference have aims close enough to truth to have a plausible analogy, whereas desires and desire-like states do not. Some expressivists may want to argue that desires and desire-like states also have a constitutive norm. Perhaps they aim at the good or at reasons. However, aiming at goodness and aiming at truth, *by themselves*, presumably offer very different forms of regulation, so the evaluations they ground are likely to be very different.[32] Moreover, this gets expressivists into a further dilemma. Perhaps moral judgments have a constitutive norm similar to aiming at truth, like beliefs, but then expressivists would have to explain how moral judgments

[31] See Pettigrew (2016) for the aim of accuracy. Moss (2018) suggests a more radical alternative according to which credences are *all* we need to make sense of our main epistemic notions, including the goal of an accurate picture of the world. In that case since credences are more fundamental the correctness conditions for expressivists to try to mimic are not based on truth but on accuracy. Since Moss' view is controversial I will not explore that possibility.

[32] The restriction here is important. As I will argue in chapter 3, the best response to the challenge for expressivists is that given certain assumptions about moral practice the aim of goodness for moral judgments results in an aim for a proxy for truth.

somehow lose their desire-like constitutive norm while remaining desire-like. Instead, expressivists could argue that moral judgments keep their desire-like constitutive norm, but then they would have no clear way of explaining their epistemic evaluability. Thus, the options of explaining moral judgments as having a similar aim to truth or as an appearance do not look promising; desire-like moral judgments seem just unfit for the kind of correctness conditions they need for epistemic evaluability. We have then a compelling version of the challenge.

## III. Expressivist Objections to Veritism

Let us summarize our discussion so far. Veritism offers a way to establish the challenge of epistemic evaluability. It shows why beliefs are (2) epistemically evaluable and (1) different from desires in that respect. Beliefs aim at truth, and epistemic evaluations stand for the good or bad means to attain that end. Moral judgments are epistemically evaluable like any other belief. The problem for expressivists is reconciling this picture of epistemic evaluability with their view of moral judgments as desire-like states. Desires do not aim at truth, so it is hard to see how they can merit the same kind of regulation and, thus, evaluations. Then, expressivists should either explain how moral judgments could be desire-like and epistemically evaluable by sharing an aim similar to truth or resist the challenge. For this section, I want to address the most pressing worries they may have with veritism and so give some reasons for expressivists not to opt for the second option.

### III.1 Objections to Normativism

Expressivists may want to reject the normativist picture of beliefs as aiming at true. Even if it is an attractive and commonly accepted view, it is not uncontroversial. This worry could be understood in two ways. One is a worry that even if beliefs should be characterized according to their constitutive norm or aim, truth should not be that aim. Other relevant options are that beliefs aim at knowledge, justification or understanding.[33] There are two things to say in response. First, the view here is not incompatible with those other options. We can think of veritism as a model for how to connect views about the nature of states like beliefs with epistemic evaluations. And thus, if the nature of beliefs is such that they aim at other things like knowledge, then a plausible way to connect that with epistemic evaluations is to think of them as standing for the means to achieve that goal. Second, it is unclear whether appealing to these alternatives will help expressivists solve their problems with

[33] For knowledge see Williamson (2000) and McHugh (2011). For justification see Feldman (2002). And for understanding see Kvanvig (2003).

epistemic evaluability. Desire-like states do not aim at knowledge, justification or understanding either —if anything, as I will argue, aiming at truth should be an advantage for expressivists since they have had many interesting things to say about truth as applied to moral judgments.[34]

Second, perhaps the worry is with normativism more generally. However, that seems unlikely to appeal to many expressivists. Some, like Gibbard (2005; 2012), accept and defend normativism about beliefs. And although expressivists have not explicitly endorsed a normativist view of moral judgments, the fact that they are subject to norms is an important element of their views. For example, Blackburn (1998a: 126) explains that desires (or desire-like attitudes in general) are not isolated and that putting them together with other mental states can lead us to change, abandon or reinforce them. They are holistically regulated. Gibbard (1990: 204) for his part argues that desires are subject to higher-order norms that govern their acceptance or rejection. Moreover, non-cognitivism is based on an essential difference between desires and beliefs, and to explain it, expressivists appeal to a Humean division of the mind in terms of *direction of fit*.[35] According to that distinction, beliefs are characterized by how their content is supposed to match the world and desires by how the world is (in a sense) supposed to match their content. Truth as the aim of belief and its characteristic direction of fit seem to be importantly linked. For example, according to Timothy Chan, "to say that a belief is to fit the world seems to express essentially the same idea as saying that belief aims at truth —for a belief to succeed in representing the world as it is, amounts to just the same thing as for the belief to be true" (2013: 2).[36] Just like aiming at truth is a normative notion, the belief side of the direction of fit distinction suggests a normative reading. Beliefs that do not match the world are *incorrect*. In that way, by appealing to veritism to articulate the challenge, we do it in terms expressivists have good reasons to accept. Critics often argue that the notion of 'direction of fit' is too vague or too metaphorical to play any important explanatory role.[37] Expressivists have reasons to

[34] A possible disadvantage for those views is that knowledge, justification and understanding are all epistemic evaluations themselves, and so we cannot explain them as corresponding with the means to attain the aim of things like beliefs. For example, it seems unlikely that the means to attain a justified belief will themselves be properly characterized as justification. Thus, whatever other virtues those other accounts have, for an account of epistemic evaluability, veritism seems to have more explanatory power.

[35] See for example Smith (1994: 92) and Blackburn (2005: 323). For an overview of the idea of direction of fit and its problems see Schueler (2013).

[36] Kvanvig (2014: 122) assumes a similar relation between both notions. See also Ortiz Millan (2018) for a proposal for substituting the idea of direction of fit with the idea of normative aims.

[37] See for example Sobel & Copp (2001).

favour views like normativism that flesh out and support the distinction they already accept.

Additionally, we are working on a minimal characterization of normativism, according to which the norm of aiming at truth is essential for the characterization of beliefs as epistemically evaluable. Still, it is neutral enough to fit the most plausible substantial views of mental states that expressivists already accept. I already suggested that normativism can be compatible with functionalism, given that the truth norm can be part of what specifies the functional role of beliefs. For example, an observational belief can have an input condition according to which it is correct only when there is a sensory experience that can count as evidence for the truth of the belief. Normativism, in this sense, is also compatible with intepretationism, according to which we can determine an agent's mental states based on what their best rationalising interpretation would attribute to them.[38] The truth norm can be understood as part of the norms determining the rationalizing interpretation.[39] While not exhaustive, these two examples are representative. Thus, normativism, at least as it figures in veritism, should not involve any problematic commitments to the metaphysics of mental states that expressivists would be inclined to reject.

A related worry expressivists may have is that while they want to keep a contrast with non-moral beliefs, there is a sense in which they do not want to deny that moral judgments are also beliefs. Perhaps they hold minimalist views according to which a belief is just the kind of mental state apt to be expressed in an assertoric claim. And moral judgments can clearly be expressed like this. Or perhaps they argue that mental states can be aptly called beliefs in a non-minimalist sense, even when they lack a primarily representational function.[40] Either way, moral judgments are just a subclass of beliefs, and so, like any other belief, they should be epistemically evaluable. While I agree that expressivists should treat moral judgments as a subclass of beliefs, the challenge does not depend on whether expressivists are committed to doing that but on whether they have earned the right to do so yet. We can think about it like this: the challenge of epistemic evaluability is an instance of the more general challenge for expressivists of how we get to treat moral judgments, being desire-like or non-

[38] See Lewis (1974) and Williams (2019).

[39] Thanks to an anonymous reviewer for Ergo for suggesting these points.

[40] For the minimalist sense see Wright (1992: 14) and Dreier (2004: 16-9). For the more robust but non-representational sense of belief see Schroeder (2013) and Brown (2021). Even proponents of the epistemic evaluability challenge, like Beddor (2019: 1-3) endorse a view like this.

representational, as beliefs. And unless they explain how they are governed by a norm like B) or offer an alternative explanation of their epistemic evaluability, they have not yet fully earned that right. Thus, a helpful stipulation for the rest of our discussion in this and other chapters is that I agree that moral judgments are a subclass of beliefs but will keep referring to them as moral judgments to signal that the discussion here is part of the project of earning the right to refer to them as beliefs.

### III.2 Objections About Epistemic Evaluations

Apart from normativism, veritism includes a view of epistemic evaluations as corresponding to the means to attain truth. This is sometimes labelled a teleological view of epistemic evaluations.[41] A problem some expressivists could have with the overall teleological framework is that it seems too similar to epistemic consequentialism and thus inherits its problems. Epistemic consequentialism can be roughly understood as the view that epistemic norms of justification rationality or knowledge should be understood in terms of conduciveness to the maximization of epistemic goods, like true beliefs.[42] One significant problem with this view is that if the aim is to get the epistemic good of true beliefs, there may be problematic trade-offs between believing falsehood in exchange for more true beliefs (see Berker, 2013: 350-7). According to epistemic consequentialism, believing that falsehood is rational and thus has a good epistemic standing, but that seems problematic. In response, while I am assuming some form of teleology, given that epistemic evaluations stand for the means to attain true beliefs, I am not assuming epistemic consequentialism. On the one hand, to establish the challenge, I only need epistemic evaluations to correspond to the means to the truth of *individual* beliefs and moral judgments, not the overall balance.[43] On the other hand, I am *not* assuming a maximizing condition. Thus, what I say here about epistemic evaluations should be neutral concerning those problems for consequentialism.

Characterizing epistemic evaluations as standing for the means to attain truth separates epistemic from other forms of evaluation, like practical or pragmatic. Some

[41] Although the label is more generally used for epistemic norms. See Friedman (2019).

[42] For this characterization and an overview of epistemic consequentialism see Ahlstron-Vij & Dunn (2018: 2). For a way to separate a teleological view of epistemic evaluations from consequentialism, in the way I am suggesting here, see Littlejohn (2018).

[43] Perhaps one could say that the resultant balance of true beliefs should matter for the epistemic status of individual beliefs. For example, a perceptual belief that there is a computer screen in front of me is reliable given the truth ratio of true beliefs generated by perception in normal conditions. I will have more to say about the role of other beliefs and moral judgments in the epistemic status of individual judgments in Chapter 4 and about reliability in chapter 5.

expressivists may reject this distinction and opt for a form of *pragmatism*. According to this view, states like beliefs are apt for both pragmatic or practical evaluations in the same way they are apt for epistemic evaluations, which could, in principle, blur the distinction between them.[44] For example, Blackburn (1984: 224) seems sympathetic to that idea. This may be helpful for expressivists since we know that desire-like states are apt for practical or pragmatic evaluations. For example, my desire to get some rest may be justified, given practical considerations that that is what is best for me right now. In response, it is important to be precise about what commitments we need from veritism to establish the challenge. We do not need to deny that there may be pragmatic or otherwise practical evaluations of our beliefs and moral judgments. We do not need to be anti-pragmatist in that sense (see Maguire & Woods, 2020: 211). What is important is that there seems to be a clear distinction between epistemic and pragmatic evaluations and that expressivists need to explain how distinctively epistemic evaluations are possible for moral judgments given their framework. Moreover, as quasi-realists, expressivists aim to vindicate our ordinary thought and talk, not revise it. And in our ordinary thought and talk, we seem to assume there *is* such a distinction. For example, we would usually want to separate the justification of my judgment moral judgment that lying is wrong based on its truth, from the justification that it is overall best for me to judge that lying is wrong. Thus, the challenge only depends on denying stronger, more controversial versions of pragmatism that expressivists are less likely to accept.

Even if it fails, this pragmatist proposal suggests an alternative. Expressivists may argue that it is too demanding to ask them to give the *same* explanation of epistemic evaluability for moral judgments when just showing how there can be non-pragmatic (but also non-veritisitic) evaluations for them is enough. Think, for example, of how we can sometimes evaluate states that are closer to desire-like states, like feelings, in terms of *warrant*. While these evaluations do not seem to be epistemic in principle, they do not seem straightforwardly practical or pragmatic. Consider a feeling of indignation on the face of wrongdoing, which we would typically assume is warranted. In this case, the agent does not necessarily need practical reasons for their feeling or be in their best interest to feel it, so the evaluation is not similar to ordinary practical and pragmatic evaluations. This may give expressivists a way to argue that these are closer to epistemic than practical evaluations. Thus, moral judgment as desire-like states may not need to be governed by a truth norm to be apt for epistemic

[44] See for example Brown (2018) and Rinard (2018, 2019).

evaluations.[45] This approach gets right that other norms that govern desire-like states more generally, and moral judgments in particular, can ground forms of epistemic evaluation similar to those of belief —and this will be my approach in Chapter 3. However, as it stands, there are two problems with that strategy. One is that it remains to be shown that these evaluations are *epistemic*, which would bring expressivists back to the initial dilemma: either the explanation resembles veritism and then expressivists need to explain how desires can fit into that picture, or it does not, and they owe us an alternative explanation of why the evaluations are genuinely epistemic. Second, this view seems to allow for just *one* form of evaluation: warrant. However, epistemic evaluations for moral judgments are diverse. Moral judgments seem apt to be epistemically justified/unjustified, rational/irrational, reliably or unreliably produced, etc. Thus, if the explanation only allows for warrant, it seems to be seriously limited.

## IV. Expressivism, Truth, and Epistemic Evaluability

Expressivists have good reasons to accept a challenge based on veritism because they have reasons to accept its main elements. However, not all is lost for them. Truth is central to veritism, and contemporary expressivists have devoted a significant amount of their efforts to articulate and defend views of truth as compatible with their views on moral thought and language. The resources expressivists developed for those views should be useful here, too.

There are two central aspects of most expressivists' accounts of truth for moral judgments and claims. Suppose we start by thinking of truth as involving an explanatorily robust role for notions like correspondence with facts or correct representation. In that case, expressivists are excluded from theorising about moral judgments as true or false.[46] They need to start their accounts with a view of truth compatible with their views on moral thought and language. Perhaps the most notable example here is *deflationism* about truth. According to this view, truth is not a substantial property which can play a substantial role in our theorising about the functions desire-like states can or cannot enter, and rather helps in generalisations or endorsement.[47] For example, the truth conditions for the judgment that 'racial

[45] Gibbard (1990: Ch. 15) adopts an approach like this but his main concern is with evaluations, not evaluability.
[46] This is of course not to say that expressivists cannot get to a point at which they earn the right to think about those notions as appliable to desire-like moral judgments. My point here is only about the explanatory role these notions can play.
[47] For more on these elements and the main varieties of deflationism, see Horwich (2010: 13).

discrimination is wrong' are nothing further than that racial discrimination is wrong, and adding the truth predicate would just show, for example, our endorsement of that judgment. Second, for expressivists, ascriptions of truth and falsehood to moral claims and judgments are, first and foremost, an element of moral practice. The primary function of a meaningful utterance that "it is true that racial discrimination is wrong" is to moralise, that is, to make a move in our practices of thinking and talking about moral matters. This is a practice that expressivists want to explain and vindicate, and that is why they have views like deflationism. However, they are careful in distinguishing what they can say as theorists from what they can say as participants in those practices.

Deflationism, with an appropriate understanding of the scope of expressivist explanations, suggests an appealing strategy to account for truth. To earn the right to talk about some realist-seeming surface feature of morality *X*, we should understand our talk of *X* as moves internal to our moral practices. Truth is then a familiar instance that fits in this general scheme. For example, if we ask expressivists *in virtue of what* a moral judgment that racial discrimination is wrong can be true, given that truth is not a substantial property, that will just amount to asking in virtue of what racial discrimination is wrong. This is a first-order normative question internal to our moral practices that can only be answered by committing ourselves to moral claims. Expressivism, as a second-order view about the nature of any moral judgment, is not in the business of offering answers to this kind of question. In that sense, expressivists have no problem with truth. This expressivist side-step is a well-known move in the literature.[48] Instead of thinking about what truth could be as applied to moral judgments, we ask about what we are doing when we think and talk of moral judgments as true. Once we are clear about the function of that thought and talk, we will see the initial problem disappear or change substantially.

This way of thinking about truth suggests a way to think about our problem with epistemic evaluability. For expressivists that endorse deflationism, once we understand what it is to think and talk of our moral judgments as true or false, we can see that there is nothing substantial for them to explain beyond platitudes like *p* is true iff *p*. The really interesting questions about which of our moral judgments are true and why only make sense from the point of view of a participant in moral practice. Epistemic evaluability may work similarly. Once we really understand what

[48] See Gibbard (2003: 6) for the classic articulation. See also Carter & Chrisman (2011), Dreier (2015) and Bex-Priestly & Gamester (2023: 10) for a more detailed explanation.

it is to think or talk about our moral judgments as aiming at true or as epistemically evaluated, there may be nothing substantial left for expressivists to explain. And so, they could leave interesting questions about how to best conduct moral inquiry for us to resolve as participants of moral practice. I will explore three versions of this side-stepping account in the next chapter.

Whether this indirect approach will be successful depends on how well deflationism can explain the aspects of the truth of moral judgments that we need to make sense of their epistemic evaluability. And unfortunately, as many critics point out, deflationism cannot explain all aspects of truth. For example, expressivists do not want to claim that just any kind of desire-like state is truth evaluable in the way that moral judgments are. The expression of my desire for chocolate in the form "yummy" or "chocolate!" are also expressions of desire-like attitudes, but no expressivist would claim they are truth-apt. Expressivists might want to argue that this is not a problem for their view of moral judgment and just a fact about the syntax of the expressions; after all, the previous two examples lack the indicative structure of moral judgments. However, Jamie Dreier (1996) convincingly argued against this. We can, for example, reconstruct a sentence like "Hey, Ana!" in indicative form as "Ana is hey" and even embed it in truth-conditional constructions like "If Ana is hey, then I'm happy." But that should not make it truth evaluable, so syntax is not the key to truth evaluability.[49] The upshot of this discussion is that having an account of truth like deflationism is not enough; just like the challenge of epistemic evaluability, we also need an account of why moral judgments would be the right fit for the truth predicate in the first place: we need an account of *truth evaluability*.[50]

Expressivists then have supplemented their views with similarly compatible but more substantial views of truth and truth-aptness.[51] According to these substantial strategies for earning a right to talk about some realist-seeming surface feature of morality *X*, expressivists should offer a meta-ethical account of the nature of our thought and language about *X*, which is compatible with its application to desire-like moral judgments. This is still meant to be an account of *X* as an element of moral practice, and most of the important questions surrounding it will still only make

[49] Although see Lenman (2003a) for an argument that to explain truth evaluability we cannot separate syntax from a form of *discipline*.
[50] To be clear this does not by itself mean that expressivists should reject deflationism. Some expressivists may argue that we still need it as the basis on which more substantial views of truth should be constructed. See, for example, Gibbard (2003: 65).
[51] For more on these issues and the development of the non-cognitivist approaches to them see Schroder (2018).

sense within it. But expressivists that favour this strategy accept the need for explanations beyond moves internal to moral practice. These explanations should make sense of the *correct* functioning of the practice, given their non-representational picture of moral thought and language. This is a broad characterization, so a substantial account of truth can take many forms. One example is Gibbard's (2003: 65) view of truth *evaluability*, according to which we should understand moral judgments as plans which are apt for agreement and disagreement in a way sufficient for any state to be truth evaluable. Another example is Blackburn's (1984: 198) view of stability as a proxy for truth. According to him, moral judgments are apt for improvement in a way that delivers a notion of the best set of attitudes —one that cannot be further improved by finer distinctions or better sensitivity to details. Such a set would be stable in the way that truth is.

Unlike indirect accounts, these views draw more interesting and substantial connections between moral judgments and truth —in a way closer to the veritist picture of beliefs. Thus, I believe it is worth exploring what aspects of veritism we could explain via these more substantial notions of truth and what the approach can teach us about how expressivists should answer questions of epistemic evaluability. I will explore this approach in Chapter 3.

## V. Conclusion

One of the main reasons expressivists want a moral epistemology is to vindicate those aspects of our practices that concern the epistemic status of our moral views. This is why their most fundamental challenge in moral epistemology is to explain how their views of our moral thought and talk are compatible with our epistemic evaluations. In ordinary moral practice, for example, we care about whether we know that it is permissible to eat meat, that we are sure that abusing a power imbalance is wrong or that the way we come up with answers to our moral questions is reliable and unbiased. The challenge is explaining how all these evaluations are possible when we consider our moral judgments desire-like. As I argued, the best way to understand why this is a problem is that paradigmatically epistemic evaluable states, like beliefs, aim at truth, and moral judgments, being desire-like, do not. This challenge is compelling since it is formulated in terms expressivists are likely to accept. But it also allows for the possibility that expressivists use the resources they developed elsewhere to account for truth.

# Chapter 2: Side-Stepping Epistemic Evaluability

Although the challenge of epistemic evaluability has only recently been articulated as such, it has many substantial and interesting connections with other aspects of moral practice that expressivists have explained, like our thinking and talking of moral judgments as true. Thus, there is a lot of potential for expressivists to apply the tools they developed for those projects to solve the present challenge. One very usual expressivist strategy is indirect: instead of explaining what things like moral truths of facts could be, we explain what it is to think or talk about them. Expressivists then argue that after successfully explaining our though and talk of moral truths and facts, we will see that there is nothing substantial left for them to explain. This indirect strategy is sometimes labelled the expressivist side-step (Dreier, 2015). Epistemic evaluability then may be apt for a similar side-stepping treatment. Since we are exploring the possible approaches expressivists can take to explain epistemic evaluability, it is important that we consider the options expressivists have for this side-stepping strategy. In this chapter, I present three different ways to do this. Two are based on truth, since truth was crucial for the veritist explanation of epistemic evaluability developed in the last chapter. The third is a specific way to apply epistemic expressivism to our problem here. Even if I believe there are good reasons to reject these side-stepping accounts, there are important lessons we should draw from them for a successful response, and so it is important to cover them.

Section I sets the stage by elaborating on the expressivist side-step, its relationship with what in the literature is called the expressivist manoeuvre and the options they give us to explain epistemic evaluability. Section II explores ways in which we can develop a deflationist account of truth to explain epistemic evaluability. If moral judgments can aim at truth in a deflationist sense, then epistemic evaluations can correspond to the means to get true moral judgments in that same deflationist sense. While this account makes sense of the correctness conditions of beliefs and moral judgments, it has important limitations in establishing those as the basis for epistemic evaluations. In section III, I explore a different approach that does not side-step the need for an account of truth but side-steps epistemic evaluability directly by reducing the conditions of epistemic evaluability to the conditions of truth-evaluability. As I will argue, while I do not wholly reject this option, I do not think it can work *as* a side-stepping account. Finally, I explore a third side-stepping strategy based on epistemic expressivism in Section IV. I expand on the version of epistemic expressivism I presented in Chapter 1 and explain how it can be turned into an

account of epistemic evaluability. Unfortunately, this option faces a different problem: it cannot explain what makes beliefs and moral judgments the right fit for epistemic evaluations but not ordinary desires.

## I. Epistemic evaluability and Side-Stepping

The most straightforward way for expressivists to solve the challenge of epistemic evaluability is to try to explain how the main elements of veritism are either compatible with their views of moral thought and talk or that they have the resources to mimic these elements. Thus, just like we can have a story about the truth, regulation and epistemic evaluation of belief, we can have a parallel story for moral judgments, even if they are desire-like. This account is ambitious in that it requires, first, an expressivist-friendly account of truth that can help explain the correctness of states like moral judgments. Second, an account of the nature of moral judgments that explains how they, being desire-like states, are apt to be correct or incorrect according to their truth. Lastly, expressivists need to show how these two elements offer a good basis for thinking of the means to form correct moral judgments corresponding to our ordinary epistemic evaluations of them: knowledge, justification, warrant, rationality, etc.

Some expressivists may think that this approach concedes too much too quickly to critics. It involves treating the correctness of moral judgments or their epistemic evaluability as a substantial property that expressivists need to explain. However, as we saw in the last chapter, expressivists are more likely to take an indirect approach, just like they do for things like moral truth or facts. Instead of asking directly what these things could be, expressivists ask what it is that we do when we think and talk about them. The response will usually involve the denial that explaining our moral thought and talk requires appealing to robust moral facts, truth as accurate representation or other notions problematic for expressivists. With that explanation, the initial problem of what these things could be in an expressivist framework either disappears or changes substantially.

Think about moral facts specifically. Say that I judge that it is a fact that exploitation is morally wrong. Some expressivists may want to explain this by arguing that this amounts to thinking nothing over and above that exploitation is morally wrong. This, in turn, amounts to a disapproval of exploitation. For this explanation, expressivists do not need to appeal to moral facts or ascribe a robust representational function to moral judgments. All expressivists need is a proper understanding of the function of our moral thinking and talking. Once we are clear about this function, we will see

the initial problem disappear or change substantially. If we ask whether there is anything else about the existence of moral facts that expressivists have not explained after taking this indirect approach, it should be easier for them to deny this. In the existing literature, this move is often referred to as the expressivist manoeuvre (Gibbard, 2003: 6; Carter and Chrisman, 2012: 331) or side-step (Dreier, 2015); however, I think it is worth being more precise about this.

The expressivist manoeuvre, as I will be using the term, refers to the strategy of accounting for a realist-seeming surface feature of morality *X* by explaining our thought and language about *X* first. This may or may not mean that we side-step or avoid the need for a substantial account of *X*. A substantial account means that we treat *X* as a substantial property that plays important explanatory roles. For example, Allan Gibbard (2003: 6) is often credited with articulating the expressivist manoeuvre, and although he applies it to solve the problem of *truth evaluability*, he does not side-step the need for a substantial account. He posits truth evaluability as a substantial property of moral judgments (as a plan-laden state) of entering relations of agreement and disagreement. The expressivist side-step, as I will be using the term, refers to a result of applying the expressivist manoeuvre and finding we have no need for a substantial account of *X*. Think again about the rough explanation of moral facts I presented before. The reason why a number of expressivists think our talk of moral facts is not representational is that they ascribe to a form of deflationism according to which saying or thinking that it is a fact that *p* is nothing other and above thinking or saying that *p*. Thus, we do not really need a substantial explanation of moral facts because the right account of our thought and talk of them does not need to refer to them to do any explanatory work.[52] Of course, deflationists do not want to deny that there are other interesting questions we may raise about our use of the concept of facts in our ordinary moral thought and talk that need answering. It is, for example, central for us as moralizers that it is a fact that exploitation is wrong. What deflationists allow us to side-step is the need for expressivists as metaethicists to give any further substantial explanation of what moral facts are. Whether exploitation is wrong is a fact that is for us moralizers to resolve.[53] In this chapter, I will explore the prospects of three side-stepping accounts of epistemic evaluability.

[52] For examples of accounts like this see Blackburn (1993: 173) and Dreier (2004: 26)

[53] C.f. Bex-Priestley & Gamester (2023: 10-2). Their interpretation of the side-step aligns more closely with how I am understanding the expressivist manoeuvre. My interpretation here does not involve changing or avoiding the question, so is not vulnerable to objections like Kyriacou's (2017) that by changing the question expressivists would use a form of

There are two ways in which we could develop a side-stepping response to the challenge of epistemic evaluability: by using a side-stepping account of truth to secure epistemic evaluability and by trying to side-step the need for epistemic evaluability directly. About the first option, the best-developed side-stepping account of truth is deflationism, so I will use it as the best basis for a strategy of this kind. The idea is that since epistemic evaluability is grounded on how states like beliefs aim at the truth, then a deflationist understanding of truth should have interesting connections that we could exploit. This should not (and will not) mean that we will use a deflationist sense of truth to *explain* epistemic evaluability. That is fundamentally against the deflationist core commitment that truth is not meant to play any explanatory role. Rather, if truth is not a substantial property, then for each role it plays, including the correctness of beliefs and its epistemic evaluability, there should be a way to eliminate the explanatory burden on expressivists. There is a well-known objection in the literature that by removing truth of its explanatory power, we will not be able to explain its epistemic roles. Those discussions have never been applied to problems of epistemic evaluability, but as we will see, I think similar worries arise. Even if a side-stepping explanation of truth fails to secure epistemic evaluability, expressivists may still be able to directly sidestep the need for epistemic evaluability. One way they could do this is by accepting there is more they need to say about truth, but once that account is secured, there is nothing else they need to say about epistemic evaluability. A second option is to focus on our thought and talk of epistemic evaluations, their function and how much it assumes about epistemic evaluability as a property of moral judgments. Perhaps once we are clear on the expressivist account of our thought and talk of epistemic evaluations, the latter problem either disappears or changes substantially.

Before we start with the discussion, I should be clear that I do not think side-stepping accounts can give us everything we want from an expressivist account of epistemic evaluability. But, on the one hand, given the recent articulation of the challenge, it is worth mapping the options expressivists have to respond to it and mark the promising routes in which we can take the discussion. On the other hand, there are aspects that I believe these approaches get right, and they help us highlight the elements that are important for a response to the challenge. Thus, there are several lessons that I will draw from these views for the direct solution I will offer in the next chapter.

---

reasoning that is characteristically non-truth conducive. In that sense, my characterization is also similar to Sinclair & Chamberlain (Forthcoming: 8).

## II. A First Deflationist Approach

Deflationism is a natural companion to the side-stepping strategy. Side-stepping an account of truth consists of explaining what it is to think and talk of something as true and then find, after offering that account, that there was nothing substantial to explain about truth itself. According to deflationism, truth is not a substantial property, so it makes sense that there are no explanatory interesting answers expressivists can offer about our use of the truth predicate or concept. To apply this view to our problem with epistemic evaluability, we need a more precise definition of deflationism. Like expressivism, deflationism can take many forms, and there are substantial disagreements among deflationists. However, their main points of agreement are well captured by two main commitments: a negative commitment that truth is not a substantial property capable of doing any explanatory work and a positive commitment that everything we need to know about the function of truth is contained in the schema:[54]

T) *p* is true iff *p*

This usually means that the function of our concept of truth in thought and talk is nothing other than a device for expressing what is already contained in T). For example, we can use the concept of truth for generalization: instead of saying *S* said that *p* and *p*, S said that *q* and *q*, S said that *r* and *r*, we can just say what *S* said is true. Or we can use truth for endorsement: instead of just saying *p* we can express our endorsement by saying that it is true that *p*. This is not to say that there is no sense in which using the concept or predicate of truth is important. In our everyday lives, it often matters very much that what we say and think is true, that we get true information from other people, and so on. But what deflationists say about those cases is that what we care about has nothing to do with truth itself. For example, when I want to know whether it is true that I deposited the money in the right account, what matters is my reasons for thinking so, like my memories about it, the ways I have to find out, and so on. Theorizing about truth will not help with any of that. Everything we can say from the point of view of someone theorizing about the nature

[54] I pick these two since I believe they capture elements that characterize most forms of deflationism and will be useful for our discussion. But for alternatives see Lynch (2009: 105-9), Horwich (2010: 13) and Armour-Garb, Stoljar, and Woodbridge (2021). It may be controversial whether the first thesis should be about a property that cannot be used in explanations or a metaphysically fundamental property. I will stick to the former since, on the one hand, it is key for the versions of deflationism I will discuss in this section. And, on the other admitting truth can be used in explanations seems closer to the form of minimalism I discuss in the next chapter.

of truth is something trivial like T). Our account of epistemic evaluability can follow a similar pattern. If truth is important for our epistemic evaluations, then that may not be because of truth itself, and everything that truth can contribute to our accounts is something similar to T).[55]

One initial problem with this is that deflationism is often presented as an account of truth *ascriptions* or truth *conditions*, but normativism, as it figures in veritism, is about the correctness conditions of beliefs. Recall that according to veritism:

B) A belief that *p* is correct iff *p* is true

But this is not a significant problem. Some deflationists like Paul Horwich (2010: 75-7). argue that just like truth was merely a device of generalization and endorsement in truth attributions, it plays a similar role in conditions like B). And so we could easily reconstruct them without appealing to truth. We would have infinite instances of norms with the form:

B*) A belief that *p* is correct iff *p*

And given that, again, truth is not a substantial property, nothing is stopping us from reformulating these as:

M*) A moral judgment that *p* is correct iff *p*.

Like the initial view, that would not mean that the truth or the correctness of our beliefs is not important. But the reasons why those are important have nothing to do with truth. We can think about it in this way: truth is very easy to get since it is not a substantial property, so it can be attributed equally to belief and desire-like states like moral judgments. It is just meant to make things easier for us by formulating things like B) in more general terms. But which beliefs are actually correct or whether a specific belief that *p* is, that is a harder question that requires us to stop theorizing about truth and inquiring about whether *p*. And it does not matter whether the *p* includes moral content or not.[56] We need to determine which of our moral judgments

[55] To be sure, this emphasis on what is important about our talking and thinking about truth is not often part of the characterization of deflationism. Classic articulations like Horwich (2010) and Field (2001) do not give it any prominent role to it. However, see Shieh (2018) for why it is important for deflationism in general. And Blackburn (2009) for why it is important when discussing expressivism. Some, like Wright (2021) argue that this element unites different *pragmatist* accounts of truth like deflationism and minimalism —more on this in chapter 3.

[56] While discussing deflationism I will not be assuming that variables like *p* stand for propositions. Some deflationists like Field (2001: 104) would endorse similarly deflationists accounts of propositions and so would deny these should figure in our analysis of thought and language.

are correct, and to do that, we need to stop theorizing about truth and moralize. There are two aspects of this account that I think deflationists get right. Expressivists want to say that M*), or at least a very similar norm, governs moral judgments. Also, the normative pressure to regulate our moral judgments according to whether they are correct comes from the internal dynamics of first-order moral inquiry. Whether our moral judgments are correct or not is a matter of what is actually right or wrong, which we cannot tell by theorizing about truth; we need to moralize. Unfortunately, normativism is just one part of a complete explanation of epistemic evaluability, and deflationism faces a few problems accounting for the rest.

According to the second part of veritism, epistemic evaluations stand for bad or good means to form true beliefs. This initially does not seem to be a problem for deflationism. After all, we can think of correct beliefs not as those that aim at a substantial property of truth but as those that satisfy conditions B*), that a belief that *p* is correct iff *p*, which does not include truth. Then epistemic evaluations can be the means to form correct beliefs for each *instance* of B*) —since we are not using truth for generalization. For example, my belief that Trotsky died in Coyoacán is correct iff Trotsky died in Coyoacán, and epistemic evaluations like justification, rationality, warrant, etc., stand for the good and bad means to form a correct belief in that matter; those that would not support a belief that Trotsky died in Coyoacán when he did not. We will have a similar story for each belief that we evaluate epistemically.

But now say that we need the conditions of epistemic evaluability for the belief that there is a computer in front of me. We have a similar story in that we determine its correctness conditions without truth, just following B*), and then with the means to get the correct belief we also get epistemic evaluations. But what would the evaluations for the belief about Trotsky have in common with the evaluations for this belief about my computer? It cannot be that, like with veritism, both evaluations are epistemic because they concern truth, since that would mean using truth in an explanatory way. Perhaps we could draw a general principle from the story about epistemic evaluations we drew from Wedgwood's remark in Chapter 1:

> P) One can only aim to have a correct belief that *p* by aiming aim to have a rational belief that *p*.

This would help at first since we can explain the correctness of the belief in terms of B*), that a belief that p is correct iff p, and so truth does not need to play any substantial role. However, the problem was not with the belief's correctness but with rationality. We need an explanation of what, other than truth, could unify the

epistemic evaluations of rationality for each belief that meets condition P); why they are all different instances of the same epistemic evaluation. Moreover, epistemic evaluations can be comparative. For example, I can be more justified in believing that my computer is in front of me than in believing that Trotsky died in Coyoacán —perhaps because of the weight of the evidence for each. Deflationism offers no way to explain that since there is nothing unifying the means to form a correct belief in the two instances we are considering of B*).

Deflationists may find other ways to unify epistemic evaluations without using truth. Perhaps they can say that what unifies different applications of P) is not truth but the kind of mental state they apply to. In our examples, both evaluations are epistemic because they are both evaluations of *beliefs:* one about Trotsky and one about my computer. And there is some precedent in the literature of views that follow this path in reducing epistemic norms and evaluations to the norms of belief.[57] However, the motivations for those accounts have been mostly explaining what makes epistemic norms *normative* and not so much on what makes them epistemic. Deflationists then need different support for the view if they want to solve this specific problem. Second, and more importantly, perhaps that would be a good strategy for some deflationists but not so much for expressivists. After all, we needed deflationism to show that even if moral judgments are desire-like, their correctness conditions would be appropriate for an epistemic evaluable state. Expressivists then would be back at the problem that moral judgments need to be more belief-like; it is just that it is unclear in what way.[58]

To be sure, expressivists would technically not be out of options. They could very well find a property that is independent of truth, beliefs and moral judgments share, and somehow explains epistemic evaluations. But the difficulties seem enough for us to look for a different option that does not side-step truth.

[57] As McHugh and Whiting (2014: 707), point out, this seems to be Ernst Sosa's view. However, Sosa has a significantly more sophisticated version of normativism than the one I have been assuming here. For him, believing is a complex activity that involves norms of competence and success. For more on this see Sosa (2021). See also Shah (2006) for a similar suggestion about epistemic evaluations and the norm of beliefs.

[58] Other arguments that deflationism cannot explain the epistemic role of truth adequately can be found in Wright (1992) and Lynch (2009: 105-14). For an argument against a particular application of deflationism for the expressivist project of earning the right to epistemic notions (although in that case for fallibility) see Gamester (2022: 5).

## III. Truth Evaluability and Epistemic Evaluability

A different but related option for expressivists is not to side-step truth to secure epistemic evaluability but to side-step epistemic evaluability directly with a more complete account of truth. As explained in Chapter 1, one of the main shortcomings of deflationism is that it cannot help us differentiate expressions of states which can figure in schemas like T) and those that cannot. Moral judgments are both desire-like states and truth evaluable: a judgment that exploitation is wrong should be understood as a disapproval of stealing and as something that is true iff exploitation is wrong. Not all mental states are like this. Thirst, a headache or even ordinary desires should not be the kind of things that, when expressed, could be truth evaluable. Thus, as most, if not all, expressivists accept nowadays, apart from views like deflationism that explain truth ascriptions, they need a further account of truth-evaluability. The notion of truth evaluability and the notion of aiming at truth seem to be closely connected. Both seem to be about what makes states like beliefs and moral judgments appropriately connected with truth. Expressivists may then rightfully ask whether conditions for epistemic evaluability are *just* conditions for truth evaluability. If so, after securing a sense of truth evaluability, we can say that moral judgments are sufficiently belief-like so that from the point of view internal to our moral practices, we can just treat them as such. Thus, from that point of view, epistemic evaluations work just like they do for beliefs, as corresponding to the means to get the true ones. Let us try to flesh out this idea.

The problem of truth evaluability is an essential part of the more general problem of explaining the propositional functioning of moral thought and language, like the enormously influential Frege-Geach problem. And there is a widespread view that we should understand this problem as asking for a way expressivists can earn the right to talk about moral *contents* using their central notion of *expression*.[59] Suppose we adopt a view like Kvanvig's (2014), according to which the primary objects of epistemic evaluations are contents. In that case, we can start to see a connection between truth-aptness and epistemic evaluability. Suppose an expressivist gets the solution to the truth evaluability challenge right by giving us a notion of moral content. In that case, that notion should also give us a proper object of epistemic evaluations. One can draw an optimistic or a pessimistic conclusion from this point. According to the optimistic view, epistemic and truth evaluability challenges are

[59] Woods (2017) and in some respects Schroeder (2008b) both endorse this idea. C.f. Yalcin (2019: 409-11).

essentially the same, so the solutions expressivists already offer to the former should, in principle, help us with the latter.[60] According to the pessimistic view, the problems call for different solutions, and if we grant that the right notion of content should guarantee epistemic evaluations, that would just mean that every solution to the problems like the Frege-Geach is incomplete until expressivists show how the notion of content they supplied can also be the proper object of epistemic evaluations. Only an optimistic view would count as a side-stepping strategy since the pessimistic view requires a further explanation of epistemic evaluability to work.

There are some reasons to resist the optimistic conclusion. Even among those who agree that for expressivists, explaining truth evaluability is fundamentally the problem of explaining moral contents, there is no agreement on what a solution would amount to.[61] The idea then cannot be that we know which property or properties determine truth evaluability, and those also determine epistemic evaluability. The best strategy is to proceed *negatively*, denying that there are examples of states that are truth evaluable and not epistemically evaluable. There are, however, some objections to that approach. When defending the distinctiveness of the challenge of epistemic evaluability, Beddor (2019: 9) briefly considers the relation between epistemic and truth evaluability. He argues that truth evaluability and epistemic evaluability must be independent because we have examples of mental states like *imaginings* and *suppositions*, which are intuitively truth-apt but not epistemically evaluable. For simplicity, let us focus on imaginings.

One way in which beliefs and imaginings are truth evaluable is that they are both propositional and representational attitudes. I can both imagine and believe that I can walk 40 km without stopping, and the proposition in the content of both states can be true or false. In that sense, they are both truth-evaluable attitudes. However, expressivists can argue that they interact differently with the truth of the proposition, and so they are not truth-evaluable in the same way. It is, for example, easy to think that both the belief and not just the proposition in the example are true. But it is harder to think of the imagining itself as true. The problem with that response is that imaginings are diverse, and in many cases, it seems plausible to think of the imagining itself as true; for example, in cases of imaginative immersion (see

[60] Sepielli (2012) defends a similar relation between propositional and epistemic problems for the case of uncertainty. He argues that if one has a solution to the Frege-Geach problem, one has then a solution for Smith's (2002) certitude challenge for expressivism.
[61] Some candidates are disciplined syntacticism (Lenman, 2003a) and a specific form of disagreement (Gibbard, 2003) but these are controversial.

Schellenberg, 2013).[62] Moreover, some of the best ways to mark a difference between how imaginings and beliefs interact with the truth of the proposition in their content would support the challenge of epistemic evaluability, not help expressivists side-step it. For example, Shah and Velleman (2005) argue that the best way to distinguish them is that beliefs, but not imaginings, aim at the truth.[63] And so, expressivists still need to explain how moral judgments can be desire-like and aim at the truth. Truth evaluability is extremely difficult to pin down, so there may be a different way to distinguish the truth evaluability of beliefs and imaginings. However, from where we stand, it is difficult for expressivists to build a case on that hypothesis.

Expressivists could argue that while we cannot distinguish between the truth evaluability of imaginings and beliefs, imaginings, contrary to appearances, are epistemically evaluable. Thus, for all we know, epistemic evaluability conditions may still be conditions for truth evaluability. For example, many authors claim that imaginings are essential in getting modal knowledge and in our counterfactual reasoning.[64] And even if it would be a stretch to think of them as knowledge, they may have a more complex and close relation with notions like justification.[65] But that option is unlikely to be helpful. As established in Chapter 1, veritism is not the view that only beliefs are epistemically evaluable. Other apt objects like credences or patterns of reasoning are epistemically evaluable by having an aim sufficiently similar to beliefs. And even if we cannot ascribe a similar constitutive aim to imaginings, we at least have a connection with truth. Imaginings can be a means to form true modal beliefs. But we do not have a similar story about moral judgments yet. Thus, by leaving these questions open, the account would not side-step the challenge of epistemic evaluability.

A third and final way in which expressivists can respond is that we have been assuming that the difference between beliefs and imaginings or suppositions is that the former have an extra element that the latter lacks. Something else about beliefs makes them epistemically evaluable, unlike the other two. However, it might also be

[62] More precisely, our *attributions* of truth and falsehood in those cases seem to be directly about the imaginings. See also Egan (2008). There are similar difficulties in distinguishing the truth of suppositions and their contents. See, for example, Elstein & Williams (Ms).
[63] See Shen-yi & Gendler (2019: 2.1) Sinhababu (2013) for other options.
[64] For an overview of this discussion see Kung (2016).
[65] See Williamson (2016). Some argue that the imaginings produced while engaging in fiction are apt for more rich epistemic evaluations. But even those are discontinuous with the epistemic evaluations for belief. See Gilmore (2020: Ch. 6).

the case that imaginings and suppositions need an extra element *not* to be epistemically evaluable. It is fundamental for the three states to have contents that can get things as they are, and imaginings and suppositions need an extra element that enables us to use this capacity in a non-centrally epistemic way. Of course, this option is controversial, but it may find some support in views that hold that imagination is part of the same continuum of states as belief.[66] In that way, if moral judgments are already close to any of these states, it would actually be harder to prove they are not epistemically evaluable. I believe that this is an interesting suggestion worth exploring. However, there are two things to consider. First, the property that can make beliefs and moral judgments epistemically evaluable and imaginings and suppositions not cannot just be propositional content. After all, there are well-known and plausible views according to which the contents of ordinary desires are also propositions. But it seems hard to accept that ordinary desires can be evaluated epistemically, and it is just an extra element that blocks that feature. If it is not propositional content, then it is very unclear what the shared property may be. Second, this suggestion would be helpful for expressivists to understand epistemic evaluability, but it is not clear that it would support the optimistic conclusion. After all, we still need to find the property that all truth-evaluable states share, but that is somehow blocked when we imagine and suppose. There seems to be explanatory work beyond truth-evaluability to make sense of this difference, so this account starts to look like a pessimistic view.

The primary idea that motivated this option that epistemic and truth evaluability are closely connected, seems to me to be on the right track. However, the optimistic articulation of this idea that conditions for epistemic evaluability are just conditions for epistemic evaluability requires controversial assumptions about truth evaluability that expressivists are not entitled to. Thus, a pessimistic view seems more promising, even if it is not a side-stepping strategy. More specifically, views on the truth-evaluability of moral judgments give us promising materials to construct a response to the challenge. This will be my approach in the next chapter. But before proceeding to that, we need to consider one last side-stepping strategy.

[66] As noted above Schellenberg (2013) and Egan (2008) both hold views like this. A more directly related option could be Sterelny's (2003: 11) view that the functions of representing emotions in imagination as parasitic of the evolutionary advantages of representing emotions and so could be explained as offline versions of those capacities.

## IV. Epistemic Expressivism

Some expressivists may argue that the problem with our two initial approaches was that they targeted normativism. But what we really want to understand is what makes our practices of epistemic evaluability possible, so it is our thinking and talking of epistemic evaluations that we should be focusing on. And expressivists already have well-established ideas about our epistemic evaluation thought and talk in the form of epistemic expressivism. The idea is that, like its moral counterpart, epistemic expressivism is the view epistemic thought and talk do not have the function of representing epistemic properties but express practical or desire-like mental states. Allan Gibbard (2003: 199, 2012), for example argues our knowledge talk expresses states like plans to rely on someone's judgment. Consider a specific epistemic evaluation, like it is epistemically justified to judge that racial discrimination is wrong. Given epistemic expressivism, we should understand this as the expression of a state of planning to rely on that judgment. As established in Chapter 1, this says little about which kinds of objects are apt for those evaluations, so it says little about epistemic evaluability. However, a better version of this approach would combine epistemic expressivism with the side-stepping strategy. If epistemic evaluability is a property at all, it is an epistemic property. And expressivists, armed with epistemic expressivism, may question whether epistemic evaluability is a substantial property that they need to explain. Thus, the right framing of epistemic expressivism should give us a way to explain away or side-step epistemic evaluability.[67] After explaining all our thinking and talking of epistemic evaluations, we can come back to ask whether there are any open questions about the property of epistemic evaluability. If there are none, expressivists would have successfully side-stepped the problem.

The key to this argument is the often-overlooked idea that given the non-representational function of epistemic thought and talk, epistemic expressivism gives us a different approach to epistemic properties. For example, Hartry Field (2018: 2) conceives of competitors to epistemic expressivism about the reliability of inductive inferences as having to posit a kind of "justification fluid" that gets passed from premises to conclusions. For an epistemic expressivist, we just need a good explanation of our *thinking* and *talking* of certain inferences as reasonable. Once we understand that thinking and talking consists of, say, planning to rely on those inference methods and expressing those states, we can come back to the initial issue.

[67] Many thanks to Will Gamester for suggesting this approach and pressing me to give it more serious consideration.

We can then ask whether there are any open questions about the substantial properties reasonable inferences have and that epistemic expressivists need to explain. At that point, epistemic expressivists are in a good position to deny there are any. Expressivists then can say something similar about the property of epistemic evaluability. After giving a complete and convincing epistemic expressivist account of our thinking and talking of epistemic evaluations for moral judgments, they can come back to the initial issue of epistemic evaluability and ask whether there are any open questions about the kind of property epistemic evaluable states have, and then they deny there are any.

Like other side-stepping accounts, the idea is that after explaining our talk and thought, there are no more open questions for expressivists to respond to. But there may be other questions about epistemic evaluability for us to resolve as participants of moral practice. Perhaps from that first-order perspective, we can ask questions about whether it makes sense as participants of moral practice to evaluate moral judgments epistemically. Even if expressivists are not in the business of responding to that question, they can safely assume that from within our moral practices, there are many instances in which it *is* appropriate to evaluate moral judgments in this way. It is very important to us, for example, whether our moral views are *justified* or whether agents *know* what they are doing is wrong. And expressivists have offered some epistemic expressivism about those judgments, so there should not be remaining questions about those.[68]

One thing to note is that epistemic expressivism is a view about *normative* epistemic thought and talk. It is clear how this side-stepping account applies to properties like *reasonable* inference since, if they are properties at all, they are normative —they are picked out by a normative concept. However, it is unclear that epistemic evaluability itself is normative. One way to respond to this worry is by arguing that we are working under the assumption that veritism, or at least something very close to it, is the right account of epistemic evaluability. And according to veritism, what explains epistemic evaluability is a normativist picture of belief according to which they are *correct* if and only if they are true. Thus, if epistemic evaluability is an epistemic property, then we have some reason to think it is a normative one. However, even if we grant that there are more serious issues with the view.

Expressivists need an account of epistemic thought and language as applied to moral judgments. And it is very plausible that the account is a form of epistemic

[68] See for example Blackburn (1996) and Gibbard (2003).

expressivism. Moreover, like deflationism, this version of epistemic expressivism gets right that the right account of epistemic evaluability for moral judgments should say that the pressure to regulate them according to their truth comes from the dynamics of moralizing. However, it is unclear whether epistemic expressivism can side-step the challenge. There seem to be at least two important questions left open after the side-step.

First, a version of Dreier's (1996) objection against deflationism may apply to this view. The idea then was that even as a convincing view about truth claims, deflationism cannot distinguish between states that are apt for truth evaluations and those that are not. Similarly, a critic can say that after a complete epistemic expressivist account of epistemic evaluations, there is a relevant and open question about why, for example, moral judgments are epistemically evaluable but *ordinary desires* are not. Expressivists could respond that an important element of this side-stepping approach is that the remaining questions here are first-order questions about the appropriateness of epistemically evaluating these states. Then, they can appeal to wide reflective equilibrium (WRE). WRE is among the best reasoning processes for moral matters. But WRE involves ensuring that our moral judgments are consistent with our best non-moral judgments. In this case, if we did not judge that it was inappropriate to evaluate ordinary desires epistemically as part of our moral reasoning, but we did as part of our best non-moral reasoning (or would if we were engaged in our best non-moral reasoning), then, given WRE, our best moral reasoning ought to fit with this view. However, this view assumes that it is part of our best non-moral reasoning that there is something that makes ordinary desires not epistemically evaluable. But why is that? Presumably, something separates them from epistemically evaluable states like beliefs and moral judgments. But this something is precisely what a response to the challenge of epistemic evaluability needs to explain. Perhaps expressivists could appeal to truth-evaluability, but then we can raise a similar challenge with the states we considered in the last section, like imaginings.[69]

Second, the strategy does not side-step the need for a view like veritism. Consider the kind of states we express when we make an epistemic evaluation. For Gibbard, these are states of planning to rely on someone's judgment. Planning states are

[69] The objection is not entirely analogous to Dreier's since for him we could get non-truth-evaluable states to look truth-evaluable by the deflationist lights with the right syntax. But here we are dealing with mental states. However, for expressivists, ordinary desires are already relevantly similar to moral judgments in that they are both desire-like states.

responsive to reasons. And if the account is to be plausible, there must be a way to distinguish good reasons from bad ones. Since this is a view about *epistemic* evaluations, determining which corresponds to the means to attain truth is a promising way to distinguish them. But this is one of the substantial commitments of veritism that we were supposed to side-step. Proponents of this side-stepping strategy would then be back to the initial problem: either they accept this version of veritism, which seems incompatible with expressivism, or offer a similarly plausible alternative compatible with their view. This problem is not exclusive to Gibbard's account. Consider Michael Ridge's more sophisticated version of epistemic expressivism. According to him, knowledge attributions express very complex states that minimally include:

> "(1) The belief that *S* believes that *p*.
>
> (2) The belief that *S*'s belief that *p* is true.
>
> (3) Epistemic endorsement of certain procedures for deciding what to believe.
>
> (4) The belief that S's judgement that *p* is causally regulated by either (a) those procedures [anaphoric reference back to those procedures the speaker endorses in (3)] or (b) procedures which are close enough to those procedures, so far as p goes or (c) more fully informed successors to those procedures.
>
> (5) The belief that no further acquaintance with the descriptive facts is such that acquaintance with those facts is sufficient for those epistemic procedures (the ones causally regulating his belief that p) to instruct him to abandon the belief that *p*." (2007: 103-4)

Ridge's account improves on other versions of epistemic expressivism in many ways. It is, for example, helpful in specifying the kind of state(s) we express when making *epistemic* evaluations. However, it cannot help us solve our problem with epistemic evaluability. Notice that (3), what he calls the plan-laden element of his view (2007: 105), faces the same problem as Gibbard's view. Epistemic endorsement is, presumably, not random; it must be in some way responsive to reasons. And either those reasons are based on truth, and so we are back to veritism, or they are not, and then we need a different explanation of what makes this endorsement epistemic. Moreover, according to (4), for an epistemic evaluation, we need to think the judgment is apt to be regulated by the procedures we endorse. But if these are the

judgment formation procedures in virtue of which the judgment counts as knowledge, it is unclear what distinguishes this from epistemic evaluability. And so, in endorsing this form of epistemic expressivism we would be assuming a response to the problem. Epistemic expressivism, even if helpful in other respects, does not side-step the need for a view like veritism.

## V. Conclusion

From veritism, we know that truth is the key to responding to the challenge of epistemic evaluability. Given that a significant focus of contemporary expressivists has been on truth, we can draw from many of the resources they developed to respond to this new challenge. However, their specific approach to truth affects how effectively they can answer the challenge. Expressivists then need a conception of truth that, like deflationism, is compatible with their views on the nature of moral judgments and so could work as their aim. However, when it comes to epistemic evaluability, deflationism has many limitations. Even so, the challenge I am addressing here, although having deep roots in existing objections to expressivism, has only recently been articulated as such. Thus, there are other possibilities for expressivists to explore. In particular, I want to consider one more option for expressivists, one that consists of taking veritism seriously and earning the right to its central elements.

# Chapter 3: Epistemic Evaluability and Stability

Veritism consists of a normativist picture of beliefs as constitutively aiming at truth and a teleological account of epistemic evaluations as corresponding to the good or bad means to attain truth. Expressivists can earn the right to epistemic evaluability by earning the right to the main elements of veritism: truth, regulation and evaluation. First, to explain the truth element of veritism, I appeal to Simon Blackburn's view (1984: 198, 1996), according to which moral judgments are apt for improvement in a way that delivers a notion of the best set of attitudes —one that cannot be further improved by finer distinctions or better sensitivity to details. Such a set would be stable in the way that truth is. This stability property will be our proxy for truth and act as a correct condition for moral judgments in the same way truth does for beliefs. Second, I argue moral judgments while desire-like are apt for those correctness conditions by appealing to an evaluationist or guise of the good view of the nature of desire-like states. The key will be in how the function of desire-like states is to help close practical deliberation, but in the case of moral judgments, they can only do that correctly when they are stable. Having secured truth and evaluation, expressivists can find plausible connections with epistemic evaluation.

Section I sets the stage by introducing two elements of substantial accounts of truth. Section II presents the account of stability as a proxy for truth and explains why stability can act as a correctness condition for moral judgments. While apt for stability, moral judgments are still desire-like states, and expressivists have good reasons not to deny that they still have the correctness conditions of any desire-like state. In Section III, I introduce the evaluationist view of desires. This view will help us explain how the correctness conditions proposed are apt for a desire-like state. After establishing the parallels, I address some possible objections to the view. Section VI tackles the issue of evaluations as the means to attain truth.

## I. Veritism and Substantial Truth

According to veritism, beliefs are epistemically evaluable because they aim at truth, and epistemic evaluations correspond to the means to achieve that. This offers a compelling challenge and a way for expressivists to make sense of what it takes to respond to it. More specifically, there are three elements of veritism that expressivists can show are not incompatible with their view and earn the right to epistemic evaluability: truth, regulation, and evaluation. If desire-like moral judgments also aim at truth (or something close enough to it) and so are regulated similarly to beliefs, then the evaluations that correspond to the means to attain the aim will warrant the

label of epistemic. My strategy here is to construct the case that expressivists can answer this challenge by accounting for each of these in order.

We already covered one way to use expressivist-friendly conceptions of truth to secure some aspects of veritism: deflationism. That strategy is either unsuccessful or limited, mainly because, according to deflationism, truth is incapable of doing substantial explanatory work. Fortunately, expressivists have developed other more substantial accounts of truth to deal with the limitations of deflationism. In Chapters 1 and 2, we covered some examples of this in the form of accounts of truth-evaluability. Two aspects of these more substantial accounts are useful for the strategy I will develop here.

First, we need a notion of truth that can do explanatory work to act as a correctness condition for moral judgments and be a plausible basis for epistemic evaluations. The main example I will be using here is Blackburn's view of stability as the truth-like property of moral judgments that cannot be further improved. The issue of how views like Blackburn's are substantial and how they differ from deflationism is complex. One way for us to think about it is in terms of the contrast between deflationism and minimalism.[70] According to minimalism, truth is the *property* designated by all our platitudes surrounding our use of the concept. Most of the ideas expressivists would reject about truth, like correspondence with facts, are controversial *as platitudes*, so they are not essential for a theory about the truth of moral judgments. And the uncontroversial and important platitudes are things that expressivists can easily accept. One example of these platitudes is the schema deflationists claim explains everything we need to know about truth:

T) *p* is true iff *p*

But, importantly, there are others, like the stability property, which I will introduce and explain in the next section. Minimalism, unlike deflationism, is not committed to truth not being apt for doing explanatory work and not committed to T) including everything we need to know about truth. Both these aspects will be helpful in our discussion. One warning here is that the terminology in the existing literature is confusing. Horwich (2010: 19) characterizes minimalism as a form of deflationism. Others, like Wright (1992) and Sinclair (2021: 170), distinguish minimalism from

[70] The distinction as I am drawing it here, and this articulation of minimalism is from Wright (1992). For an overview of minimalism see Sinclair (2021: 170).

deflationism in the way I do here. When I use the term 'minimalism', it will be in this latter sense in which it contrasts with deflationism.

A second aspect of substantial views of truth is a commitment to giving more substantial explanations of what it is for a *mental state* to be apt to figure in platitudes like T), to be truth-apt. Take Allan Gibbard's account of truth evaluability. Gibbard's account works mainly at the meta-semantic level since we are looking for the property *in virtue of which* moral judgments have the semantic feature of being truth-evaluable.[71] Gibbard looks for a property that different mental states, like beliefs and moral judgments, have in virtue of which the claims we use to express them receive a propositional treatment. And he concludes that the most plausible candidate is being able to enter into specific kinds of relations of agreement and disagreement (2003: 65-71).[72] His claim is that if you disagree with a mental state, you can make sense of a negation of the claim that expresses it and then start to construct the propositional functions form there. This view gives a kind of psychological plausibility to the account and also demarcates truth-evaluability nicely. There are certain mental states for which it does make sense to disagree with, like beliefs, and some which do not, like plain desires or headaches. According to Gibbard, it makes sense to understand moral judgments in terms of states of normative governance like approvals, plans, or acceptances of systems of norms, for which it makes sense to agree or disagree. Since we need a similar account for beliefs, Gibbard claims that moral judgments do not mimic their truth evaluability but get to it in the exact same way.[73] Furthermore, states like plans are also desire-like in guiding action by being composed of sets of intentions. In that sense, moral judgments can have belief-like and desire-like features.[74]

---

[71] At the level of *semantics* he also proposes a modification to the usual parameters of evaluation in the possible world framework so that it includes hypeplans. That way if we evaluate a descriptive claim we do it relative to a world and if we evaluate a normative claim we do it relative to a hyperplan (2003: 53). I take this as part of his view on truth evaluability, but not one that connects as clearly with our issue here.

[72] Gibbard (2003: 184), like Blackburn before, claims that if we cannot think of anything more we want from an account of beliefs apart from being a state that has these features, then moral judgments are effectively beliefs. Thus, a different way in which my proposal is connected with Gibbard's is that both are parts of a broader project for expressivists to earn the right to a sense of moral beliefs.

[73] This appeal to a unified account of judgments is what he calls *broad* expressivism (2003: 82).

[74] Just as not all representational attitudes result in truth-functional or propositional judgments, Gibbard thinks only some of our practical plan-like attitudes result in truth-functional judgments. Disagreement plays a significant role in this jump from attitude to judgment, but plausibly other elements like our needs of expression and reasoning inside our moral practices are central too.

To be sure, my strategy will *not* be to use Gibbard's view to explain epistemic evaluability. For one, the proposal is controversial. As some critics have pointed out, a general sense of disagreement might be just assuming the notion of inconsistency Gibbard is trying to explain (Schroeder, 2008a). There are also widespread cases of (metalinguistic or pragmatic) disagreement that do not assume a shared notion of content, and thus, disagreements are not a reliable indicator of truth-evaluability (Plunket and Sundell, 2013). My idea is that we can use this as a blueprint for a substantial view of epistemic evaluability that is sufficiently belief-like but grounded in the desire-like properties of moral judgments. I will start with truth to specify the property we are looking for and then move on to the account of the nature of desire-like states that can explain why that is a property of desire-like moral judgments.

Given that this strategy involves exploring the desire-like nature of moral judgments in more detail, we will make more heavy use of the non-cognitivist resources of expressivism. Non-cognitivist views are complex, and there is not always agreement among non-cognitivists. To keep our discussion manageable, I will work with two restrictions. First, there is no general agreement among non-cognitivists about which desire-like state moral judgments consist of. Some influential candidates are approvals (Blackburn, 1998a), preferences (Dreier, 2006; Silk, 2015), and plans (Gibbard, 2003). Here, I will speak initially only of approvals as I believe it is the most general of the candidates —we will come back to this in Section V. Second, there are significantly different kinds of moral judgments: some specify actions we ought to perform, others say that a thing or a state of affairs is good or bad, yet others say that a person is virtuous or vicious, and so on. I cannot offer a fully general account here, and given that expressivists only need to argue for the possibility of epistemic evaluability, I will focus on just one specific kind: first personal judgments of right or wrong actions. These will, additionally, help us highlight the expressivist concern with the connection between moral judgments and intentional actions.

## II: Truth: Stability Through Improvement

The most promising substantial views of truth for an expressivist version of veritism are those based on how expressivists think of our practices of forming and settling on moral judgments such that the goal of moral inquiry is to have moral judgments with truth-like features. In this respect, there are a few examples in the literature, and we do not have to commit to one for our purposes here.[75] I will, however, use a view

[75] See, for example, Crispin Wright's (1992) superassertibility view of truth for discourses for which anti-realist metaphysics seems appropriate or Gibbard's (2003: 248) view that our

inspired by some of Blackburn's (1984: 198; 1996; 1998a: 318) ideas just for exposition. According to this view, moral judgments are truth-apt because they are apt for improvement in a way that delivers a notion of the best set of attitudes. Given the specific improvements that Blackburn mentions —being more informed, imaginative, coherent, sensitive to finer distinctions, etc.— we should expect that, like true beliefs, only correct moral judgments would not be unsettled or would be stable through these improvements.[76] Correct moral judgments then would be stable in the way that true beliefs are.[77] This stability property (hereafter SP) is thus a proxy for truth. We can now ask whether moral judgments are regulated by SP in the way that we have seen beliefs to be regulated by truth. If yes, we should expect moral judgments also to be epistemically evaluable in the way beliefs are.

Expressivists then can argue that SP regulates moral judgments, and that is so because of the internal dynamics of moralising. Moral thinking is one kind of thinking concerned with what to do, think, and be. Insofar as our answers to those questions are not stable, those questions remain unsettled by our answers. But, given the practical orientation of moral judgments, helping to close deliberation is part of their job. The internal dynamics of moralising might, in this way, call for moral judgments that cannot be further improved. Consider a case in which an agent judges racial discrimination is morally permissible. We can plausibly think that the judgment meets at least one of the criteria for an unstable judgment: it is based on morally arbitrary distinctions.[78] Since we can recognise getting rid of that arbitrariness as an improvement, then the judgment is not stable. Thus, insofar as ordinary moral thinkers do not count moral judgments based on arbitrary distinctions as true, expressivists can explain this by saying that such judgments lack SP. Moreover, as long as this judgment is known to be unstable because of arbitrariness, it will not settle our questions about what to do, think or be.[79] In this example, the

judgments about the ideal end of inquiry are judgments about deferring to ideal hyper-decided agents. See also Sinclair (2021: 193) and Gamester (2022) for different applications of a similar view.

[76] Here it may be important to mark the contrast with imaginings and suppositions. While it is not impossible to think of instances of these states that cannot be further improved, these are unlikely to be the *same* improvements beliefs and moral judgments are apt for.

[77] I call this a Blackburn-inspired view because it captures what he believes to be the truth-like aspects of the objectivity we want to reach in moral inquiry. However, it is unclear whether he would accept it as a complete view of truth. See Blackburn (1998b).

[78] Some may argue that racism can be based on partiality which, even if morally problematic in other ways, is not arbitrary. However, the sense of arbitrariness that I suggest here we care about our judgments with SP lack, is that of distinctions that are not *morally relevant*. Arguably, that is also what is wrong with racism based on partiality. See Crisp (2018).

[79] Notice that the view is not that non-stable moral judgments cannot close moral deliberation, but that they cannot in principle close moral deliberation without *aiming* at being stable. Just

internal dynamics of moralising provide the standards or improvements by which moral judgments are deemed stable. Arbitrariness is just part of the standards that I suggested ordinary moral thinkers would endorse. Thus, the sense of 'settling' here is not just psychological; it is normative. In this sense, SP has substantial moral content.[80] But this is not unlike other minimalist views of truth. As Crispin Wright, who proposed a similar stability property called superassertibility, explains:[81]

> "[T]the crucial reflection is that superassertibility [or SP] is, in a clear sense, an internal property of the statements of a discourse a projection, merely, of the standards, whatever they are, that actually inform belief formation and assertion within the discourse. It supplies no external norm —in a way that truth is classically supposed to do— against which our ordinary standards might themselves be measured sub specie Dei and might rate as adequate or inadequate." (Wright, 2021: 590, my clarification)

SP is a proxy for truth in this general minimalist sense. Thus, it should give us a good analogy between moral judgments and beliefs. From veritism, we knew that:

> B) A belief that *p* is correct iff *p* is true,

Now we also know that:

> M) A moral judgment that *p* is correct iff it has SP.

A disanalogy between M) and B) is that M) includes stability, whereas B) includes truth. This may be problematic since the correctness of beliefs is a property primarily of their contents, the proposition that *p,* but the correctness of moral judgments is a property of the *attitude*. However, first, expressivists are likely to say that they do not want to deny that there are moral propositions that are the contents of moral judgments. What is tricky for them is to give a convincing characterization of moral propositions whose function is not to represent moral properties —which is not to say that they have not offered plausible candidates.[82] And given how well SP models the truth-like properties of correct moral judgments, whatever the right

---

like an agent cannot close theoretical deliberation with a belief that is false *by their own lights*, an agent cannot close moral deliberation with a moral judgment that is not stable by their own lights.

[80] This is why the thought and talk of the relevant improvements is often characterized by expressivists second-order *normative* judgments about how to best form our moral judgments. See Sinclair (2021: 194) and Gamester (2022: 451-3).

[81] According to Wright this element is what makes both minimalist and most forms of deflationism *pragmatist* accounts of truth. See Shieh (2018: 454) for a similar point.

[82] See for example Schroeder (2013), Ridge (2014: 124-31) and Brown (2019).

characterization of a moral proposition is, it should draw a very tight connection between moral judgments with true contents and stability.[83] Second, the distinction between SP being about the attitude or the content is not as straightforward. As we have been treating SP, whether the judgment that racial discrimination is wrong survives through the relevant improvements depends on the wrongness of racial discrimination, not on aspects of the attitude itself —more on this later.

There are interesting questions about the difference between the role of truth and SP in B) and M), but for expressivists that appeal to minimalism, the difference should not be problematic. One thing to note, however, is that expressivists are likely to have different views about the relation. Some are happy to say that beliefs and moral judgments sharing SP is indicative of some general conditions for applications of a single truth-predicate. Applications of this predicate to moral judgments and beliefs, even if different in the details, are just instances of the same general schema. Others would say that SP just indicates something similar enough to truth to merit the similar language. But ultimately, what we get from it is just a proxy for truth in the case of moral judgments.[84] Both views would have a way to establish a parallel to veritism, but let us stick to the latter as it is the less committal. Using '$\text{true}_p$' as a notation for a proxy for truth, we could restate:

M2) A moral judgment that p is correct iff p is $\text{true}_p$

However, this would just be a notational variation to highlight the parallels between M) and B). What does the real explanatory work for the analogy to work is SP —this is what captures the platitude of stability.[85]

The only remaining disanalogy between them is the kind of questions beliefs and moral judgments are meant to address, and thus what kind of reasoning it is at issue. Whereas M) seems pertinent for practical reasoning, B) seems pertinent for

[83] How tight the connection must be is controversial. A plausible articulation is that any view of the content of moral judgments should meet the following condition: a moral judgment that *p* is true iff the judgment has SP. However, as suggested in footnote 75, that depends on whether expressivists, as anti-realists should accept that truth is epistemically constrained. That is a more general problem about truth, not specific to correctness.

[84] Gibbard (2003: X), for example, is explicitly non-committal about these two options.

[85] There is a well-known objection to similar views based on a form of stability. The idea is that they cannot make sense of fundamental moral fallibility: that a judgment may be wrong even if it is stable through our best ideas of improvement. See Egan (2007) and Köhler (2015). However, expressivists have a wide array of responses to that challenge. See for example, Blackburn (2010), Ridge (2015), Horgan & Timmons (2015), and Bex-Priestly (2018). My own view is similar to Gamester's (2021: 461-2) that it is unclear expressivists need to make sense of such a sceptical scenario when they are trying to vindicate ordinary moral inquiry. See also Cosker-Rowland (2017) for a related point about fundamental moral disagreement.

theoretical reasoning. And one of the main elements of the challenge is that desires are apt for practical evaluations, not epistemic. In response, given the expressivist picture of moral thought as practical, the difference between M) and B) is by design but should not be problematic. According to expressivists, moral thinking is practical, so the questions we address in moral inquiry will always have this practical content: they will always be about what to do, think or be. In this sense, the regulation of moral judgments will be essentially practical in its *subject matter*. However, what ultimately decides whether a state meets the correctness conditions is truth or the SP. Thus, the reasoning will be theoretical *in its issue*.

There is a separate, more pressing issue. Desire-like states plausibly also have correctness conditions that specify a constitutive norm —just like B) is a correctness condition *and* a constitutive norm of beliefs. This makes sense if beliefs are, by their nature, epistemically evaluable states. But if desire-like states have a constitutive norm, it cannot be M). While M) is fitting for moral judgments and compatible with their desire-like nature, it seems like the wrong fit for desire-like attitudes in general. Ordinary desires, for example, are not correct or incorrect, if they ever are, when they have or lack SP. Even if we could think of desires that would be unsettled by improvements of consistency and so would be incorrect by some standards, like prudential or pragmatic, it is difficult to think if it as incorrect *as a desire*. The correctness or incorrectness does not seem to be constitutive like it is for beliefs and moral judgments. It would seem then that moral judgments do not have the constitutive norm of desire-like states, since they are correct when they have SP. But expressivists have strong reasons to say they *are* governed by the norms of desire-like states, given their commitment that the desire-like nature of moral judgments is the best explanation of the practical function of moral thought. Perhaps M) is not a constitutive norm of moral judgments. But then, expressivists would lose a strong parallel with veritism to secure epistemic evaluability.

Expressivists could argue that this is another non-problematic disanalogy. Perhaps we *treat* moral judgments as epistemically evaluable in moral practice, following M), but their constitutive norm is desire-like. But that is unlikely to convince critics. After all they can reframe the challenge of epistemic evaluability as one of asking why we can treat moral judgments like this in moral practice, given that they are desire-like states. A different, more promising option is to argue that M) is specific to moral judgments but necessarily connected to a constitutive norm that applies to desire-like states more generally. It is not uncommon for expressivists to argue that the belief-like surface features of moral judgments, like epistemic evaluability, can

ultimately be explained via their practical desire-like function.[86] Gibbard's view of truth evaluability is a good example. For him, even if disagreement is what explains some logical properties of plan-laden states, like moral judgments, it is fundamentally an element of their desire-like practical nature. We need to be able to disagree in plan in order to respond to our practical questions of what to do, think or be. This function, even if it can explain belief-like aspects of plans, is still very much characteristic of the desire-like side of the Humean mind. And that strategy should be useful here, too. But to see why, we need to know more about the nature of desire-like states.

## III: Regulation: The Aim of Desire

A fruitful way to explain how M) is not only not incompatible but necessarily connected with the desire-like nature of moral judgments is to appeal to a substantive theory of desires. The idea would be that just as beliefs aim at truth, desires also have a constitutive norm that, in the case of moral judgments, affords a relevantly similar form of regulation. And a theory that fits the bill is the evaluationist or guise of the good theory of desire. According to evaluationism, desires present their object under a positive mode of presentation —a "guise of the good".'[87] Think of an ordinary desire for coffee. A proponent of an evaluationist view would say that even after listing all the relevant psychological (and in general causal) facts of how and why I ended up acting to get the coffee, to understand my action as intentional, we still need evaluative of normative concepts to explain why that action in particular made sense to me, why it was desirable. By appealing to evaluationism, expressivists can argue that moral judgments, like all desire-like states, aim at what is desirable. But, unlike the rest of the desire-like states, for moral judgments to aim at what is desirable, they must aim to have SP.[88]

[86] For a list of representative examples and an explanation of this strategy see Sinclair (2021: 37).

[87] For more on the role of this view in the literature on the nature of desire see Lauria & Deonna (2017: 6). For a general overview of the position see Orsi (2015) and Tenenbaum (2021). For existing applications of evaluationism in moral epistemology see Schroeder (2007: 170), Tenenbaum (2008) and Milona and Schroeder (2019).

[88] Appealing to some forms of evaluationism can also be also a good way for expressivists to flesh out their Humean division of the mind. While they mostly appeal to a difference in direction of fit (see, for example Blackburn 2005: 323), nowadays that distinction is controversial, and it is often suggested that we should replace it with a form of evaluationism. See Gregory (2012), Chan (2013: 2) and Ortiz-Millán (2018).

For our purposes here, the relevant form of evaluationism has some important parallels with normativism.[89] According to both views, normative and evaluative concepts are essential to our accounts of the nature of mental states, like beliefs and desires. For normativism to believe that *p* is to think of *p* as true, thus, *correct* beliefs are true. Evaluationists think something similar about desires; to desire that *p* is to think of *p* as having something in its favour, as being, so to speak, desirable.[90] And so correct desires are for what is desirable. However, unlike normativism, for evaluationism, normative and evaluative concepts have a further role: we need them to explain this sense of desirability. Perhaps *p* is in some sense *good* or is such that there are *reasons* in its favour.[91] We do not have to settle for one of these options, so I will keep using 'desirable' as a neutral stand-in. What matters for us is how this sense of desirability helps explain the correctness of desire-like states:[92]

D) A desire-like state that *p* is correct iff *p* is desirable.

Desires are the inputs in their practical reasoning, things that other things being equal should make a difference in what the agent decides to do.[93] But for the desire to help in *rationalising* whatever the agent decides to do, to think of that as the product of practical reasoning, the desire needs to be for what is desirable. In the case in which I got a coffee, to explain that as a product of my practical reasoning, we need to appeal to a desire to get coffee. But that desire will not help us rationalize the action if we do not think that in desiring it, getting coffee enjoyed some positive standing in practical reasoning, i.e., it was desirable.[94] Maybe there were reasons to get coffee, like that I was tired, and thus being desirable in that sense can help us make sense of that action as the product of my practical reasoning. It is in this sense that desires *aim for* what is desirable or are *correct* when they are for what is desirable. This is a

[89] My view here is based in Sergio Tenenbaum's (2007, 2008, 2020) but differs in some crucial respects.

[90] I will understand this content as propositions that include some course of action or response. This course of action or response is what enjoys the positive standing in practical reasoning. See Sinhababu (2015) for the advantages of this approach and Milona & Schroeder (2015) for discussion.

[91] For this variation in terms of reasons see Bedke (2009) and Gregory (2013; 2021).

[92] Talking about correct desires may seem counterintuitive. However, we often use concepts like *fittingness* to talk about this same sense of correctness, e.g., a desire is fitting when it is for what is desirable. See Howard & Cosker-Rowland (2022). Everything that I say here can be restated in terms of fittingness.

[93] It is controversial whether desires figure in practical reasoning as premises or in the background. See Pettit & Smith (1990). As long as desire-like states are needed to reach the conclusion of practical reasoning, both views should be compatible with the account I present here. Thus, I will use the term 'input' to remain neutral.

[94] This does not mean that the desire figures in the foreground (as noted in footnote 87). This is why it is the *object* of desire, getting coffee, which enjoys the positive standing, and not the desire itself.

rough sketch, and we will flesh it out, but for now, it should be helpful to see why evaluationism is a good framework to think of the nature of desire-like states in general and moral judgments specifically. Like all desire-like states, the function of moral judgments is to help us conclude practical reasoning; thus, they should aim for what is desirable. But to conclude practical *moral* reasoning, being desirable in any general sense is not enough; we also need SP.

As per my stipulation, I will focus on first personal judgments of right and wrong actions, understood as approvals and disapprovals. Think of the case in which I judge that it is wrong for me to lie, so I disapprove of my lying. Like all other desire-like states, it is an input of practical reasoning: all things being equal should make a difference in how I act. And thus, it must present its object, that I do not lie, as desirable.[95] So much is clear from the commitments of both evaluationism and expressivism. What is important for epistemic evaluability is how this desirability works for the specific case of moral judgments. As I explained, being desirable is to have some positive standing in practical reasoning, so it makes sense to rationalise the agent as picking that course of action. This is how SP and M) get into the picture. Practical *moral* reasoning is special in that it is restricted by the dynamics of moralising.[96] In moral practice, we ordinarily want correct responses to our practical moral questions of what to do, think or be. Only those can *correctly* close practical moral reasoning. And we know correct responses to those questions have SP —this is just M). But being the kind of thing that can correctly close practical reasoning is just part of the correctness conditions for all desire-like states. Thus, for moral judgments to play their constitutive role as desire-like states, they must have SP. In our example, my disapproval of lying is correct, as a desire-like state, iff not lying is desirable, but whether that is the correct outcome for my reasoning, given that this is a moral judgment, depends on whether the judgment has SP —being such that it would survive through improvements of coherence, imagination, etc. D) and M) are thus both fitting for desire-like moral judgments because:

MD) For a moral judgment that *p*, *p* is desirable iff the judgment has SP.

[95] As the example illustrates, this way of integrating evaluationism with expressivism should make room for judgments on what is morally *wrong*. Expressivism would want to say that such states are closer to disapprovals and do not seem to aim at the good in any interesting way, but they are nonetheless meant to influence our practical reasoning.

[96] The restriction to moral practical reasoning is important because moral judgments are plausibly not overriding. Practical reasoning is more holistic; the conclusion of it often depends not just on moral judgments but the other relevant mental states of the agent.

Say that my judgment that I should not lie is only backed by the reason that I do not feel like it. Given the dynamics of moralising, the judgment is not about something desirable, precisely because it does not have SP. The reason is arbitrary and thus should not close practical moral reasoning. Or, to think of this differently: say that a moral judgment could not have SP but be correct as a desire-like state. This would amount to recognizing that improvements would unsettle the judgment, so it should not close practical reasoning, but at the same time, thinking it should close it anyway. It may not be impossible to think that, but it makes it harder to see that judgment as *correct*. This is the picture I argue expressivists should use to explain why moral judgments share the constitutive norm of desire-like states, D), while having correctness conditions appropriate for an epistemically evaluable state, like M). The key, then, is that the dynamics of moralising restrict practical moral reasoning such that for a moral judgment to be correct, to aim at what is desirable, it necessarily aims at having SP. Before we move on to the third and final element of the view, evaluation, we need to deal with a few possible objections.

### III.1. Problems With Mixing Expressivism and Evaluationism

There may be some worries about the compatibility between expressivism and evaluationism. The evaluationist strategy is to account for desires in terms of evaluative or normative concepts, but expressivists want to explain our use of evaluative and normative concepts by appealing to desire-like states. Thus, for expressivists, evaluationists have the order of explanation backwards. There are a few things to say in response on behalf of both expressivists and evaluationists. Take expressivists first; they generally accept that they have good reasons to accept that one way or another, they will need to appeal to normative or evaluative concepts to establish their theories of moral thought and language.[97] For example, theories of meaning and content that use normative terms seem independently plausible, and expressivists would not want to deny themselves the useful theoretical tools they offer.[98] This would perhaps mean abandoning foundationalist projects, according to which, at a fundamental level of explanation, we do not use any normative concepts. However, expressivists are free to adopt a more holistic approach for which there are no fundamental levels of explanation, and they are allowed to use different parts of the theory depending on which element of moral practice they are explaining (see

[97] Some central examples are Gibbard (2005; 2012), Toppinen (2015) and Dreier (2022). Blackburn (2002: 125) even acknowledges that there is something fundamentally right about the guise of the good view of desires and expressivists would do good in explaining it in their own terms.

[98] For more on these theories see Glüer, Wikforss, and Ganapini (2022).

Toppinen, 2015). This also would not mean that the theory is circular. Explaining the judgment that it is wrong to lie using concepts like correctness is not to appeal to the wrongness of lying; it is not explaining something in terms of itself (Dreier, 2022). The issue is not whether expressivists would use normative concepts in their theories of the correctness of moral judgments but *how.*

As Jamie Dreier (2022) explains, it would be problematic if expressivists would appeal to normative concepts in a way that can only be explained by the existence of normative *facts* or *properties*. But it is unclear that expressivists would be doing that when appealing to evaluationism, as I suggest. Rather we are talking about natural psychological facts like the existence of desire-like states, but given that we are explaining their correctness, we need normative concepts to pick them out. The constitutive norm of desire-like states is essentially connected to practical reasoning. But how we think about the correctness of the reasoning and the outcomes is not a matter we can resolve by theorising about the nature of desire-like mental states; it is a matter of first-order normative inquiry. For example, as a desire-like state, my judgment that I should not lie should help me conclude practical moral reasoning. But whether that is the correct outcome, expressivists should say, is a first-order matter that cannot be solved without moralising.

Of course, some evaluationists think of the correctness of desires as accurately representing normative or evaluative properties. After all, the correctness in D) can be read as the match between how the desire-like state represents the normative properties and how they really are. But there is no general agreement among evaluationists about that. Tenenbaum (2020: 235) offers an informative taxonomy. A content version of evaluationism holds that to rationalize desire-like states, we need to think of them as representing normative or evaluative properties as part of their contents. In the example in which I want coffee, the idea would be that the content of my desire perhaps implicitly includes the proposition that getting coffee is good or that getting coffee is something I have reasons to do. Correctness is just a matter of whether the evaluative or normative concepts that figure in that proposition represent normative properties. These are the evaluationist views that expressivists should reject. An attitude version of evaluationism would instead propose that we need normative concepts to understand the way the agent thinks of the content as desire-like. This version seems, in principle, compatible with expressivism and could be fleshed out in a few different ways. One is the one I am suggesting here that rationalize an agent's intentional actions by attributing desire-like states; we need to employ normative or evaluative concepts to understand how that state fits in the

agent's practical reasoning. For example, for the desire to get coffee to rationalize the agent's behaviour, we need to think of getting coffee as the kind of thing that can correctly conclude practical reasoning; as desirable.[99]

A different worry here could be that the way a norm like D) seems to govern moral judgments in a way too different from ordinary desires for us to think that it is the same norm governing all desire-like states. Desires aim for what is desirable, so they present their object as such, as something that could conclude practical reasoning. However, ordinary desires involve no specific commitment to the actual desirability of the object. In desiring more coffee, it may appear good or as something I have reasons to get, even if I, all things considered, do not think it is. In that respect, desires seem closer to perceptions than beliefs (see Tenenbaum, 2008: 38-42). Moral judgments are not like this; they involve a personal-level acceptance or commitment (Björnsson, 2001; Björnsson & McPherson, 2014: 12). If I judge that it is wrong that I lie, I am not merely thinking that it appears or seems wrong. Rather I am committed to the action actually being wrong. But this seems to be part of their desire-like nature, too, since it seems that in judging that lying is wrong, we are committed to the desirability of not lying, which is something that characterizes desire-like states in general. In response, recall that D) is supposed to be a norm that governs different desire-like states, not only ordinary desires. Some evaluationists argue that a similar personal-level acceptance or commitment is important to explain other desire-like states like intentions (Tenenbaum, 2007: 51; Gregory 2012: 607, 2021: 45). When I intend to get coffee, my intention is also correct when it is for something desirable, but in intending it, I manifest a commitment to its desirability; that getting coffee is good or there are reasons in its favour.[100] Moral judgments are closer to intentions than ordinary desires in this respect —although they are not completely similar either.[101]

[99] Tenenbaum's (2008) own view is an attitude version of evaluationism. However, it is different from the view that I present here in that he argues that we need to posit a further attitude of thinking of the content of desires as good.

[100] A version of evaluationism based on reasons is better equipped to explain this than a version based on goodness. Consider, for example, that in a trolley problem a deontologist can *intend* to not pull the lever to save five people even if they judge that it would be best to pull it. See Cosker-Rowland (2019a: Ch. 2). This however is a conflict between versions of evaluationism on which expressivists can remain relatively neutral.

[101] It does not seem like stability is important for the correctness of intentions. They can even reflect characteristic forms of instability like arbitrariness. And even if be some limits to this arbitrariness (see, for example, Broome, 2013: 177), these are not substantial enough to pair them with moral judgments. On the other hand, intentions are often thought of as conclusive of practical reasoning, whereas moral judgments are not necessarily. See Toppinen (2017) for a more detailed explanation of this aspect.

Lastly, one element some think to be central to normativism is the transparency of beliefs. If I wonder what I believe, I do not consult my mental states but the elements of the world they represent. This is especially important to veritism since it makes sense that the means for a correct belief are means to get how things are in the world right; epistemic evaluations stand for those means.[102] The worry, is that it is hard to make sense of the same phenomenon for moral judgments given this picture of their correctness so tightly connected to their desire-like nature. The first thing to note here is that expressivists already want to account for a similar phenomenon. If I wonder whether racial discrimination is wrong, it better not be that expressivists claim I should consult what I disapprove of. That would mean that what is right or wrong depends on what I approve and disapprove of, which is a form of subjectivism that most expressivists would reject.[103] This is also consistent with evaluationism, since, according to it, when I consider whether I want something, I do not consult my mental states but the features of the object. There is then a good question of what, according to expressivists, we consult in the specific case of moral judgments if it is not the judgment itself either. I believe this is a very fruitful area to explore further. For example, one important question here is whether we should talk about moral *evidence* and whether expressivists can also earn the right to that notion. I will come back to the issue of evidence for moral judgments in Chapter 4. For our purposes here, it suffices to say expressivists are already committed to the view that we should not consult our moral judgments themselves, and the picture I present here supports it.

### III.2. Problems Specific to Expressivism

One concern about this proposal is whether this picture can be generalised to other kinds of moral judgments beyond first personal judgments about right and wrong actions. Think of a different example of a judgment that racial discrimination is wrong. There is no single action in the content, so nothing that, in principle, can be desirable. In response, expressivists have some existing resources to deal with this problem. Even if a judgment that racial discrimination is wrong has no specific action in the content, given its practical nature, there are certain things that we would expect the agent to do and refrain from doing when sincerely uttering it or judging it. Perhaps avoiding that kind of behaviour, feeling or responding in certain ways in the presence of it, approving of other people's rejection of it and disapproving of their approval, etc. Many expressivists then prefer to understand moral judgments as

[102] See for example Shah (2006)
[103] See for example Schroeder (2014).

complex dispositions to act and think.[104] Even if these dispositions are not part of the content of the approval, it is those that expressivists should say are desirable: they should make a difference in how the agent ends up doing.[105] Thus, the evaluationist picture offered should be, in principle, extendable to other kinds of wrongness and rightness judgments.[106] That is not to say that this exact same story can be extended to *all* kinds of moral judgments. This is not the right place to offer a fully general account. However, on the one hand, my goal here was modest, just to argue for the *possibility* of moral judgments being desire-like and epistemically evaluable. Second, expressivists should have some licence for optimism since, assuming their non-cognitivism, it is very plausible that moral judgments are generally unified by their desire-like nature. Thus, if they are all governed by a norm like D) in one way or another, then there are plausibly similar ways to connect them with M).[107]

A different worry is that the account here could prove too much; that moral judgments are too similar to beliefs. This would, problematically, blur the distinction between cognitivism and non-cognitivism. Normativism is a view about what is constitutive of beliefs, which should fundamentally distinguish them from other mental states. If my argument here works, however, it could invite well-known problems of creeping minimalism.[108] If moral judgments can mimic beliefs even in their constitutive norms, then one can start wondering why they are not just beliefs. This should be a welcome result for expressivists; after all, they are trying to earn the right to treat moral judgments as beliefs. However, if there is nothing to

[104] See for example, Blackburn (1998a: 68), Gibbard (1990: 132), Björnsson & McPherson, 2014) and Toppinen (2017).
[105] Following Björnsson & McPherson (2014: 15) some of these dispositions like the social hostility to people engaged in the act judged to be wrong, are what characterizes these judgments as *moral*.
[106] As suggested above this may be easier to articulate in an attitude version of evaluationism, since the focus is not on the contents but in how we think of them. Still, an interesting question is how those dispositions fit in that picture. This is a broader question about attitudes and contents for non-cognitivists. One option is to think of those dispositions as elements of the conceptual role needed to determine the content of the judgment in the first place. See Köhler (2017).
[107] A trickier case is that of non-atomic moral wrongness and rightness judgments. Think for example, of conditionals like if one breaks a promise then it is morally right to apologize. Although not completely analogous to other moral judgments (it cannot be expressed in an indicative form), it seems it can be epistemically evaluated. I can, without any evident conceptual confusion, think that I know that if one breaks a promise then it is morally right to apologize. Although I agree that it seems unproblematic that complex states with conditional structure are epistemically evaluable, what is controversial is whether we need a unified account for the epistemic evaluability of conditional and non-conditional beliefs and moral judgments. Against this see, for example, Lewis (1976). I will not address this issue here as to not side track the discussion. Thanks to Michael Ridge for pointing out this issue.
[108] See Dreier (2004) for the classic articulation of this worry and Cuneo (2020) for one closer to these epistemic issues.

distinguish beliefs from moral attitudes, then there seems to be nothing that separates expressivism from its cognitivist counterpart. This problem seems more pressing in this context since usual responses to this problem appeal to the explanatory indispensability of a notion of moral facts. Even if expressivists earn the right to a cognitivist talk, their account makes no substantial use of moral facts. However, neither veritism nor the expressivist counterpart I offer here appeals to facts either, so the difference cannot lie there.

I already hinted at a possible response before. Expressivists adopting this account can say that even if moral judgments and beliefs both respond to a similar aim of truth, the kind of regulation they are subjected to differs. Moral judgments aim at presenting their object not as it is but as desirable. In that case, even if the regulation is belief-like in its issue because it responds to the aim of truth, it is still essentially practical in its subject matter: it is still about practical matters of what to do, feel or be. Thus, when we distinguish between beliefs and moral judgments according to their norms, the difference is not in the norms themselves but in the subject matter and how we should understand its correctness. As I explained, the Humean division of the mind is at the heart of expressivism, but how to draw the distinction across the different contexts in which it is relevant is not often discussed.[109] I hope that the way I connect it with evaluationism here to respond to a specific challenge opens the door for further discussion.

## IV. Epistemic Evaluation

We now have the two first elements of the expressivist version of veritism: truth and regulation. Let us call the combination of these two elements *stabilism*. To turn stabilism into a complete analogue of veritism, we are still missing the last piece: evaluation. In our ordinary moral practices, we think of moral judgments as epistemically justified, warranted, rational, etc. Expressivists then need to explain how the aim of moral judgments, SP, grounds *those* evaluations. This is not the right place to offer a full account of all these evaluations. Each one has distinctive aspects that require individual discussion in the context of expressivist moral epistemology, and the discussion in the second part of the dissertation will focus on some central examples. Instead, I propose to explore the general but substantial connections between stabilism and veritism. This will give us a bridge to connect my discussion here with what will come in the next three chapters.

[109] See, for example, Sinhababu (2017: 194-7) for an objection in this vein.

The means to form moral judgments with SP are recognisably epistemic in at least two respects. Truth helps us demarcate epistemic evaluations from the rest. Epistemic evaluations, unlike practical or prudential ones, concern truth. And SP is just our proxy for truth: it captures what is platitudinous about truth. But for expressivists, this proxy should be enough, given that they are likely to endorse a form of minimalism or deflationism. Moreover, SP captures the truth platitudes that correct moral judgments satisfy but leaves the substantial issue of which judgments are true and why open for first-order moral inquiry. In that sense, and following the rules we set out, the evaluations that stand for those means will necessarily be epistemic. One may argue that, unlike beliefs, moral judgments aim only indirectly at a proxy for truth, and primarily aim at what is desirable. However, means are not hyperintensionally individuated. If *m* is a means for *x*, and something has *x* iff it has *y,* then *m* is a means for *y*. And we know that a moral judgment is about something desirable iff it has SP. Thus, even if the regulation is different in this respect, the means to form correct moral judgments should still count as epistemic evaluations via our proxy for truth.

Even so, I believe that expressivists can claim an even tighter connection between their views and veritism. The central point of veritism is that beliefs are states that involve a kind of *commitment* to things being a certain way and are regulated by how they get it right or not. This was important for epistemic evaluability since, from the epistemic point of view, we want to get things right in matters that are important to us, and having true beliefs is an essential element of that goal. But truth is not always transparent to us, so we need evaluations in terms of the means to get it right. Expressivists want to say something very similar to desire-like moral judgments. Even if they have a different desire-like function, from the epistemic point of view, we also want true moral judgments —perhaps we want them even more since moral matters are, by definition, important. And, moral judgments involve commitments too, not about how things are, but on what morally to do, think or be. And they should be regulated by how they get these things right. Moreover, since moral matters are generally very complex, what is the right response to them is not always transparent to us. Hence, we plausibly also need evaluations that correspond to the processes, methods, norms, or grounds that could plausibly produce the correct judgments. Given these parallelisms with veritism, the means to form moral judgments with SP should be distinctively epistemic.

There are two further specific connections we can draw with epistemic evaluations. First, the view here can, in principle, make sense of complex cases, characteristic of

the epistemic evaluations of beliefs. Think, for example, of true but unjustified moral judgments. One example could be a judgment that racial discrimination is wrong held on the basis that I do not like it. The judgment is likely to be stable; it would survive our best improvements. However, the means to get it right are characteristic of unstable judgments: it is based on arbitrary considerations. Similarly, even if they are controversial, we can make sense of the possibility of justified but false beliefs. These are, plausibly, judgments formed in ways characteristic of stable ones but that are not members of the best set of attitudes. For example, I can form the judgment that it is morally permissible to eat shrimp based on misleading evidence that they cannot experience pain. This would be a judgement formed in the way characteristic of stable judgments, but that is not itself stable.[110] These are, of course, just general outlines. What expressivists should say, in addition, depends on how we should think of epistemic justification for moral judgments more generally. That is a complex issue that, as indicated, I cannot address here. One thing to note, however, is that, as we know, most expressivists are likely to already have some views on epistemic evaluations. They endorse a form of epistemic expressivism. Hence, our second connection, stabilism is compatible and complementary to epistemic expressivism.

We discussed epistemic expressivism in further depth in Chapter 2. This is the view that normative epistemic talk, like that of epistemic evaluation, has the function of expressing desire-like or non-representational mental states. Thus, if we want our account of epistemic evaluability to fit what expressivists generally think about epistemic evaluations, epistemic expressivism is the right place to look. A problem presented in Chapter 2 was that epistemic expressivism assumed elements of views like veritism that expressivists have not earned the right to yet. But with our account of epistemic evaluability in place, we can see how these views complement each other. As Sinclair (2021: 57) explains, even if expressivists want to use the same framework to explain all normative talk, like epistemic, moral and prudential evaluations, they also do not want to deny these evaluations are substantially different. To mark some of these differences, expressivists can index the evaluation to a goal specified by the context. And for epistemic evaluations, the goal can be truth. Thus, an epistemic evaluation expresses an attitude in favour of the means to get to the goal of truth —or an attitude against means that hinder it.[111]

[110] A question here would be whether there could be false but justified *fundamental* moral judgments. This is an important and well-known problem for expressivists that I cannot address here. For responses to this problem see Blackburn (2009) and Gamester (2021).
[111] In terms of Ridge's (2007) version of epistemic expressivism this is what should guide our epistemic endorsement of belief/judgment formation procedures.

## V. Conclusion

In conclusion, beliefs are epistemically evaluable because they constitutively aim at the truth. My proposal in this chapter was for expressivists to argue that moral judgments, even if desire-like, can also aim at a proxy for truth. For this, all we needed is, on the one hand, a more substantial sense of truth as stability for moral judgments with which expressivists are already familiar from their responses to other problems. And, on the other, a sense of regulation that, even if grounded on the desire-like nature of moral judgments, still fits perfectly with the expressivist account of truth. This gives us a sense of aiming at the truth for epistemic evaluations to get a grip on and helps expressivists solve their challenge. Individual epistemic evaluations and some of their central aspects can present different challenges to expressivists, but with this account of epistemic evaluability, they can rest assured that they are not committing a category mistake in pursuing those. And now, with the problem of epistemic evaluability out of the way, we can proceed to discuss epistemic evaluations.

# Part II: Epistemic Evaluations

# General Introduction to Part II

Epistemic evaluations are diverse. To take just a few examples: knowledge, ignorance, understanding, justification, warrant, entitlement to believe, epistemic (ir)rationality, epistemically virtuous/vicious, etc. —we can even have subcategories for these examples like the different epistemic virtues and vices. For a complete moral epistemology, expressivists need to show how they can earn the right to notions of epistemic evaluation and not just epistemic evaluability. Moreover, an account of epistemic evaluability, like the one offered in the first part of the dissertation, is conditional on being the right fit for our best accounts of epistemic evaluations. If moral judgments aim at the truth, in a sense that expressivists can explain, it better be that the means to attain that truth correspond to our best accounts of things like justification, warrant, rationality, and the rest.

Accounting for epistemic evaluations is complex, and finding the best approach is difficult. One option would be to list all the relevant epistemic evaluations we use in moral practice and then offer an expressivist-friendly account of each. This would be a slow-track account in Simon Blackburn's (1993: 184) terminology. For Blackburn, this strategy was for expressivists to earn the right to the different logically complex or non-atomic constructions of moral thought and language. However, it is easy to see how we can use a similar approach for epistemic evaluations. Even so, offering a slow-track account of epistemic evaluations would be very difficult considering the variety of evaluations we use in moral practice. A different option would be a fast-track account. For Blackburn, this amounts to a simple argument that since we need moral judgments to play all the propositional functions of belief for our ordinary moral practices but also have very good reasons to think these judgments are desire-like, then what is more likely is that our ideas that desire-like states are unfit for these roles are wrong. Regardless of how effective that approach may be for propositional functions, epistemic evaluations are so diverse and complex that building a general but simple case like that is unlikely to be convincing.

I propose a middle ground between these two options. The idea is to look into concepts central to our epistemic thought and talk. We typically use these concepts to determine whether moral judgments are justified, warranted, rational, and so on. While not exhaustive, the list will give us a comprehensive and representative case that expressivists can have adequate accounts not just of epistemic evaluability but

of epistemic evaluations. And I believe evidence, reliability and a prioricity are among the best candidates. These concepts offer different advantages for my project here. Evidence and reliability are concepts that tap into a central element of a number of our epistemic evaluations: a non-accidental connection with truth. We ordinarily think of the truth of evidence *e* to be non-accidentally connected with the truth of the belief/moral judgment that *p* for which it is evidence, such that the former indicates the truth of the latter. And we ordinarily think that when a process *P* is reliable, it can systematically produce or regulate true beliefs/moral judgments in the right circumstances. Thus, evidence and reliability can give us representative examples of how expressivists think of this kind of epistemic evaluation for moral judgments. A priori moral knowledge is different since epistemic evaluations cut across different sources like a priori and a posteriori. Just as we have unreliable and reliable processes to form true beliefs about external objects, we have reliable and unreliable processes to form beliefs about mathematical, modal, or logical truths. We have good reasons to think that the sources of at least some of our moral knowledge are a priori, too. And expressivism would be in trouble if it could not account for that.

Evidence may seem like the odd candidate among the three. Existing moral epistemologies rarely appeal explicitly to the notion of evidence for moral judgments. However, when I claim that expressivists should account for evidence, the phenomenon I have in mind is our use of pieces of moral and non-moral information to support our moral judgments. Think of cases when one, for example, epistemically supports a judgment that eating shrimp is permissible with considerations that they do not feel pain or do not have a central nervous system. This notion of support is widespread in moral practice. The decision to call it evidence may be merely terminological. However, one benefit of using a framework of evidence to explore this relation of support is that evidence is helpful in fixing facts about various epistemic statuses like justification, knowledge, rationality or certainty. These are statuses that we do ascribe to our moral judgments in moral practice. Thus, while we rarely use the *term* in moral practice, the *concept* should be helpful to fix facts about these statuses, and so it should be relevant for expressivists that want an explanation of epistemic evaluations.

Think of justification first. For a number of substantive conceptions of justification that can help us nail down the concept we use to evaluate moral judgments, what anyone is justified in believing is determined by facts about the evidence. The main example here is evidentialism, according to which justification is determined solely by a subject's evidence (i.e. Conee and Feldman, 2004). But even if we think that

evidence plays a less decisive role in justification, it still plausibly plays a substantial one in determining whether a belief/moral judgment is justified. This partial but central role can work the same for other epistemic evaluations like rationality or certainty. Some might object that something that plays a similar role in the epistemic evaluations for moral judgment and would be less controversially applied to them is reasons for belief. However, the relation between reasons for belief and evidence is complex. On some conceptions, and given that we specify we are talking about epistemic (or the right kind of) reasons for belief, they are equivalent. And so, the choice of discussing evidence would be, again, merely terminological. But even in most conceptions where reasons for belief and evidence mismatch, evidence seems to be the broader notion. For example, views according to which a small raise in probability would amount to evidence but not substantial enough to be a reason for belief. Even in that case, we have a good reason to frame the discussion in terms of evidence.[112] In any case, I will have more to say about this notion of epistemic support that I am characterizing as evidence in Chapter 4.

Reliability, in the sense expressivists should explain it, plausibly consists of judgment-forming processes that could *systematically* and *non-accidentally* produce true moral judgments. This notion is more prominent in our ordinary epistemic evaluations for moral judgments. We care a lot about how we form our moral judgments: we, for example, reject judgments formed with biases or formed partially when the situation calls for impartiality. Expressivists face challenges in securing this sense of reliability. In fact, if one issue on expressivist moral epistemology has received attention in the literature, it is reliability and debunking challenges. In short, expressivists do not want to say —at least in the first instance or in a substantial sense— that our reliability consists of the systematic tracking of moral facts. But if that is not their view, it is unclear what is. Moreover, our best evolutionary accounts of our moral judgment formation processes seem to conflict with some of their main quasi-realist commitments about reliability and truth, restricting their options even more. These challenges will be the topic of Chapter 5.

Some of our most important moral judgments, like those about fundamental or basic moral principles, seem to be knowable a priori. This fundamentality is not about the structure of moral principles or their metaphysical priority but about how we come

[112] For more on the relation between reasons for belief and evidence see Lord (2018, Forthcoming). For a classic argument for the separation see Foley (1991). And for a different argument that evidence is the broader (or even explanatory prior) notion see Star & Kearns (2008, 2009)

to know them. More specifically, it is about the *source* of our knowledge of them. Given the plausibility of this idea, any comprehensive view on epistemic evaluations of moral judgments must offer a way to explain or explain away. Although there are no existing challenges that expressivists cannot account for a prioricity, it is unclear what strategy they could employ. It seems that traditional rationalist accounts based on rational intuitions or self-evidence are not available to them. Thus, if expressivists want their accounts of epistemic evaluations to be compatible with *all* of our thinking and talking of epistemic evaluation for moral judgments, they must say something about this. I will explore these issues in chapter 6.

Before we begin, one important clarification here is that while my main focus is going to be on evidence, reliability and a prioricity as *concepts* we use in epistemic evaluations, which is the right concept we use in moral practice not always clear. Thus, I will often need to rely on more substantial *conceptions* of evidence, reliability and a prioricity to nail down a rough characterization of the concept and work initially. These conceptions will often be those offered by substantial accounts like reliabilism, evidentialism, moral rationalism, etc. However, the idea will not be to make expressivist moral epistemology dependent on these substantial views. Once we have an initial characterization in place, the options will be clearer, and the commitments can be relaxed. Thus, I will mark these points in the discussion when needed.

# Chapter 4: Evidence

The second part of the dissertation builds on my account of how expressivists should explain the epistemic evaluability of moral judgments. I will explore how expressivists should make sense of the central notions involved in the epistemic evaluations of moral judgments. As in the case of non-moral beliefs, what determines how we should evaluate our moral judgments often depends on whether they are supported by our best reasons or evidence. This notion of support will be the topic of this chapter. More specifically, I think there is much for expressivists to gain from theorising about this relation of support as evidence. Thus, in this chapter, I will explore the challenges and options for expressivists to account for this notion of evidence for moral judgments in a way compatible with their picture of moral thought and talk.

The structure of the chapter is as follows. Section I is about the specific sense in which evidence supports our moral judgments. There, I will introduce some distinctions that will be useful for our discussion. In Section III, I present what I take to be the main obstacle to an expressivist understanding of evidence: the problem of evidential fit. The idea is that expressivists need to explain what it is for both a proposition with moral content to be evidence for anything and for moral judgments, being desire-like mental states, to be supported by evidence. In section IV, I explore three options to respond. Unfortunately, these options either do not work or leave aspects of the evidential relation unexplained. In response, in section V, I propose that whatever the right account of evidential support for moral judgments is, for expressivists, it has to incorporate elements of first-order moral inquiry. That *e* is evidence for *p* bears on the question of whether *p*, which is a question for first-order moral inquiry. By incorporating that element, expressivists can overcome the limitations of other accounts. Finally, section VI applies some ideas from the holistic account to help solve the existing problem of certitude for expressivism.

Evidence is a complex subject, and many of the issues we need to discuss are subject to different controversies. To make our discussion manageable, I will identify a few theoretical choice points that will help us delineate the options and restrictions expressivists may have in understanding and answering the different challenges. The first choice point (I) is about the kinds of things that can be evidence, that is, the ontology of evidence: propositions or mental states (factive or non-factive). Second (II), the kinds of things that evidence is supposed to support, the objects of evidential support: doxastic states or propositions? Third (III), what is the right way to think

about the relation between evidence and its object of support: abductive, probabilistic, or explanatory? Fourth (IV), what is the right way to think of the relation between evidence and epistemic statuses like justification or rationality? Is it that only evidence can determine such statuses, as evidentialists would have it or something else? Lastly (V), should we think of evidence as individual relations of support or holistically?

## I. Evidence for Moral Judgments

We ordinarily appeal to independently credible considerations to support our moral judgments. In the right conditions (e.g., with the right background assumptions in place), a judgment that eating shrimp is morally permissible can be supported by the consideration that they do not have a central nervous system or support a judgment that depicting minorities with stereotypes is wrong based on the consideration that those depictions wrong members of minorities. Some may argue that the best way to frame this relation of support is not epistemic, perhaps in terms of practical reasons or wrong/right makers —more on this later. But expressivists (and metaethicists, more generally) have good reasons to consider this as a relation of *evidential* support. As we have been assuming, moral judgments get their epistemic statuses in ways similar to ordinary beliefs. And often, how we can evaluate our beliefs epistemically depends on the available evidence. Whether I am, for example, *justified* in believing that today it is going to rain is likely to depend on the evidence, at least in some capacity. A sensible way to think about evidence is as considerations that indicate truth. Moral judgments do not seem to be different from non-moral beliefs in these respects.[113] Now, to see the more substantial parallels, we need to clarify the sense of evidence and evidential support we are discussing.

How to flesh out this concept of evidential support for moral judgments is a complex issue, and much of our discussion will focus on elucidating the best way to approach that task. Thus, I will stipulate three conditions for a piece of information to count as evidence for a moral judgment. By stipulating these conditions, we will have a rough characterisation to work initially and think about different cases. Also, some notation will be helpful for the discussion here since we will often need to mark propositions that are the content of evidence. I will use '<p>' for a noun phrase referring to the proposition that *p* or the mental state with the content that *p*. For

[113] Given that the best way to start our discussion is with these ordinary relations of evidential support, I will assume initially that the units of evidential support are attitudes like beliefs and moral judgments —one of the options in choice point (II).

example, <racial discrimination is wrong> refers to the proposition or the judgment with that content.[114] The three conditions are:

<*e*> is evidence for <*p*> in circumstances C when:

(1) <*e*> supports <*p*> by indicating its truth,[115]

(2) <*e*> has an epistemically creditable standing itself.

(3) There are no undercutting defeaters for (1).

The most central of these conditions is (1). The best way for us to initially think about this relation of support is in two parts. First is that the truth of <*p*> and the truth of <*e*> are not accidentally connected with each other. Even if it is both true that <it is raining> and that <I am perceiving sounds of raindrops>, one would not be evidence for the other if it was not because their truth is non-accidentally connected. Think of how this contrasts with a case in which it is true that <I was born on May 27 1992>, and <I am easily distracted>, but those truths are not connected in the right non-accidental way, so it is wrong to use one as evidence for the other. Sometimes this connection is called evidential *fit,*[116] that <*p*> fits with the evidence <*e*>. The term is meant to illustrate the idea that the evidence paints a picture of things such that <*p*> fits well in that picture, so their truth is non-accidentally connected. This non-accidental relation can be fleshed out in different ways, for example, that <*e*> explains <*p*>, that <*e*> raises the probability of <*p*>, etc. But for now, I will refer to it generically as a relation of evidential fit. Second, the notion of evidential fit helps make sense of other things we say of evidence: that *e* indicates the truth of *p,* that evidence is a basis for different evaluations like knowledge, justification and some forms of rationality —since these, in general, seem to depend on a non-accidental connection with truth.[117] Take the first of our examples, that evidence E) <shrimps do not have central nervous systems> supports the judgment that M) <it is morally permissible to eat shrimp>. It is at least initially plausible that

[114] I take this notation from Rosen (2010).
[115] It is controversial whether there are other ways to support *p* like pragmatically. I will have nothing to say about that possibility, so I frame the presentation here to exclude it.
[116] The main example here is McCain (2014: Ch. 4). But see also Fumerton (2018) and Steup (2018).
[117] 'Fit' may erroneously suggest that the relation is of *coherence.* However, the idea is to get a more informative gloss that captures all these roles the non-accidental connection between the truth of <e> and <p> is supposed to play.

if the truth of E) and M) is non-accidentally connected, E) can epistemically support M).[118]

Although <*p*> fitting with <*e*> could be understood discretely, perhaps in terms of <*e*> being a reason to believe <*p*>, it seems to be at its most plausible when we consider how other considerations determine whether <*p*> fits with <*e*>. Perhaps that <I was born on May 27 1992>, is not evidence that <I am easily distracted> if that is bundled with the consideration that <I am a gemini>, but it may be if it is bundled with considerations about the parenting techniques popular at that time. This way of not thinking of pieces of evidence in isolation is a key element of evidential *holism* —this is choice point (V). Many views fall under that banner,[119] but what unites all of them is that they all claim relations between a discrete piece of evidence and what it is said to be evidence for underdetermine the evidential relations. Think of a different example in which you go to the doctor to check a pain in your ear, and after examining your ear canal and finding nothing damaged, the doctor concludes that you probably grind your teeth when you are asleep; you have bruxism. You cannot see how the considerations about your earache can support the judgment that you grind your teeth. But your doctor can, because of their background assumptions. For example, the doctor's beliefs, that earache without visible damage to the ear canal is often connected with grinding or about the percentage of bruxists in a specific population, etc. Holists think all evidential relations are like this: you change the background assumptions, you change the evidential relations.

Although it is a common assumption in epistemology and philosophy of science, holism has not played a significant role in discussions of evidence in metaethics.[120] I think this is a mistake. The claim that E) supports M) is at its most plausible when combined with further assumptions. Suppose we bracket bridge principles such as B) <It is morally permissible to eat creatures without central nervous systems>. And suppose we bracket other assumptions about the link between central nervous systems and conscious experiences of pain or about the badness of pain. How plausible would it then be that E) supports M)? Not very. I will assume this form of holism for the discussion here. It will, additionally, play a substantial role in solving

[118] This rough outline of what is evidential fit is partly inspired by McCain (2014: 96) and Fumerton (2018).
[119] For a taxonomy of these views see Morrison (2010).
[120] Diggin (2022), McGrath (2019), Setiya (2012) and Sinhababu (2021) all seem to assume we can understand relations of evidential support individually. There are, however, notable exceptions to this tendency like Boyd (1988) or Sayre-McCord (1996) that too endorse forms of holism.

the problems expressivists will face in making sense of evidence. How background considerations interact with <e> and <p> is contentious among holists. There are, for example, two natural ways to integrate background assumptions in the evidential relation. One is that '<e> is evidence for <p>' expresses a dyadic relation, where <e> is properly understood as a complex item that includes things that we sometimes call background considerations or auxiliary assumptions —and so given that it is a set of considerations {e} will often be a better notation. Another possibility is that '<e> is evidence for <p>' expresses a triadic relation, such that <e> is always evidence for <p> relative to the background assumptions {b}. Throughout the discussion, I will assume the dyadic interpretation, but I believe nothing substantial for our purposes hangs on it, so everything I say here can be restated in the other framework. Adopting holism does not mean talking about a single piece of information as evidence never makes sense. The most plausible explanation of why we typically call a single piece of information evidence, even when bundled with other considerations, is unlikely to be substantial. What is more plausible is that we sometimes call one consideration evidence for conversational purposes: we need to single out one member of the bundle to call attention to it, and we assume the rest are just part of the common ground.[121]

How substantial a constraint (2) is, that <*e*> needs to be creditable itself, will vary across different accounts of evidence. For some views like subjective Bayesianism, the only restriction for evidence is coherence, which is a form of creditability but one that is, at best, derivative of (1). For other views, evidence may need a more substantial sense of creditability, like knowledge or justification.[122] At this point, we do not have to worry about which of these views to pick, so I will keep referring to a generic sense of creditability. What is important to keep in mind about (2) is that it is meant to capture the plausible idea that for most ordinary cases, for pieces of information to support anything epistemically, they must have good epistemic standing themselves[123] —I will come back to it in section VII.

As for (3), we know that epistemic evaluations often depend on the evidence; however, which evaluation is licenced by the evidence will also depend on whether

[121] Cf. Lewis (1986) view about our talk of single events as causes.

[122] Williamson's (2000) view that S's evidence equals what *S* knows is one of the main examples.

[123] Even for cases of using sense perception as evidence for a perceptual beliefs most views appeal to either a general sense of reliability of sense perception of which individual cases are just instances, or to a sense of default entitlement to base our beliefs on perception as we would do with other evidence —where default entitlement is itself an epistemic status. For examples see Audi (2003) and Pryor (2000). For an overview see DiFate (2007: § 1.b)

there are defeaters. Defeaters are conditions that, in one way or another, decrease or take away the positive epistemic status of a belief or, in this case, a moral judgment. Different kinds of defeaters will interact differently with the status of a piece of information as part of the evidence.

Undercutting defeaters show the evidence and the judgment are not connected in the right way.[124] Think of cases where evidence from a perception of hearing raindrops is undercut from supporting the belief that <it is raining> by the consideration that <my wife likes to play recordings of nature sounds around this time of the day>. In our characterisation, the evidential fit is cut; there is no non-accidental connection between {*e*} and <*p*>. Thus, the presence of undercutting defeaters would conflict with an important element of (1) and so we have reasons to make that part of the initial characterisation —hence (3).[125] There are other idealised or generic senses of evidence that some pieces of information still count as evidence in the presence of undercutting defeaters, perhaps by being the kind of thing that could support <*p*> in different conditions. But since that generic sense will make no difference for our discussion, we will focus on evidence that does meet condition (3).

Not all defeaters are relevant for a general characterisation of evidence. Rebutting defeaters, for example, present considerations against the truth of the judgment. In our picture, we could think of cases of rebutting defeaters as supporting a proposition incompatible with <*p*>. These typically make a difference in what kind of epistemic evaluation we can make based on the evidence. It is less clear whether rebutting defeaters will make a difference in what counts as evidence since they are just competing considerations —and on some accounts, they even count as evidence too, but for a proposition incompatible with <*p*> (Piazza, forthcoming). There is then no need to include them in condition (3), but it will be illustrative to discuss them later.

Conditions (1) to (3) then give us a general but plausible characterisation of evidence that seems useful for thinking about how considerations like E) can epistemically support moral judgments like M). Despite these parallels, evidence in moral epistemology differs from general epistemology in a few ways. Two are especially

[124] Pollock (1986) is the source of the distinction but see also Moretti & Piazza (2018: 2847) for an overview of the issues surrounding it. Like a number of other notions relevant for evidence, defeaters are not widely discussed in metaethics, perhaps because there is already a prominent notion of defeaters for moral obligation. When I discuss defeaters here, I will be referring to epistemic defeat exclusively.

[125] For similar considerations authors like Sturgeon (2014) and Melis (2014) think that undercutting defeaters are by nature different from rebutting ones, and more central to our conceptions of evidence.

relevant. One is that the evidence supporting our moral judgments often has moral content.[126] Consider the second example in which the judgment that M2) <depicting minorities with stereotypes is wrong> is supported by the evidence that E2) <depiction with stereotypes wrongs members of minorities>. E2) has moral content; 'wrongs' is meant morally. Still, since moral judgments can have epistemic statuses, E2) can be independently creditable. E2) also seems to be the kind of thing that makes a difference for the epistemic status of a judgment like M2). The difference between E) and E2) is central for expressivists, but we need a more precise distinction to see exactly why.

E) and E2) are pieces of evidence, but what kind of *entity* is a piece of evidence? Choice point (I) is whether a piece of evidence is a proposition or a mental state.[127] Expressivists probably do not need to pick a side in this debate. But in either case, they are committed to a specific interpretation of the distinction between E) and E2). This is clear in the case of mental states. Attitudes with moral contents that can play the role of evidence are moral judgments, and so they are desire-like states with a primarily practical or non-representational function. Attitudes without moral content that can play the role of evidence, like non-moral beliefs, do not share that function; they are representational.[128] Thus, E) and E2) as mental states should also differ in their function.

There is a similar distinction for propositions, although we have less straightforward ways to draw it. Critics of expressivism used to think that one of the main commitments of expressivism is denying the existence of propositions with moral content. There was a widely shared assumption that propositions are necessarily representational entities, so they cannot be the contents of non-representational moral judgments. But nowadays, most expressivists accept that there are propositions with moral contents; it is just that some propositions do not have a

[126] To be more precise it is at least controversial in general epistemology that beliefs or hypothesis can be supported by evidence that has moral or normative content, although there are good reasons to think it may. For example, it is a common assumption in social and moral epistemology that adopting specific values can and often does affect the epistemic quality of the products of research. And some, like Anderson (2004), argue that the best explanation of this is that values can figure among the evidence. See Longino (1990) for a different example. See also Anderson (2020) for an overview of these positions and Campbell (1998) for a way to draw some metaethical implications.

[127] A classic example of propositionalism is Williamson (2000). Against it see for example Fumerton (2005), McCain (2014) and Wedgwood (2023).

[128] A problem here may come from the distinction between factive and non-factive mental states, and whether expressivists can make sense of it as applied to moral judgments. However, the distinction is difficult to mark and in the less controversial sense, states that can only link an agent to the truth are factive and those that can fail to do so are non-factive, there does not seem to be a problem for expressivists to adopt it. See Nagel (2017).

representational function.[129] To clarify this sense of function for propositions, consider the proposition that is the content of the belief that Ana is asleep. Some argue that to play this role of content, the proposition needs to represent Ana as being asleep; this is its function.[130] Since expressivists do not think that moral properties and facts play substantial roles in thought and talk, they must say that propositions with moral content are different; they are not representational in this sense. But since expressivists do not want to give up on the benefits of having an entity that plays the role of propositions, they offer different non-representational roles for them.[131] What else they would need to say about propositions *as* evidence is unclear, partly because whether all evidence really is propositional is contentious. What matters for our purposes here is that expressivists want to account for propositions with moral content and have some options, but these options would all share the commitment to a non-representational function. Thus, even if we characterise them as propositions, E) and E2) have different functions. Since expressivists can pick either side in this debate between mental states or propositions, here I will speak neutrally about *pieces* of evidence. But since expressivists are committed either way to a fundamental distinction between the nature and function of evidence with and without moral content, I will speak mostly of pieces of moral evidence and non-moral evidence, respectively.

There is a further difference between evidence in moral and non-moral epistemology. Some might object that in our ordinary practices, we do not often call these considerations evidence or think of the form of support I am discussing as evidential support. We, for example, use the term to talk about evidence for beliefs that are relevant to make a moral judgment. Think, for example, of claims like: "If they said that, then that was wrong, but we need evidence that they said it." But we do not really talk about evidence for a moral judgment.[132] This should not be a significant problem; the fact that we do not use the term does not mean that we do not ordinarily deploy the concept —as the cases I presented suggest.[133] However, some

[129] See Schroeder (2013, 2015), Ridge (2014: 124), and Woods (2017). Like all these authors I do not think that the task here is to offer a deflationist account of our talk of meaning or 'that' clauses.

[130] See King (2014).and Soames (2015).

[131] Perhaps they instead play a role of categorization of concepts. See Brown (2019).

[132] An alternative could be to think of more sophisticated practices like moral first-order theorizing. But even there people rarely talk about evidence for moral judgments. It is notable that Campbell (2019) and Zimmerman, Jones & Timmons (2019) both being recognized general references in moral epistemology only include one mention of evidence in the sense we are discussing.

[133] It is a contingent fact about languages that some of them are better equipped than others to be explicit about our assumptions related to evidence. A good example of this is that

expressivists may have an objection along similar lines: when we appear to use evidence to support our moral judgments, we are likely just appealing to something closer to a *practical* reason in favour of the moral judgment. For example, think about E2) <depiction with stereotypes wrongs members of minorities>. What E2) is about seems closer to a wrong-maker for the fact that depicting minorities with stereotypes is wrong or a practical moral reason not to depict minorities with stereotypes. The objection then is that the support for moral judgments we are after may be explained just by appealing to these things, and so there is no need for expressivists to worry about evidence.

I will come back to this issue in section V. For now, notice that I could think that E2) supports thinking that M2) even if I did not think that what M2) is about, stereotypes, is morally wrong *because* of how it wrongs minorities, as E2) is about. Maybe, for example, I think that stereotypes are wrong because they are a form of discrimination, or because they are based on arbitrary distinctions, or something else. Still, I could think that E2) is a good reason for M), not a reason *why* it is wrong but a reason *to think* it is.[134] An even clearer case is testimony from a trusted informant that the use of stereotypes is a form of discrimination. The testimony is clearly not a wrong-maker for stereotypes, but it seems it can unproblematically be evidence for M2).[135] To be sure, I am not saying that these will never overlap. After deliberation, I could conclude that the way it wrongs minorities is indeed what makes stereotypes wrong. And so, E2) would work as both evidence and a moral reason against stereotypes. Instead, my point is that there is a distinctive role for evidential relations in moral practices that is not fully explained by just practical moral reasons or wrong-makers.[136]

Now, we have a better understanding of the relation of evidential support that expressivists need to explain, and so are better positioned to develop, clarify, and respond to the two problems expressivists face. My main focus will be on what I call

English, unlike other languages lack syntactic markers for a claim that is supported by evidence —or evidentials as linguists calls them. See Aikhenvald (2004) for more on this.

[134] Some may explain this difference in terms of evidential relations and substantially explanatory relations. See Star & Kearns (2008: 47).

[135] Pure moral testimony is controversial. However, that should not affect the case I am presenting, since most of its problems with it seem to be moral, not epistemic. See Hills (2013) for an overview. C.f. Fletcher (2016) who proposes an alternative explanation of the problems compatible with non-cognitivism, but that is still compatible with its epistemic possibility.

[136] For a similar reason, I do not think that the distinction between evidence and practical reasons tracks different *normative kinds*, like different kinds of reasons. See Schroeder (2012) for a different argument against that view.

the problem of evidential fit. The idea is that expressivists need to explain two things: what it is for information with moral content to be evidence for a moral judgment if moral judgments are desire-like states, and what it is for moral judgments so understood to be supported by evidence. Apart from this problem of evidential fit, there is a different problem for expressivists: a specific strand of their well-known *certitude* problem. The idea is that expressivists cannot explain degrees of certainty for moral judgments. Evidence should be important to understand and respond to some aspects of this problem, given that our level of certainty for the truth of a proposition should depend on the available evidence. Despite this, no critic or defender of expressivism has said much about evidence in discussing certitude. Thus, once I present what I consider is the best response to the problem of evidential fit, I will draw some lessons for how expressivists should understand certitude.

## II. The Challenge of Evidential Fit

Given their commitments on the nature of moral thought and language, expressivists have two things to explain. First, to explain cases where some moral consideration seems to be evidence for another moral consideration, they need to explain how a non-representational entity, like moral considerations, can figure among the evidence for anything. Second, expressivists need to explain how evidence can support moral judgments if these are non-representational desire-like states. It seems to be a category mistake to think of ordinary desires or a non-representational moral consideration as the kind of thing that can offer evidential support. Of course, some desire-like states, like moral judgments, are both truth and epistemically evaluable, and we have some good ideas of how that evaluability works for expressivists. Still, the fact that the evidence and the judgment can be true, even simultaneously, would not be enough to explain how moral evidence and moral judgments can enter relations of evidential support. To explain that, we need something like evidential fit; a non-accidental connection between their truths. Thus, even if the problem has these two aspects, they are both connected by the fact that, given the expressivist picture of moral thought and talk, we do not seem to have a good explanation of evidential fit for moral evidence and moral judgments. Since we are working in a holistic framework in which the evidence works in bundles, these two challenges interact. For example, it is at least logically possible for a judgment to be supported by a set consisting entirely of moral considerations, entirely of non-moral considerations or a mixed set of moral and non-moral considerations.

One natural way for expressivists to think about evidential support is to model it on inferential justification. But the case of mixed moral and non-moral evidence will both show why this does not suffice and sharpen our sense of the dual challenge that expressivists face. One example is our case of evidence E) <shrimps do not have a central nervous system> for the judgment that M) <it is morally permissible to eat shrimp>, and the bridge principle that B) <It is morally permissible to eat creatures without central nervous systems>. The bundle would minimally include B) and E), which are moral and non-moral, respectively. One may think that expressivists have at least one straightforward way to explain cases like this. The judgment may fit a bundle of mixed evidence because it follows our well-established ideas of inferential justification or reasonable inference. This proposal has two parts. First, we need a principle of inferential justification adapted to the case of evidence. Here is one way to articulate it:

> Inference principle: if (a) <if *e* then *p*>, has a creditable standing and (b) <*e*> has a creditable standing, then <*p*> too has a creditable standing.[137]

Evidence then can support a moral judgment by following this pattern. E) meets condition (b) since we can assume it is creditable. B) is a moral judgment, so it is epistemically evaluable. Unless scepticism about moral knowledge is true, B) can be creditable. There are good questions about how B) might get to be creditable. Is it itself supported by evidence, or is it creditable on some other basis? I will return to some of these questions later. For now, however, all we need is the non-sceptical assumption that B) can be creditable. B) can also plausibly be restated as the conditional that <if a creature does not have a central nervous system, then it is permissible to eat it>, so it meets condition (a). If both B) and E) figure among the evidence, it seems like they can support M) by just following this principle.

The Frege-Geach problem may make it difficult for expressivists to make sense of the conditional in (a) and of the inference that figures in the principle. This is a problem for expressivists to explain how moral contents work given the existence of logically complex constructions like negations, conditionals, conjunctions, etc. And the problem is especially thorny when complex constructions include representational propositions, like E) in this case. Still, this would not represent an additional problem specific to the expressivist account of evidence. Assuming expressivists can solve the Frege-Geach problem, they could have an initial way to

[137] This characterization of inferential justification and the structure of the principle is based on, but not completely similar to Fumerton's (1995: 36).

explain cases of evidential support when evidence consists of mixed bundles and could try to generalise from there. Unfortunately, this approach fails, but in an illustrative way.

First, recall that expressivists have two challenges: explain how a moral non-representational consideration can be evidence for anything and how a moral judgment, being a desire-like state, can be supported by evidence. This inferential approach may give us a good way to solve the second problem: moral judgments can be supported by evidence via the evidential correlate of inferential justification. However, we still do not know how considerations with moral content, being non-representational, can be evidence. Moreover, the proposal presupposes we have a solution to that further problem. For this inferential solution to work, the conditional in B) must also be part of the evidence. But B) has moral content, so it is moral evidence if it is evidence at all. Perhaps expressivists could say that parallel to their other response, B) gets to be evidence just by having a creditable standing and being the kind of thing that could figure in an application of the inference principle. However, that leads us to the second problem.

Not all inferences that follow the pattern of the inference principle should count as relations of evidential support. Whether inference is sufficient (or necessary) for this kind of epistemic support is controversial.[138] This is not a problem for most views of inferential justification because they always supplement their views with plausible ways to interpret (a). Consider, for example, how Fumerton (1995: 36) interprets it with the relation that <*e*> makes <*p*> probable, not with a conditional. Moreover, theories of inferential justification are not about the nature of evidential support but about the *structure* needed for that support in most ordinary cases.[139] Thus, (a) does not help explain the non-accidental connection characteristic of evidential fit. Instead, (a) just makes it explicit that inferential evidential support requires such a connection to obtain. But this is precisely what expressivists cannot assume; they need to explain it. They could respond that what is important about this case is that B) is a moral judgment, so the non-accidental connection we seek can be guaranteed via how some moral judgments with good epistemic standing connect the truths of things like E) and M). However, that by itself is unlikely to be helpful. There are many ways in which moral judgments can connect the truth of things like E) and M),

[138] For why it may not be sufficient there are possible cases of failure of transmission of justification. See, for example, Wright (2002a). For why it may not be necessary, see cases of analytic entailment (Smithies, 2021: 552)

[139] It is important that this does not cover all cases, since most proponents of inferential justification are also foundationalists that defend the existence of non-inferential justification.

and not all of them are going to have evidential import. The general lesson here is that it is difficult to get a good response to the two problems without a better sense of how expressivists should understand the non-accidental connection between the truth of <*e*> and <*p*>, characteristic of evidential fit.[140] Thus, in what follows, I will explore some options that directly address that issue.

## III. The Options

The problem of evidential fit is a new problem for expressivists, so we need to spend some time mapping the available and possible options to respond. There are at least two general ways expressivists can proceed: denying there is a problem or offering a positive account. There are many ways in which expressivists can deny a problem, but few are plausible. I will only explore one: denying the problem is specific to expressivism by a companions in guilt argument. As for offering a positive account, expressivists also have two ways to proceed. First, although this is a new problem, expressivists might think they can draw on their views on supervenience or explanation to develop a positive account. As I will argue, these proposals have important limitations. But we can draw valuable lessons about what we need for a good expressivist account of evidence. My proposal is that the right account of evidential support for moral judgments must incorporate elements of first-order moral inquiry. That <*e*> is evidence for <*p*> bears on the question of whether <*p*>, but the latter is a question that we cannot address without moralising. Thus, claiming that '<*e*> is evidence for <*p*>*'* either is or incorporates substantial moral claims. Based on that idea, we can construct a framework that recaptures what is right about the other views.[141]

## IV. Companions in guilt

First, perhaps expressivists can accept that they do not have an explanation of how moral considerations and moral judgments can enter into relations of evidential support given their characteristic non-representational functions. But they can deny that that problem is specific to their view. Many authors in moral epistemology agree

[140] Although related this is different from Dorr's (2002) wishful thinking objection to non-cognitivism in many respects. For example, the challenge here is not about forming beliefs based on desire-like states. Also, whereas he thinks that adjusting your beliefs to your desire-like states is a form of irrationality, here I am just asking how that would work, given a plausible but mostly neutral notion of evidential fit.

[141] An alternative that I will not consider here is that the probability of the moral judgment raises given the evidence. The reason is that this view would involve ideas that expressivists cannot appeal to until they solve the problem of certitude. I will discuss this problem in section VI.

that moral judgments are very different from ordinary beliefs, so it is difficult to explain how they can be supported by empirical evidence —or by non-moral evidence, more generally. But so are many other relations of evidential support for attitudes without moral content. Take Sarah McGrath's (2019: 116-20) recent articulation of this point. As she explains, empirical evidence can support theoretical hypotheses, judgments about how things were, how things will be, or how things may be. Like moral judgments, these all seem very different from beliefs with contents that are straightforwardly *about* how things are.[142] As she explains:

> "Our observations deliver information about how things are, while our moral beliefs concern how things ought to be. In this respect moral beliefs and non-moral observations seem to have distinct subject matters. […] However, it is now generally argued that relations of confirmation can obtain between statements about different subject matters" (2019: 116, 118).

Thus, if our accounts of evidential support are too restrictive to exclude moral judgments, we risk excluding many other uncontroversial examples, like the ones McGrath presents. Hence, expressivists have the basis for a companions in guilt argument: since attitudes that are not straightforwardly about how things are can be supported by evidence, then plausibly, moral considerations as expressivists conceive of them, too.[143]

Perhaps this companions in guilt argument is a good starting place to rethink how we explain relations of evidential support in general. However, it all depends on how we explain the difference between ordinary beliefs, the attitudes McGrath mentions and moral judgments. She does not say much about this beyond marking their different subject matter, but given what we know about how expressivists distinguish the function of desire-like and belief-like attitudes, we can see that this argument is unlikely to help them.[144] Though the attitudes McGrath cites seem different in their subject matter, most of them are still belief-like. We can think of this in terms of direction of fit. Attitudes about how things were, will be, and even theoretical

[142] McGrath talks mostly of claims, statements, and their contents. But I take it that all her points can be unproblematically translated to beliefs and moral judgments since they can presumably have the same contents as the claims she refers to.
[143] See Cowie and Rowland (2020) for more on this form of argument in metaethics.
[144] To be fair McGrath would probably agree that the argument will not help expressivists. But mostly because she shares the common misconception that expressivists deny moral judgments can have epistemic statuses (2017: 79). As for how she marks the distinction she endorses a specific version of confirmation theory according to which these are not analytically or definitionally equivalent (2019: 117).

hypotheses all seem to share the world-to-mind function characteristic of beliefs. Moral judgments are not like this since their primary function is non-representational; they have the desire-like or mind-to-world direction of fit. At least for expressivists, moral judgments cannot be companions in guilt with the other examples; they are different in a deeper but relevant sense. Moreover, given that the desire-like direction of fit of moral judgments was central for the challenge of evidential fit, this response would not be dialectically effective.

In response, expressivists could say that although seemingly belief-like or representational, some of these cases may be closer to moral judgments.[145] For example, there are expressivist or expressivist-like accounts of modal thought about how things may be. Even so, one case is unlikely to help expressivist establish a strong companions in guilt argument. Thus, we need to look for other options.

## V. Covariation, Supervenience and Explanation.

A more promising option for expressivists is based on relations of covariation. When two things, *e* and *p*, stand in a relation of covariation, we generally think of <e> as evidence for <p>. Expressivists then can argue that evidence for moral judgments works similarly. This option is especially promising to account for support for moral judgments from bundles of non-moral or mixed evidence, so I will focus on it initially. Let us use '[p]' for a noun phrase referring to the fact that *p* —assuming the right quasi-realist gloss of moral facts.[146] We start with these two claims:

> 4) the relation between [e] and [p] gives us a non-accidental connection between the truth of <e> and <p>, such that <e> is an indication of the truth of <p>,
>
> 5) <e> is evidence for <p> if and because [e] and [p] stand in a relation such as covariation, supervenience or explanation.[147]

Think of a simple form of causation. Evidence that <e> points to things being such that [e] obtains and given our background assumption that [e] covaries with [p] because one causes the other, evidence <e> is an indication of the truth of <p>. Say that a belief that I still have coffee in my travel mug is supported by the evidence that it makes a splashing sound when I shake it. That there is coffee in my mug

[145] See Yalcin (2011) and Thomason (2020).
[146] See, for example, Blackburn (1998a: 78) or Sinclair (2021: 187).
[147] For presentation I will speak loosely of covariance of facts. However, in supervenience relations, it is classes of properties that stand in covariance relations.

typically causes it to make a splashing sound when I shake it. Against this background, the splashing sound that my mug makes as I shake it is evidence that it still has coffee in it. The two covary: if there wasn't coffee in my mug, it would not make the sound. Given this background assumption about a metaphysical relation it makes sense that the evidence supports the belief. In this way, we can draw a good epistemic principle about indications of truth from a metaphysical relation.[148] This is compatible with the holistic framework we have been assuming since we may need background assumptions to determine that the covariation relation is in place. This view connects with choice point (IV) in that the relation that explains relations of evidence is a form of covariation.

There are some ways for expressivists to mimic this covariation story. Expressivists accept that moral and non-moral facts covary with each other in relations of *supervenience*.[149] Thus, they can appeal to their best views on supervenience to construct their own version of claims 4) and 5). Supervenience, roughly speaking, means there cannot be a difference in normative facts without a difference in non-normative facts.[150] The response to our challenge here could be that evidence can support moral judgments because the non-moral base facts (or a belief about them) would be evidence for a moral fact when and because the latter supervenes on the former. Thus, we have:

> 5S) <e> is evidence for <p> if and because [e] is a member of the supervenience base for [p][151]

Unfortunately, this view has a significant problem: the evidential order need not match the metaphysical order that supervenience puts in place. Very often, the supervenience base for a given property is highly complex, and it is not clear every element is evidentially relevant. (E.g., suppose it is a vast disjunction of its

[148] Classic examples of this strategy are Lewis (1980). See also Rayo (2014). A more sophisticated version of this account can be found in McCain (2013, 2014). He appeals to a form of explanation instead of the simple causal covariation I present here. But my idea is just to introduce the option in the simplest terms and build an expressivist analogue from there.

[149] See for example Blackburn (1984: 186) and Gibbard (2003: 88).

[150] Supervenience is compatible with nihilism. From the supervenience thesis and the premise that the base holds, we cannot infer that the supervening property is had by anything. For the present purposes however, I will just assume that some things are right or wrong —and so there is at least one thing that has a supervening moral property.

[151] There are related existing views. Kieran Setiya (2012: 46) argued that supervenience is the key to explaining all relations of evidential support for moral judgments. And Steven Diggin (2022: 7) suggests this is why even expressivists can make sense of moral evidence. Since they have different goals than mine here what I offer is what I take to be a related option for expressivists to solve the challenge of evidential fit.

instances). Many moral properties have a supervenience base that involves social properties or psychological properties. E.g., the wrongness of racial discrimination in social hierarchical relations or the psychological effects on its victims. But these, in turn, supervene on some yet lower-level properties, like neuro-physical properties. And nothing at these lower levels is likely to have much epistemic import. Consider how this would work for other relations of supervenience. Psychological properties supervene on physical (perhaps too, neuro-physical) properties. Appealing to considerations about the physical properties at the supervenience base will typically not be helpful as evidence for beliefs about psychological properties —and we are more likely to appeal to behavioural considerations as evidence for them. The general point, then, is that even if supervenience can give us covariation and so it can sometimes have evidential import, it just does not follow that all members of a supervenience base will be evidence. Thus, this supervenience proposal *overgeneralises;* it implies the existence of evidential relations when there are none, as with irrelevant members of the supervenience base.

Supervenience is not the only option for expressivists. They could instead opt for a relation of *explanation* between moral and non-moral facts.[152] Consider a *simple* version of this view:

5E) <e> is evidence for <p> if and because [e] explains [p].

This would be a better option since considerations that explain why <p> seem to always carry evidential import, so there is, in principle, no risk of overgeneralisation. Suppose the fact that we ought to donate to charity is explained by the fact that it maximises utility. In that case, the fact that it maximises utility is evidence that we ought to donate to charity. The problem with this option is that there are many cases of considerations <e> that evidentially support <p>, but [e] does not explain [p]. For example, the fact that a nutrition book says that spinach is good for me is evidence that I should eat it, but that does not explain why I should eat it (Broome, 2004).[153] Thus, even if there are some instances in which explanations can count as evidence, too, and that is something that expressivists should explain, this will not solve our

[152] Perhaps as unifying moral generalization (Berker, 2021) or a form of grounding or backing (Berker, 2020).

[153] A different relevant example here could be moral testimony. Hearing testimony that a colleague did something wrong may be evidence that supports the judgment that they did. But it clearly does not explain it. For more on moral testimony and evidence see Lee, Sinclair & Robson (2020).

problem of evidential fit. In contrast with the supervenience proposal, this simple version of explanationism *undergeneralises*: it does not cover all evidential relations.

Not all is lost for explanationism. I believe that a different, more sophisticated version of this explanationist view would have an iterative structure. I propose then, first, that core cases of evidential relations are explained in the way just described: <e> is evidence for <p> if and because [e] explains [p]. Second, secondary cases of evidential relations, like testimony or evidence consisting of non-explanatory considerations, are explained by reference to the base explanatory case. In our case, if the book is reliable, it says spinach is good for me only insofar as that is true, so it is connected with the facts that explain why we should eat it. Thus, the evidential claim ultimately rests on an explanatory claim. We can then modify the simple version:

> 5E*) <e> is evidence for <p> if and because [e] explains [p] or <e> is about a further fact [q] which explains [p]

Consider three objections to this view to develop it further. Some may worry that it seems to go against some of the deeper commitments of expressivists. These explanatory relations are metaphysical relations between moral and non-moral facts. By explaining evidential relations based on them, expressivists would seem to be endorsing a picture of moral inquiry (or at least the parts of moral inquiry involving evidence): that it is about discovering the right metaphysical relations are in place. That picture of moral inquiry is the kind of thing expressivists need to earn the right to, using expressivist-friendly tools, not something that plays a substantial role in their accounts. However, expressivists do not need any robust metaphysical commitments to opt for explanationism. For the view to work, they just need two assumptions. First, claims of the form '[e] explains [m]' (where m is a moral) are themselves a kind of moral claims, whatever else they may be —and so that the right reading of [m] as a fact is not metaphysically robust. For example, a claim that a '[e] explains [m]', just amounts to the moral claim that [e] is a wrong/right maker for [m]. Second, that '[e] explains [m]' need not be read as picking out or presupposing a robustly metaphysical relation.[154] Both are plausible claims that expressivists would want to accept anyway. We can call the combination of these two claims *non-substantialism*. This is a better approach for expressivists who want to adopt

[154] On many views of explanation, explanation is not a purely metaphysical relation even if it is an objective relation. And even if it is a metaphysical relation, there are deflationary accounts of metaphysical explanation. See for example Dasgupta (2017) and Baker (2021) for examples, and Väyrynen (Forthcoming) for an overview.

explanationism. However, it raises worries that the view cannot really account for the possibility of bundles of only non-moral evidence.

Relations of evidential support would always need the right moral considerations in the background to be in place for the explanatory relation to work. Suppose both 5E*) and the non-substantialist approach outlined in the last paragraph. When <p> is moral, the claim that '[e] explains [p]' is a moral claim, whatever else it may be. Thus, for any moral <p>, the claim that <e> is evidence for <p> presupposes a moral claim. Thus, bundles will need to be mixed to include the assumption that the explanatory relation is in place.[155,156] One way for expressivists to address this worry would be to point out that we may have independent reasons to reject the possibility of only non-moral bundles supporting moral judgments. One is to appeal to the view known as the autonomy of ethics —or, more specifically, the epistemic autonomy of ethics. There are many ways to construct this view, but a general and plausible one is that there is no reasonable inference or epistemically permissible inference from purely non-moral evidence to moral conclusions.[157] The typical argument for this claim is that it is an instance of what is known as Hume's law, that we cannot logically derive moral conclusions from non-moral premises but applied to inferential justification. Although explanations do not reduce to inferences, whenever [e] partly explains [p], there are possible inferences from <e> to <p> Hume's law may block the inference if the right moral considerations are not in place. The autonomy of ethics and Hume's law are not uncontroversial, but expressivists may already have some affinities with them anyway. As Nicholas

[155] While there may be some further issues, this is in principle compatible with a triadic interpretation of the evidential relation. For cases where the evidence itself is wholly non-moral, the view would be that when <e> is evidence for <p>, this *presupposes* a moral claim. Under explanationism, it would presuppose that some suitable explanatory relation is in place (either that [e] explains [p] or some iterative case). Even if that claim would strictly speaking not be part of the evidence, evidence would always be evidence relative to that principle and the bundle would be wholly non-moral in an unsubstantial sense —more so given the autonomy of ethics I explain below.

[156] One important question here could be how to think of the holding of the explanatory relation as part of the evidence, if it supposed to be that in virtue of which the evidence is evidence. However, that connects to further general issues about evidence that I cannot discuss here. For example, whether any proposition which is evidence is evidence for itself. See Brown (2015) for discussion.

[157] For more on the autonomy of ethics and the distinct senses of autonomy which include epistemic, see Maguire (2017). For an endorsement and a defence of the epistemic autonomy of ethics see Sturgeon (2002) and Huemer (2005: 72-87). For examples of views on moral evidence that deny Hume's law see Smithies (2022) and Sinhababu (2022). Blackburn (2017: 59) suggests expressivists should reject it too. But from his claim that the movement from non-moral to moral principles is not mediated by logic but by principles that we need to figure out in moral inquiry, we can assume he would also appeal to the existence of bridge principles in the sense outlined here.

Sturgeon explains, one reason for thinking ethics autonomous "would just be the idea, familiar from Hume and especially influential among non-cognitivists, that it is distinctive to ethical thought to be intrinsically action guiding, and that, as Hume remarks, 'an active principle can never be founded on an inactive'" (2002: 195). These may be independent reasons for expressivists to rule out bundles consisting of purely non-moral considerations.

One last worry here is that the view is incomplete, given the specific challenge of evidential fit for expressivists. We know that relations of explanation can be sufficient for relations of evidence given that they guarantee a non-accidental connection between <e> and <p>. We also know that expressivists want to and have offered plausible ways to understand relations of explanation compatible with their framework. However, we are looking for an explanation of how that relation can be of truth *indication*. That element is crucial given that, on the one hand, the challenge of evidential fit was to explain how that indication was possible for a desire-like state. On the other, truth indication seems to be central to how evidence figures in epistemic evaluations like justification or rationality. Thus, I believe sophisticated explanationism can only help us solve the challenge of evidential fit against the background of an account of how expressivists think of these relations in moral inquiry.

## VI. Expressivism, First-order Moral Inquiry, and Evidence

So far, we have been looking at accounts that try to explain claims like "<e> is evidence for <p>" by appealing to relations like supervenience or explanation that could underwrite them. What these views get right is that, on the one hand, these relations will often make the truth of <e> relevant for whether <p> such that one can be evidence for the other —this was our main condition (1). On the other hand, these are relations that expressivists want and have tried to explain within their framework. If we recall, the challenge of evidential fit was to make sense of claims like "<e> is evidence for <p>" in an expressivist framework where what is supported is a moral judgment, and so a desire-like state and the evidence includes considerations with moral content. It makes sense then that this view goes some way into solving the challenge. However, as we saw, all these accounts have limitations. What I believe would help us make sense of and overcome these limitations is that most expressivists would argue that providing evidence is one way in which, from the perspective of first-order moral inquiry, we can respond to the question of whether <p> for a moral <p>. Using our examples, these are questions like whether eating

shrimp is permissible or whether depicting minorities with stereotypes is wrong. These are first-order moral questions, so we need to moralise to respond to them. Accounts that leave that crucial element out will likely be limited by the expressivist's lights.

Consider how expressivists, in general, conceive of moral inquiry. In moral inquiry, we aim to get the correct answers to our moral questions of what to do, think or be. And, as we know, for expressivists, part of what determines correct responses is that these are moral judgments that cannot be further improved or are stable through the relevant improvements of information, maturity, imagination, sensitivity to finer details, etc. In Chapter 3, we called this property of the judgments we aim to get out of moral inquiry a stability property. From our initial gloss of evidence, we know that what we need to account for is evidential fit: a non-accidental connection between the truth of <e> and <p>, such that <e> indicates the truth of <p>. My claim is that our best strategy to offer that account is locating the explanation of claims like "<e> is evidence for <p>" in the broader issue of how expressivists think of the question of whether <p>. Claims like "<e> is evidence for <p>" help us respond to that question: given <e> is appropriately connected with <p> and <e>, one can judge that <p>. But the question of whether <p>, for expressivists, depends on whether it belongs to the best set of attitudes. So possible responses, including those based on claims like '<e> is evidence for <p>', should be in some way indicative of that, indicative of which moral views are epistemically *better*. And epistemically better here is a matter of which judgments have a better chance of being stable through the relevant improvements. Thus, we could understand claims like "<e> is evidence for <p>" as themselves moral claims of the form:

> a) the epistemic endorsement of judging <p> given that <e> and <e> and <p> stand in relation *R*.[158]

To see how this kind of judgment works, we can start with a bad case. Take our initial example of E) < shrimp do not have central nervous systems> and M) <it is permissible to eat shrimp>, but they stand in relation *R*, that the former usually causes people to think the latter. We can even stipulate that E) is creditable and that M) is true. Still, the problem with this way of understanding *R* is that we can expect the resulting moral view to be suboptimal. *R* seems to be the kind of relation that

[158] I am borrowing the term of "epistemic endorsement" from Ridge's (2007) epistemic expressivism. We can understand my interpretation of evidence claims here as a similar form of epistemic expressivism.

draws arbitrary distinctions since we generally cannot rely on what people are usually disposed to think to determine what is morally right or wrong. We should expect the truth of M) to be, at best, accidental and accidental in the relevant sense that we should not expect it to be part of the best set of attitudes —even if it ultimately is. The indication we have, its relation with E), draws arbitrary distinctions that should normally lead us to discard the attitude. *R* is thus constrained by what it tells us about the stability of <p>. But this is also why views like the sophisticated explanationism I proposed in the last section seem so plausible. That view gives us a good way to understand how <e> is relevant to judge whether <p> such that we should expect a judgment that <p> not to be unsettled by the relevant improvements.[159]

According to sophisticated explanationism, <e> is evidence for <p> if and because [e] explains [p] or <e> is about a further fact [q] which explains [p]. To flesh this out, we need a better sense of how expressivists understand relations of explanation. Consider Derek Baker's (2021)[160] view that for expressivists, moral explanations should work by *unifying* moral generalisation. The judgment that M3) lying is wrong, can be explained by the more general consideration that EX) lying treats people as mere means —and possibly other background considerations like that treating people as mere means is wrong.[161] As before, we will treat background considerations as part of a bundle of evidence for M3). In our case, the bundle or set of considerations {EX), p1, p2, p3…} is what really supports M3). We can then understand the relation of evidential support in two parts. First, the fact that EX) explains M3) makes the considerations in the bundle relevant to judge whether M3), because explanations are generally relevant as evidence. However, what makes this a claim of evidence for a moral judgment is that the relation *R* between EX), other considerations in the bundle, and M3) is such that we would find judging M3) permissible in moral inquiry; judging it would make our moral views epistemically *better*. Using Baker's term, it would make our moral views more *unified*: instead of

[159] An appropriate relation with truth is also why explanationists outside of metaethics consider explanations to be evidentially relevant. See for example Belkoniene (2019).

[160] There may be substantial differences with other accounts of explanation, like Berker's (2020) based on grounding. However, going over the different accounts would side track the discussion significantly, and the main point that I want to capture here using the proposed framework is that explanation relations often have evidential import.

[161] As Baker notes this is only a partial explanation, but it should work for my purposes here of illustrating what can count as an explanation *and* evidence. Also, Baker's view is that a good expressivist account of explanation is not about metaphysical relations but about practices of explanation constituted only by attitudes. But we can reconstruct everything I say here without facts given that we are assuming there is a right quasi-realist gloss for them available.

including many different considerations that specific acts are wrong, which may seem arbitrarily put together, bundles with a general piece of non-moral evidence like EX) can present all these considerations as instances of a single unified pattern.[162]

Reconstructing sophisticated explanationism in this first-order framework has a few consequences for our discussion. On the one hand, while not completely off track, views like explanationism worked at the wrong level. Relations of explanations are a great candidate for *R*, but whether they really are the relations that underwrite relations of evidential support depends on how we think of the right conditions to judge that <p> given <e> as a matter of first-order moral inquiry —the conditions that constrain *R* in the way just explained. On the other hand, that means that expressivists can be relatively ecumenical about accounts like supervenience or explanationism. Given that which relation is the correct one is a matter of first-order moral inquiry, expressivists do not have to commit to one interpretation of *R*, their job is just to explain how our thinking and talking of those relations is possible in their framework; similarly to how I just proposed they could do with explanationism. Some interpretations of *R* will be more controversial than others. It is, for example, unclear that expressivists can make sense of interpretations of *R* according to which <e> raises the *probability* of <p> —more on this later. But that option is in general controversial when we think of evidence for moral judgments. And, as the supervenience and explanationist accounts show, expressivists can explain other important candidates.

Most of our discussion has focused on getting the first condition of the initial gloss right: 1) non-accidentality. However, a complete picture of evidence should also include conditions 2) and 3): that <e> is itself creditable and that there are no undercutting defeaters for <e> and <p>. About 2), with a view about the relation between <e> and <p> in place, we can see that most questions about how <e> gets to be creditable should not be problematic for expressivists. For one, <e> could be creditable by being itself supported by further evidence because most plausible candidates for *R* can be iterable —e.g., explanations can themselves be explained.

[162] One question here is whether there is one general property of relations R between <e> and <p> which shows that <p> is part of the best set of attitudes —of which Baker's case of unity is one instance— or rather there are many properties that are not reducible to each other. While I think that this is primarily a first-order question that expressivists do not need to answer, given how they think of the many improvements that can lead you to the best set of attitudes, like coherence, sensitivity to finer details, imagination, etc., they would be inclined to favour the second option.

And even if they are not, we have no reason to think the problem would be specific to an expressivist interpretation of *R* as explanation and not for explanationism, more generally understood. If the question, however, is whether <e> can be creditable by other means or sources, then that is a more general question that we want to answer in this second part of the dissertation. We will explore two further examples in Chapters 5 and 6. But from what we can see, there do not seem to be problems specific to expressivism about 2).

A problem that may be specific to expressivism is that different moral views can fare equally well in terms of improvement, incorporate the same evidence but include different moral judgments. In a non-expressivist framework, we could say that one of these views is likely to be wrong, given that there is truth beyond evidence. Truth could, in principle, break the epistemic tie. However, the same response is not available in the expressivist framework, given that improvement can result in two different sets of attitudes that cannot be further improved. In response, it is unclear that expressivists need to accept that this is a problem specific to their view when it comes to evidence. In non-moral epistemology, whether the same body of evidence can support only one or more doxastic attitudes is controversial. Many permissivist views would side with expressivists in saying that evidence by itself does not have to break the epistemic tie.[163] Thus, even if there is a general disanalogy between realist and anti-realist views regarding truth, the disanalogy may not make a substantial difference when theorising about evidence.

The third condition of our initial gloss of evidence is the absence of undercutting defeaters. More generally, if evidence is central for fixing facts about epistemic evaluations, which evaluations hold will typically depend on defeaters. Expressivists must explain how defeaters fit into this account for a complete picture of evidence. However, to see what is distinctive about undercutting defeaters, it may be helpful to contrast them with rebutting defeaters first. Rebutting defeaters are considerations that can take away or negate the positive epistemic status of the judgment that <p> by supporting a judgment incompatible with it. Following our account, this would amount to, one, <r> initially supporting a judgment <q> given *R*. Second, <q> being compatible with <p>. But third, our moral views being better if we accept <q> based on <r>. More concretely, say that in our initial case about the permissibility of eating shrimp, we add to the bundle the evidence that R) it is not morally permissible to eat creatures with faces. This can work as a rebutting defeater for the evidence that

[163] See for example Kelly (2013). Thanks to Simon Graf for suggesting permissivism.

includes E) shrimps do not have a central nervous system for the judgment M) it is morally permissible to eat them. R) is the kind of thing that, if true, it could improve our moral views by including it. It would make them, for example, sensitive to finer details —e.g., independent conditions for moral status. It is also at least consistent with E) and relevant to settling the question of whether M) eating shrimp is permissible. However, it would show that there are improvements that would lead us to abandon M), the best possible set of attitudes would not include it. In this way, R) is an indication that M) is false in the way we expect rebutting defeaters to be. But the way we explained it is perfectly compatible with expressivism; more specifically with the fact that M) is a moral judgment, and so desire-like, and with R) having moral content.

While a similar story can work for undercutting defeaters, these do not concern the epistemic status of the judgment directly. Still, they give us a reason to think the evidence is not connected to the judgment in the right way. Thus, in our framework, an undercutting defeater <u> would target the relation *R* between <e> and <p> directly and leave open whether the judgment that <p> is likely to belong to the best set of attitudes. In our example, U) some creatures with no central nervous systems can feel pain, could be an undercutting defeater for E) and M). Adding U) to our set would not make a difference on whether it is permissible to judge that M) but would show that E) is not connected in the right way with M): it should not bear on the question of whether M) either. It may still be the case that shrimps do not have central nervous systems, and it is morally impermissible to eat them, given other background considerations about the badness of pain. Notice how this does not depend on a specific interpretation of *R* since any correct interpretation would make <e> relevant to whether <p>.

## VII. Certitude

There is a different, well-known problem for expressivism relevant to the issues I have been discussing: the problem of *certitude*. While I do not intend to discuss the problem in depth, the account of evidence I just presented has important implications for how expressivists should think about certitude. Think of a case in which you ask someone who believes *p*: how confident are you? The appropriate answer seems to express a state that is either a belief with a certain degree or a different mental state, like a credence. This seems to be the same when the question concerns a proposition with moral content. For example, I can be very sure that we should not eat cows but less sure that we should not eat shrimp. Thus, either moral judgments come in

degrees, or a distinctive gradable mental state with moral content comes in degrees, like credences.[164] Since they are committed to the view that moral thinking is desire-like, either way, expressivists should explain this gradability. One way to do this is to argue that, like most ordinary desires, some moral judgments come in degrees; they can vary in their *intensity*. For example, one's desire for a break can vary in intensity depending on how much time they have passed without one. If expressivists need a form of gradability for moral judgments to explain certainty, they can appeal to this notion of intensity. However, moral judgments have a further dimension of gradability. Think of the judgment that <it is morally bad to eat cows>. The normative predicate 'bad' can come in degrees too. One can be *very* sure that it is *very* bad to eat cows. Let us call these two senses of gradability, *certitude* and *importance*, respectively. While desire-like states can accommodate one sense of gradability with their intensity, they lack the structure to accommodate the two.[165]

This problem of certitude has typically been formulated as one about the structure of moral judgments as mental states. However, the gradability of moral judgments has implications for how expressivists can think about evidence. There is a fairly uncontroversial epistemic principle that a rational agent should *proportion* their doxastic attitudes to the amount of evidence they have in favour or against it. How sure I am that <it will rain> should vary according to the evidence I have. If this is right, then in an important sense of the term, certitude depends on evidence: the credence one should adopt depends on the available evidence. Of course, certitude has a different non-epistemic sense of how confident an agent is. And confidence can be irrational and so not responsive to evidence. Some critics and proponents of expressivism are explicit that they are concerned exclusively with the latter sense. My claim here, then, is not that they are wrong. Perhaps there are two senses of certitude that expressivists should worry about —although more on this later. Still, the epistemic sense should be important for expressivists who care about evidence and want to explain the principle of proportionality. Some of the strategies expressivists developed to think about the psychological sense should also be useful for that task.[166]

---

[164] Most proponents of this certitude challenge frame it in terms of degree of belief. But since it is controversial whether they exist, I think we should be explicit that what we designate by this gradable state may be a different mental state.

[165] Smith (2002) is usually cited as the source of this challenge. But as Bykvist & Olson (2009) point out, it appeared first in Bergström (1990: 35).

[166] Although, to be sure, it is controversial whether there is a further psychological and substantial sense of credence beyond the epistemic one. See, for example, Rayo (2014: 34).

Thinking about certitude as depending on evidence should give expressivists reasons to favour some ways of thinking about the problem over others. There are two ways expressivists have explained certitude. One is to accept the critics' terms and think of it as a problem of desires lacking enough structure to accommodate different dimensions of gradability. The strategy, then, is to appeal to two different attitudes to account for the two gradability levels. Andrew Sepielli's (2012) response is a good example. He appeals to a second-order attitude of *being for*.[167] In the case of a moral judgment that <racism is wrong>, expressivists should understand this as the attitude of being for disapproving of racism. How much the agent disapproves explains the importance, and how much they are for disapproving explains the certitude.[168]

Michael Ridge (2020) offers an alternative. According to him, expressivists should not grant that the missing structure needs to be part of the moral judgment as a mental state. There is no general agreement about the metaphysics of credences in the literature. And often, they are thought of as a more holistic property of the agent's cognitive system. More specifically, credences reflect the counterfactual dispositions of the agent, given their estimation of utilities. For example, we can think of this as the agent's betting on the truth of the proposition in different counterfactual situations and their dispositions to act according to their bet. Here is Ridge's gloss on this:

> "How confident am I that lying is wrong, then? This will be fixed by how much of what I care about I am disposed to risk on the truth of that proposition. We determine this by going to the nearest possible worlds in which I believe that the truth of the proposition determines how much utility I will derive from acting in one way rather than another and then "see what I choose" in that world." (2020: 3334)

Thus, attributing a credence to an agent is less about a single attitude and more about patterns of attitudes we can attribute to the agent in different situations. That is what best rationalises their behaviour. To put the view in simpler terms, to say that I have low confidence that M) <eating shrimp is morally permissible> is to say that I would not bet on the truth of that proposition in most relevant situations. Thus, expressivists

---

However, here I will assume with most defenders and critics of expressivists that there is a substantial and important difference.

[167] He takes this from Schroeder's (2008a) response to the Frege-Geach problem. Sepielli's main point is that any response that provides enough structure for moral judgments to respond to the Frege-Geach problem will provide enough structure to respond to the certitude problem.

[168] See Bykvist & Olson (2012) for an argument against this option.

would not need more structure for moral judgments themselves, just a notion of moral propositions and a way to model an agent's response to them in different situations. And we know that expressivists accept that the contents of moral judgments are truth-functional moral propositions, so they should be confident that they do not have problems with certitude.

With these two options in mind, my view is that the holistic picture of evidence just offered favours a strategy like Ridge's. To be sure, Ridge is not concerned with evidence. He is clear that he's concerned exclusively with the psychological and not the epistemic sense of certitude (2018: 3328). Still, I think my account here complements Ridge's view in two ways. First, one of the main options he considers for expressivists is interpretationism, according to which we can determine an agent's mental states based on what their best rationalising interpretation would attribute to them.[169] Although which sense of rationality is the correct for a rationalising explanation is controversial, it is very plausible that it should follow something like the proportionality principle. A good rationalising interpretation interprets the agent as proportioning their beliefs to their evidence.[170] Thus, the two senses of certitude may have more substantial connections than they initially seem. My proposal here then can be understood as a way to fill out that part of interpretationism. Second, even if an account of certitude as a mental state does not need to make sense of the epistemic principle of proportionality, a good expressivist explanation of the principle does need a way to make sense of the gradability of moral judgments. And even if that is a further problem, it is still one that expressivists should solve. And looking at options like Ridge seems to be one of their best available strategies. Either way, what I offer here should be understood as *independent* support for a strategy like his, one that does not locate the missing gradability in the structure of individual moral judgments.

According to the view of evidence claims proposed in the last section, evidence is related to our judgments by a relation *R*. When <e> and <p> stand in relation *R* and <e> is creditable, <e> shows the judgment <p> is likely to be part of the set of attitudes, those that cannot be further improved. And improvement is a gradable notion. A judgment can be *more or less* informed, coherent, imaginative, etc. This should be helpful for expressivists to theorise about the sense of certitude needed to

[169] See Lewis (1974) and Williams (2019).

[170] What is controversial is whether the sense of rationality should be structural or substantive. While the principle of proportionality may seem substantial it is perhaps the less controversial elements of substantial rationality. For more on this idea of substantial rationality see for example Lewis (1974: 336) and Williams (2019: 15-9).

explain how moral judgments can follow the proportionality principle. First, the gradability of improvement seems to be the right place to locate the gradability of certitude. The principle is about how a rational agent should think about the truth of a proposition given the available evidence. In our picture, the stability property is a proxy for truth. Thus, our thinking about how likely it is that M) <eating shrimp is morally permissible> can be explained by expressivists as thinking about how likely it is that M) survives the relevant improvements. Second, this gradability is *not* part of the structure of the attitude. Again, the issue concerns the right response a rational agent should have to the evidence. And it is not obvious that we should think about this response as adopting an attitude with a certain degree. Instead, we could think of whether the agent would drop or keep the attitude in different situations with different evidence. These two consequences of our account of evidence and the principle that connects certitude with evidence are good independent reasons for expressivists to locate the structure they need outside of moral judgments.[171] This would also mean that expressivists are free to use intensity to explain importance. Importance also seems to be a better fit for the intensity of moral judgments. That I judge that *p* is very wrong seems to go hand in hand with how strongly I disapprove of *p*.[172] Combined, these two senses of gradability can secure the double structure we were looking for.

This is not a complete solution to the problem of certitude —not even for the specific epistemic sense of certitude. There are aspects of the problem I am not addressing here. For example, I have said nothing about the issue of motivational maladies. Our desire-like attitudes vary in degree in response to changes in the subject's psychology, like depression. And this should not affect the certitude of our moral judgments, and it would be *irrational* if it did. But it is unclear whether expressivists can say that, given their view that moral judgments are desire-like states.[173] However, my goal here was not to solve the problem but to show how thinking about the relation between evidence and certitude can give expressivists different ways to think of certitude and support a view like Ridge's. In this way, I hope expressivists can

[171] This proposal has important affinities with Lenman (2003b). But they are independent insofar as he is not concerned with the proportionality principle or evidence, and he frames his response as locating the missing structure in the judgment itself.
[172] This intensity of desire-like states is a sense of gradeability critics already accept and that simplifies our discussion here. However, I do not think the desire-like intensity is the best explanation of importance. Instead, thinking back of our discussion of the nature of moral judgments in Chapter 1, the level of first-personal commitment involved in moral judgments seems to be a much better fit.
[173] For a good overview of other problems see MacAskill, Bykvist & Ord (2020: Ch. 7).

find fruitful connections between the different aspects of their moral epistemologies (like evidence and certitude) and move the discussion forward.

## VIII. Conclusion

Although we do not refer to them as such, relations of evidential support are ubiquitous and important for moral practice. Thus, expressivists have very good reasons to offer an account of evidence compatible with their picture of the practical function of moral thought. Like many others, this aspect of expressivist moral epistemology is underexplored in the existing literature. The problem of evidential fit has never been articulated as such, and though the problem of certitude is well-known and discussed, evidence is rarely (if ever) mentioned in relation to it. I hope this chapter gives expressivists a better sense of why they want an account of evidence, the obstacles to offering one, and the most promising options. As for my specific project here, we have a clearer picture of what is involved in making some of our main epistemic evaluations for moral judgments and how that connects with the view of epistemic evaluability offered in Part 1. The next step is to explore the notion of reliability and how expressivists can make sense of it.

# Chapter 5: Reliability

In this chapter, I discuss the concept of reliability and how expressivists can make sense of it as applied to desire-like moral judgments. In this second part of the dissertation, I shifted the attention from epistemic evaluability to proper epistemic evaluations. But as I explained in the last chapter, I am not interested in giving accounts of the usual examples like justification, warrant or knowledge. Instead, I propose to explore how concepts central to these evaluations would work, assuming the expressivist picture of epistemic evaluability offered in chapter 3. In this way, I can give the basis for expressivists to think more systematically about epistemic evaluations and have the tools to make sense of each case. In the last chapter, we covered evidence since, together with reliability, they are some of the main considerations we appeal to when deciding on the epistemic evaluations of beliefs. Thus, if moral judgments are epistemically evaluable in the same way as beliefs, we should expect the same considerations to apply to them.

The structure of the chapter is as follows. I start in section I by explaining the importance of reliability for epistemic evaluations in general and for our project of an expressivist moral epistemology in particular. Section II introduces the reliability and evolutionary debunking challenges (Street, 2011; Dreier, 2012; Golub, 2017). According to these challenges, even if expressivists may need an account of reliability, we have good reasons to think they cannot provide one. Section III concerns existing responses to these challenges and why they are limited. In sections IV and V, I introduce an alternative. This alternative has two parts. First, I argue that expressivists should accept that there is a disanalogy between the means to form true moral judgments and true beliefs, but one that keeps a close enough parallel to think of both instances of a general but plausible account of evaluations of reliability. Second, I argue that this proposal can be supported by empirical research on reliability by Mercier & Sperber (2011, 2017).

## I. Reliability and Expressivist Moral Epistemology

Reliability is central to our epistemic evaluations. Whether a belief was formed by a reliable process is often crucial for how we can evaluate it epistemically. For example, whether my belief that it is raining outside is *justified* depends on the fact I am basing it on my perception of the sound of raindrops. Likewise for unreliability and negative epistemic evaluations. If my belief that it is raining outside was formed by flipping a coin, that should, at the very least, count as a defeater for my justification. Epistemic evaluation of moral judgments works similarly. There are

certain moral judgment formation processes that we trust and others we distrust. We care very much, for example, that our moral judgments are formed without biases, impartially (when the situation calls for it), or with all the relevant information. More specifically, these processes can make our judgments *epistemically* better or worse.[174] Think about a specific case in which *S* is deciding whether one of their friend's lying was all things considered morally good, and the process by which *S*'s judgment was formed was *biased* —perhaps *S* had a personal benefit to gain that is independent of the rightness or wrongness of the lie or was angry for similarly independent reasons.[175] The problem with judgment formation processes that include these things is that *S* is unlikely to produce a true moral judgment. There is an important parallel between processes of belief formation based on perception and processes of forming moral judgments without biases. These processes are reliable when they give us a reasonably systematic and non-accidental connection with truth and unreliable when they do not. To see this more clearly, it will be helpful to establish some stipulations to nail down a rough characterization of reliability —as we did with evidence. This characterization will help us see the parallels and work through the different arguments and examples, but we will refine it throughout the discussion.

> A process *P* is epistemically reliable in circumstances C when:
>
> (1) *P* can systematically produce true beliefs/moral judgments.
>
> (2) The connection between the process *P* and the truth of individual judgments that *p*, in (1), is not accidental.
>
> (3) There are no undercutting defeaters against the connection between *P* and the truth of *p*.

Evaluations based on reliability are primarily about belief and moral judgment formation *processes*. These evaluations have a specific structure, and these conditions are meant to capture central aspects of that structure. Take Jason Baehr's

[174] We could also evaluate these methods morally. For example, if our there is a lot at stake in the outcome of our deliberation, we are morally responsible for not using unreliable methods. But this is compatible with their role in our distinctively epistemic evaluations.

[175] This will be my main example throughout the chapter so it is worth noting that I will be using a very general notion of bias as dispositions to form judgments based on factors *irrelevant* to the correctness of the judgment —like truth for beliefs or stability for moral judgments. See Kelly (2019) for more on this view. This will also help us separate cases in which things like anger do not count as biases since they are relevant for the correctness of the judgment —think of accounts of blameworthiness in terms of fitting anger. This is also not different from most evaluations of reliability like those based on perception since even those can be unreliable in unfavourable conditions.

(2019: 97) claim that reliability is not *primarily* an epistemic concept. We evaluate processes as reliable if they are apt for the systematic achievement of certain ends. What makes evaluations of reliability epistemic is the ends, for example, true beliefs in matters that are important to us. With that end fixed, we can establish conditions like (1) to (3) for good or bad ways in which our judgment-producing processes achieve (or help us achieve) the aim.

Condition (1) is meant to capture the intuitive idea that methods that give us good results most of the time, or at least have a good ratio or frequency for producing true beliefs, merit an evaluation based on reliability —and methods that do not, merit an evaluation based on unreliability. Think about belief formation processes based on perception under normal conditions. Part of the reliability of that process is that most of the time our beliefs were formed in this way we ended up with true ones. We will call this *systematicity*.[176]

Although systematicity captures an intuitive and important element of reliability, it can only give us a partial characterization. Even at a point where one never used their perceptual capacities, the process of forming beliefs based on them should be reliable. Reliability is a sort of dispositional property: even if a process never manifests it, it would show up if the right conditions were met. Reliabilists offer some good examples of what this further condition could amount to. An important aspect of the reliability of a belief-forming process is that it is based on the aetiology of beliefs and how in the right processes, like those based on perception, it forms a path from the objects they are about to their formation. This path is the one we should expect the process to be disposed to repeat in a more or less systematic way, given the right circumstances. This dispositional property explains why there is a non-accidental connection between the process and the truth of the belief, as (2) indicates. We will call this non-accidentality.[177]

---

[176] Something similar to this condition is sometimes labelled a statistical frequency conception of reliability and used to explain some problematic cases for broader reliabilists views. See Lyons (2013). See also Sinhababu (2021) for an example of how to apply a similar idea to moral judgments based on non-moral facts. Here I just take this as an important element of a characterization of reliability.

[177] For more on why it is important that reliability is a dispositional property see Alston (1995), Graham (2017: 43), and Greco and Reibsamen (2018: 726). Notice that non-accidentality is more fundamental but does not obviate the need for systematicity. A process that produces a true belief once by ways which we can explain, but that we cannot rely systematically may merit a good epistemic evaluation but probably not one based on reliability. For a similar reason something like systematicity is often cited as necessary to individuate reliable processes —and solve what is called the generality problem. See Conee and Feldman (1998).

Even the most reliable processes that meet conditions (1) and (2) cannot merit a positive epistemic evaluation in unfavourable conditions. We covered some similar cases in the last chapter. A belief that there is a red paper in front of me based on the perception that there is one will not be reliable, considering there is a red light over it. How to think about the connection between these considerations and reliability is a contested issue. However, one useful way to think of them is as defeaters that undercut the truth of the judgment from the process producing it.[178] Thus, we have condition (3).

Similarly to evidence, reliability should be important for an expressivist moral epistemology because, as the example illustrates, the concept is important for our ordinary moral practices of epistemically evaluating moral beliefs. These are the practices quasi-realist versions of expressivism want to explain and vindicate. More specifically, reliability, like evidence, is representative and central enough that, by securing it, expressivists can build a case that they can account for the variety of our epistemic *evaluations*, not just epistemic *evaluability*. Unlike evidence, expressivists have independent reasons to want an account of reliability. There are existing and influential challenges that expressivists cannot adequately make sense of it: reliability and evolutionary debunking challenges. I will explain these challenges in more depth in the next section. For now, the main point is that expressivists do not want to say —at least in the first instance or in a substantial sense— that our reliability consists of systematic tracking of moral facts. But if that is not their view, it is unclear what is. Moreover, our best evolutionary accounts of our moral judgment formation processes seem to conflict with some of their main quasi-realist commitments about reliability and truth, restricting their options even more. An account of reliability that addresses these issues would then be an important addition to an expressivist moral epistemology and a natural next step for our project here.

Some expressivists like Blackburn (1996, 2001) have suggested some affinities with *reliabilism*[179] as a broader view in epistemology, so it may be important to note how my goals in this chapter do and do not overlap with that other project. Reliabilism, in this sense, is the ambitious project of explaining or reducing our main epistemic concepts, like justification or knowledge, to non-epistemic notions. Reliability is especially promising in this respect because, to make sense of it, we just need the

[178] Goldman (2009: 324; 2012: 82) is a good example of how some reliabilists integrate a no-defeaters condition in their characterizations of reliability.
[179] See Blackburn (1996, 2001) and Gibbard (2003: 226-62) for some suggestions in this respect. Although see Jenkins (2015: 70) for some reservations expressivists should have about this approach.

relevant end of having true beliefs and the mental processes that produce them. Thus, we could explain things like justification in terms of reliability when the latter can be explained using these non-epistemic concepts. Apart from pure reliabilism, the main examples here are reliabilist virtue epistemologies.[180] These projects are not meant to account for an aspect of our epistemic practices, as expressivists usually do, but to establish a completely different approach to systematic epistemological theorizing. There are some good reasons why expressivists may want to endorse one of these more ambitious projects, but that is not the approach I will be adopting in this chapter. My goal is more restricted: explain how evaluations based on reliability are possible in an expressivist framework.[181] Nevertheless, it will be useful at many points of the discussion to use some reliabilist tools to flesh out the concept of reliability that expressivists want to explain and vindicate.

## II. Reliability and Evolutionary Debunking Challenges

Reliability is perhaps the aspect of epistemic evaluations that expressivists have given more attention to;[182] so have their critics. Reliability and evolutionary challenges have received considerable attention in the literature in recent years.[183] These are generally regarded as independent challenges; however, given my goal of explaining reliability as it figures in our epistemic evaluations of moral judgments, it makes sense to address them together —I will come back to this. First, I will explain which aspects of expressivism these challenges target, then introduce their main structure and explain how both reliability and debunking challenges fit into it.

### II.1 The Target of the Challenges

Sharon Street (2011), who first introduced the challenge for expressivists, argued that by trying to mimic everything realists can say about things like truth and objectivity, quasi-realist expressivism inherits its epistemological problems. But to be more precise, the problem arises not from different explanations of similar

[180] Classic examples of these can be found in Goldman (1979) and Sosa (2009). For an overview see Goldman & Beddor (2021) and Greco & Reibsamen (2018).
[181] One reason for this is that the more ambitious project assumes a successful undertaking of the less ambitious one. If expressivists cannot make sense of reliability, they clearly cannot endorse a reliabilist epistemology.
[182] Though they expanded on their accounts in response to these challenges, both Blackburn (1993: Ch. 2; 1998a: 318-19) and Gibbard (2002: 226; 2003: 224) present reliability as the first issue to address for an expressivist moral epistemology.
[183] See Sinclair and Chamberlain (Forthcoming) for an overview of both and their relations. For more on evolutionary debunking challenges for quasi-realism see Street (2011), Gibbard (2011), Kyriacou (2017), Korman (2019: 5, 17) and Blackburn (Ms). For reliability challenges see Dreier (2012) and Golub (2017).

*interpretations* of moral practices.[184] Both expressivists and moral realists accept that among the fundamental assumptions of moral practice are:

> **Truth**: moral judgments and claims can be true or false.
>
> **Objectivity**: the truth of our moral judgments does not counterfactually depend on us.
>
> **Reliability**: some of our judgment-forming processes are reliable, and others are unreliable.

While expressivists and moral realists agree that these are fundamental assumptions of moral practice, they disagree in their *explanations*. Roughly speaking, moral realists typically appeal to moral facts to do substantial explanatory work in their accounts of truth, objectivity and reliability.[185] Moral judgments are true or false when they represent moral facts accurately. Moral facts stay the same regardless of what we think of them; in that sense, they are objective. And we are reliable in producing true moral judgments because we are reliable in tracking moral facts. Expressivists have different explanations for both truth and objectivity. Moral judgments do not represent moral facts because they are desire-like states with a primarily practical (not representational) function. But they can be true or false in a deflationist or minimalist sense. As for objectivity, expressivists think of it as a substantial first-order issue, so they only need to explain our thinking and talking of moral facts as objective. For example, consider again *S*'s judgment that it was all things considered good for their friend to lie. This judgment can be explained as accepting the conditional that if *S* did not approve of their friend's lying, it would still be good. That conditional is a substantive first-order moral claim. While expressivists are not in the business of endorsing it as metaethicists, they can make sense of it as an element of our practices that is still compatible with their picture of moral thought and language. These explanations of truth and objectivity are not uncontroversial, but it is at least clear how expressivists have available options. What we are still missing is an explanation of the third element: reliability.

---

[184] Sinclair & Chamberlain (Forthcoming: 14) make a similar point about this distinction.

[185] Some may argue that quietist versions of realism, even those that address reliability worries, do not do this. See for example, Dworkin (1996) and Scanlon (2014). Although I think that there is a difference between moral facts doing explanatory work and having a metaphysically quietist interpretation of them, the distinction will not matter for us since I am using realism merely as an illustration.

## II.2 The Structure of the Challenges

Reliability and evolutionary debunking arguments give us two different reasons to think expressivists cannot explain reliability. While they differ in some aspects, we will understand them as sharing a similar structure for our purposes. Following Daniel Korman (2019: 4), we can understand these arguments as based on two main premises:

> E) Assuming quasi-realist expressivism is true[186], all our moral judgment-producing processes *P* are not explanatorily connected with moral truths.[187]
>
> R) The fact that a moral judgment that *p* and moral truths are not connected in the right way by process *P*, is a *defeater* for the reliability of the judgment that undercuts *P* from the truth of *p*
>
> And so
>
> C) Given that all our moral judgments are produced by a process *P* quasi-realist expressivists are committed to a general defeater against the reliability of all our moral judgments.[188],[189]

R) is a very plausible epistemic principle that the lack of an explanatory connection between the truth that *p* and a judgment that *p* is a defeater for the reliability of the judgment. The principle is general, so it does not owe its plausibility to its use in debunking arguments or the specific case of moral judgments. If the explanation of why I believe it is raining outside has nothing to do with the truth that it is raining outside, maybe because I flipped a coin to decide that it was, that should be a defeater for the reliability of the process that produced the belief. More specifically, relations of explanation between truth and judgment-forming processes are paradigmatic non-

[186] As suggested here my focus will be on conditional debunking arguments that aim to prove that by assuming the truth of a theory, we can prove that moral judgments have a negative epistemic status. See Korman (2019: 3). Only conditional arguments are relevant for expressivism. Conditional arguments should not be confused with local arguments that only prove that a subset of our moral judgments have a negative epistemic status or are false. See Cosker-Rowland (2019b).

[187] The explanatory connection here is related but not the same as the one presented in chapter 4 to explain evidence. The main difference is that that the explanatory relation then was between facts, whereas this is one is about moral truths and judgment formation processes.

[188] My formulation here is largely based on Korman (2019: 4) and Korman & Locke (2020: 327). There are many ways to formulate debunking and reliability challenges, but not all of them are relevant for expressivists or for our focus here on epistemic evaluations.

[189] Many debunking arguments aim to conclude that evolution plus the theory in question (expressivism or realism) entail moral scepticism. There are ways to frame the conclusion here like that, too. The lack of explanatory connection via expressivism would mean there is a defeater for *each one* of our moral judgments, which would entail moral scepticism. I will, however, stick with the more restricted version to keep the discussion focused on reliability.

accidental relations of the kind we need for reliability —think of our condition (2).[190] Moreover, as it was part of our initial conditions (3), there should be no undercutting defeaters for the process *P* and the truth of the belief that *p*. Lack of an explanatory connection is such a defeater.[191] R) would be no problem for expressivists if it was just about instances of moral judgment formation processes. What is problematic for expressivists is that according to E), there will be a defeater for *each* moral judgment given their specific commitments. Which of these commitments entail there is no explanatory connection varies according to whether we are dealing with reliability or debunking challenges.

E) is an explanatory premise about what expressivists can or rather cannot say about the connection between the truth of our moral judgments and the processes that produced them. E) is often formulated in terms of explanatory connections between moral judgments and *facts*, but my formulation is about moral *truths*. The main reason for this is that many expressivists are committed to not using facts in their explanations of reliability, and we do not want the challenge to beg the question against them. I also make explicit that the challenge is about the *processes* for forming true moral judgments and not just about why *each* of our actual moral beliefs is true. As I noted evaluations of reliability are primarily about those processes, and we are interested in how these challenges apply to them.[192] Reliability and debunking arguments provide two ways to support or articulate this premise.

The main point of *reliability challenges* is to ask for an explanation of why we are good at producing true moral judgments. And then, point to the nature of moral truths and moral judgments, given expressivism, to deny there is, in principle, a way to make sense of the explanatory connection between them. Given the nature of moral thinking and truth, it is easy to see why there are problems on both sides. Moral judgments are desire-like; thus, they do not represent or track moral truths, so no moral judgment formation process can establish that connection. And since expressivists accept the existence of moral truths only in a minimalist or deflationist sense, it is hard to see how they can be explanatorily connected with moral judgments. But there must be a connection since expressivists are committed to our

[190] For a convincing defence of this and similar ideas see Jenkins (2006) and Faraci (2019).
[191] While I take the principle to be relatively uncontroversial what *is* controversial is whether lack of an explanatory connection defeats reliability directly or by showing that other modal relations between truths and judgments, like safety or sensitivity, are not in place. See for example, Faraci (2019) and Korman & Locke (2020) for convincing arguments against the latter.
[192] For the distinction see Golub (2017: 797) and Sinclair and Chamberlain (Forthcoming: 12)

moral judgments being true or false and to the fact that their truth does not counterfactually depend on us. Of course, expressivists can deny that this is the right explanation of moral thought and talk, given their views. But if that is not the correct account, it is unclear what is.[193] To make this approach to premise E) more explicit, we can restate it as:

> ER) Given the nature of moral thought and language in an expressivist framework, moral truths and moral judgments are not connected by any process *P*.

Instead of just denying there is a connection, *evolutionary debunking challenges* provide an alternative but very plausible explanation of our moral judgments that makes no use of moral truths. More specifically, this is an account of our capacities to form moral judgments and our tendencies to pick certain moral judgments over others. Our capacity to make judgments about what is morally right or wrong was selected because it was enormously advantageous for our ancestors. It made cooperation and life in society possible. Not only that, but a significant amount of our ordinary moral judgments can be explained by this advantage. Tendencies to value things like honesty and generosity seem to be precisely the kind of things that would best serve the purpose of social cooperation and thus would help the members of our species survive. Those explanations are uncontroversial, so no metaethicist, realist, expressivist or otherwise would want to reject them.[194] But then the best explanation of our moral judgment-producing processes *P,* the evolutionary explanation, does not include moral truths or any way these processes could track it; there is no explanatory connection. And, again, according to expressivists, there *are* moral truths, and they do not counterfactually depend on us. Thus, evolution and expressivism together entail that there are moral truths, but our processes were not

[193] Some could argue that the reason why reliability challenges are independent from debunking challenges is that the former does not need any claims about the negative epistemic standing of moral judgments given the theory. They could be formulated as a plain a request for an explanation. See Sinclair and Chamberlain (Forthcoming: 2). While I agree that this is possible, I think the most interesting reliability challenges do incorporate this element. Consider why there is no substantial reliability *challenge* about beliefs based on perception.

[194] What is controversial perhaps, is whether evolutionary accounts of our judgment formation processes and tendencies would provide a complete account of individual moral judgments. As Sinclair (2017: 102) and Sinclair and Chamberlain (Forthcoming: 2) explain we may need other psychological, historical explanations to supplement this evolutionary account. However, as debunkers point out the other elements of our explanation are unlikely to include truth either. Thus, even if we had a complete explanation, or as they call it, an evolutionary+ explanation, that is unlikely to help expressivism. When I talk about evolutionary accounts I will assume these are evolutionary+ accounts but will not make it explicit to simplify the discussion.

selected for tracking them. As R) states, this lack of explanatory connection is a defeater for the reliability of those processes.[195] And so, again, expressivists cannot block the problematic conclusion. To be more explicit, we can again restate premise E) as:

> ED) Assuming both quasi-realist expressivism and evolution, there are moral truths that do not counterfactually depend on us, but our moral judgment formation capacities were not selected to track them. Thus, moral truths and our moral judgments are not explanatory connected by a process *P*.

A different way to think of debunking arguments is that they restrict what expressivists can say about the explanatory connection between moral truths and our processes of moral judgment formation in response to reliability challenges. Their accounts must be compatible or at least as plausible as evolutionary ones. As Street (2011: 5) explains, the constraints from debunking challenges to expressivist accounts of reliability are especially pressing since expressivists are naturalists. One of the main attractions of an expressivist picture of moral thought and language is that it sits better with our best scientific pictures of the world, which is likely to include this evolutionary account of our capacities and tendencies to form moral judgments.

Some may find the conflation of reliability and debunking challenges problematic[196]. However, my goal here is not to offer a complete study and response to the challenges but to establish a way to understand evaluations of reliability given an expressivist picture of moral thought and language. This unified framework is useful to see which elements of the challenge need to be addressed for that purpose. Think of some of the reasons we may find against the conflation. One is that reliability challenges just ask why we are good at producing true moral judgments, while debunking challenges deny we can do that if not by an extraordinary coincidence. Offering an account that responds to debunking challenges by denying there is a coincidence will not be useful for our purposes here if that account does not also explain why we are good at producing true moral judgments. A different reason to think that the challenges should be dealt with separately is that reliability challenges assume the truth of some

[195] One way to think about the difference between reliability and debunking challenges I am trying to mark here is in terms of what Korman (2019: 4) calls negative and positive approaches to premises like E). He suggests his characterization, like mine here, aims to give a unified account of reliability and debunking challenges.
[196] See for example Sinclair and Chamberlain (Forthcoming: 2)

of our moral judgments, but debunking challenges would deny we can assume that. However, that view is very controversial.[197] Moreover, the truth of some of our moral judgments is not something that expressivists want to prove but an element of their interpretation of our moral practices that they want to explain and vindicate. Expressivists then are likely to respond that assuming that some of our moral judgments are true is just part of what moralizing involves, and so it is not their job to convince these debunkers otherwise.[198]

To sum up, reliability and debunking challenges teach us different things that we need from an expressivist account of reliability. A good response should explain why there is an explanatory connection between moral judgments producing processes and moral truths given the expressivist picture of moral thought and language —to block ER) — and the account must be compatible with or as plausible as the evolutionary story of those processes —to block ED).

## III. Existing Options and Their Limitations

Expressivists have offered a few responses to these challenges. The main line of argument has been that given the expressivist picture of moral thought, language, truth and objectivity, we should expect a disanalogy between what expressivists can say about reliability and what we can say about reliability in other domains —like forming beliefs based on perception. But this disanalogy may not be problematic. Although I agree that this disanalogy is key to solving the challenges, the way existing responses have been constructed makes the accounts problematically limited. Still, because of what they get right about the disanalogy, it will be useful to cover two of the most promising ones: Jamie Dreier's (2012) and Camil Golub's (2017).[199]

Dreier's strategy is to show that given how expressivists explain truth and objectivity, there was no way for the problem of reliability to even arise in their framework. More specifically, moral judgments are true or false only in a deflationist

[197] For convincing arguments against it see Sinclair (2017) and Korman and Locke (2020: 314-6)

[198] See for example Blackburn (1993: 149). C.f. McPherson (2022).

[199] Other options include Blackburn's (ms) and Gibbard's (2011, 2015). Expressivists may even appeal to some realist responses to debunking arguments based primarily on first-order judgments of reliability, like Enoch's (2009). Options that are relevant to my discussion here but will not address directly are Sinclair and Chamberlain's (Forthcoming) and Jenkins (2015). Sinclair and Chamberlain's solution is based on our reliable methods to track non-moral facts our moral judgments are based on. While I think there is much to be said in favour of that view, it seems to fit better with the account of evidence offered in Chapter 4. And Jenkins' was just meant as a suggestion for expressivists and not a full account. Still, as it will be clear, many of her points will inform my positive proposal in the next section.

sense. This means that truth is not a substantial property but a device for generalization of instances of schemas like *p* is true iff *p*.[200] Thus, moral truth is not something for our moral judgments to match or mismatch or something that we can be good in tracking. And perhaps more centrally, since truth is not a substantial property, it cannot figure in relations of explanation, so it cannot be connected in that way with our moral judgment formation processes. Dreier proceeds by offering a detailed account of the emergence of truth and objectivity in our practices. These concepts emerged as a way for agents to express their non-cognitive attitudes of approval and disapproval in the sophisticated ways that the practices demand. He proceeds in stages, introducing evaluative predicates, truth predicates, and claims of objectivity successively. He claims that at no point a worry about reliability could arise because the materials are the same as the story proceeds: non-cognitive attitudes and the language needed to express them. Think again about ER):

> ER) Given the nature of moral thought and language in an expressivist framework, moral truths and moral judgments are not connected by any process *P*.

Lack of an explanatory connection would normally be a defeater because we would have a reason to think our judgments were not produced to match moral truth. Dreier's response is not that there *is* such a connection but that the question of whether there is one does not arise for expressivists, given how they understand truth and objectivity. These are just ways to think and talk about our moral judgments, so there is nothing strictly speaking for them to match or mismatch in the relevant explanatory way. In our initial case, *S* judged that it was all things considered good for their friend to lie. Whatever the process that produced it, to say that it is true, following deflationism, is nothing other and above saying that it was all things considered good for their friend to lie. And so, there was no way for expressivists to be committed to this explanatory thesis or its negation. Even thinking of that truth as objective does not open this gap since that just amounts to thinking that even if *S* did not approve of their friend's lying, it would still be wrong. Thus, the argument does not follow.

[200] Dreier writes of both deflationism and minimalism in the paper. However, as his argument goes, he seems to be focusing exclusively on deflationism, at least in the way we have been distinguishing these views. An interesting question is whether the argument would work under a minimalist framework in which truth is a property that we perhaps can track. I will not pursue that argument here, although the response I offer in the next section assumes stability as a proxy for truth, which is an important element of minimalisms like Wright's (1992).

As Dreier himself admits, the response is limited. We know there is no reliability problem for expressivists since there was no way it may arise in their framework. Still, there must be something for them to explain, given that they accept reliability. Or, as formulated here, the reliability argument does not follow since we do not have premise ER), but there seems to be something for expressivists left to explain about reliability. The problem that I think is still open is that the lack of an explanatory connection still seems like a strong indication that the process in question is unreliable. But if explanatory considerations seem irrelevant for the reliability of moral judgments in an expressivist framework, then either what we are talking about is not reliability, or we should reject R) that lack of this explanatory connection is a defeater for reliability. Both options seem problematic for expressivists. Following the expressivist interpretation of moral practice, we do think and talk about our moral judgment formation processes as reliable, and that reliability seems distinctively epistemic: it depends on whether the processes produce true moral judgments. Thus, by denying that we are talking about reliability properly speaking, the kind that is defeated by lack of explanatory connection, expressivists seem to go against their quasi-realist commitments of vindicating moral practice. Expressivists could deny the epistemic principle R). But as I explained, we have good independent reasons to accept it. Thus, that seems like a difficult task beyond Dreier's negative argument. Either way, there is clearly something left for expressivists to explain.

Golub's (2017: 806) solution has two parts: a negative thesis about the nature of desire-like moral judgments and a positive thesis about what it is to evaluate something as reliable from the first-order perspective of moralizing. He argues that since expressivists believe that moral thought does not represent moral facts, we should not expect an expressivist account to mimic the best pictures of their reliability of other representational states like beliefs. However, expressivists can explain the connection between moral judgments and truths by appealing to the kind of processes we would normally endorse as reliable while moralizing. We are reliable, for example, when we form our moral judgments carefully, considering the relevant facts, or without biases. This would not be part of the expressivist picture of the nature of moral thought and talk, but rather just pointing out the internal commitments of moralizing that expressivists would not want to rule out. Still, these responses are genuinely unifying and illuminating, given that we are pointing to processes that we regard as securing connections with truth, not just claims about the truth of our individual moral judgments. Golub concludes that if expressivists try to say more, their only option is to try to mimic a problematic 'tracking epistemological

model' (2017: 808) of the reliability of beliefs incompatible with their other commitments. But it is not clear that they need to, given the disanalogy.

As I understand Golub, the idea then is that expressivists can deny ER), that given the nature of moral thought and language in an expressivist framework, moral truths and moral judgments are not connected by any process *P*. From the perspective of moral practice, we can support certain moral judgment formation processes as those that can get us closer to the truth, and so there is a sort of explanatory relation between the process and the truth. In our case, *S*'s judgment that it was all things considered good for their friend to lie was produced without biases. Expressivists can point out that the process of forming judgments without biases are among those we are likely to endorse when moralizing. Two aspects of this approach are on the right track. Expressivists should expect an important disanalogy between our accounts of reliability for moral judgments and beliefs because of the non-representational nature of the former. Second, the fact that we are reliable in producing moral judgments is not part of the expressivist picture of moral thought and talk; it is a commitment internal to the dynamics of moral practice that they want to explain and vindicate. It makes sense, then, that whatever illuminating account expressivists can give about reliability appeals to some of our first-order commitments.

However, recall that a complete account of the reliability needs to meet conditions (1) systematicity and (2) non-accidentality. And it is not entirely clear how these fit into Golub's account. Expressivists could say that a story like Golub's just explains our historical reliability and so meets condition (1). And that seems right; the idea seems to be that we know that some of our moral judgments are true and know which methods produced them. This is illuminating and unifying. Thus, we can explain our reliability by citing those methods that historically have produced moral judgments we know are true. But we also need (2), non-accidentality, and it is unclear how to explain it with the present tools.[201] Given that historically, these processes have produced true moral judgments, we may infer that there is a non-accidental connection between them or argue that when we endorse these processes in moral

[201] The strategy used in chapter 4 to say that explanation claims are themselves normative is not available here, since we are talking about truths explaining why processes produced judgments, not explanations between facts. The strategy back then was to reinterpret claims like 'capacity to feel pain explains the wrongness of harming non-human animals' like 'capacity to feel pain is a wrong maker for the wrongness of harming non-human animals.' See Dasgupta (2017). In the case here the judgment formation process is not a right-maker for the lie; that would be the wrong explanatory connection.

practice, it is because they produce moral judgments with certain properties that connect them with the truth. However, it is at present very unclear what that connection *is* or what makes it non-accidental. Expressivists could say that this is the right place to locate the disanalogy between moral judgments and beliefs, and so we need different conditions for the reliability of beliefs that presumably do not include (2), only (1): systematicity. But without something like (2), it seems very difficult to deny premise ER), that moral truths and moral judgments are not connected by any process *P*, so the claim that there is a general defeater for our moral judgments seems pertinent. In that case, this picture of reliability is not just disanalogous but limited.

Golub (2017: 807) does point out that if expressivists needed a further condition for the reliability of our beliefs, they *could* rely on modal conditions like sensitivity or safety. These can give us a sense of non-accidentality by testing the belief in nearby possible worlds. For sensitivity, *S* should not believe *p* in nearby worlds where *p* is false, and for safety in all nearby possible worlds where *S* believes that *p*, *p* is true.[202] However, as Golub rightly pointed out, it is very controversial that safety and sensitivity are conditions for the reliability of moral judgments. Since fundamental moral truths are necessary, safety is unlikely to be informative. And many proponents of reliability challenges propose counterexamples that show sensitivity is a bad guide for reliability. Think, for example, of cases where *S* believes that *p* and, on that basis, comes to believe that they are not mistaken that *p*, but acknowledging that they would still believe it even if it were false does not seem to *defeat* this further belief (Korman and Locke, 2020: 317-23).[203] And defeat, as we know, *is* a good guide to reliability. In any case, given that Golub does not think of this as a good option for expressivists, I will not get further into the details. What matters for us is that we need a different, not accidental (but also not modal) connection between moral truths and judgments.

## IV. Reliability and Improvement-Means

Expressivists do not have to settle for a limited account of reliability. The best approach for them is to adopt what Sinclair calls an ancestral strategy:

[202] These are just meant as rough outlines. The right characterization of sensitivity and safety is likely more complex. See Clarke-Doane (2020: Ch. 5) for more on this.
[203] They also offer examples of safe beliefs that are unreliable: cases in which *S*'s belief could not easily have been wrong but are still defeated by concession that there is no-explanatory connection in place, and so do not seem reliable. See also Vogel (2012) and Faraci (2019) for other defences of this idea. C.f. Clarke-Doane (2020: Ch. 5), although his view is framed explicitly for realists, not for expressivists.

> "According to this strategy a given feature of a discourse is domesticated insofar as it is shown to have more than one version, one of which applies to the case of moral discourse understood expressively, and where all versions can be understood as particular instances of a perfectly general 'ancestral' account." (2021: 220)

My proposal is that expressivists use this ancestral strategy to solve their problem with reliability. As Dreier and Golub rightly point out, there is a disanalogy between our best tracking accounts of the reliability of beliefs and what expressivists can (or want to) say about moral judgments. The disanalogy is to be expected, given expressivists' views on the desire-like and non-primarily representational nature of moral judgments. But this does not mean the expressivist account in question is ad-hoc or limited. Instead, expressivists should argue that there is a perfectly general and plausible ancestral account of which the expressivist account of moral judgments and our best accounts of beliefs are instances. Consider Baehr's view of evaluations of reliability as having a means-ends structure. That account is both plausible and general. A tracking account of reliability for beliefs fits with it, but they are not equivalent. Thus, as long as expressivists offer an account that fits the general structure, they do not have to either settle for a limited account or try to mimic a problematic tracking account for moral judgments.

Consider the end of true moral judgments first. One reason expressivists may not be able to earn the right to that part of the ancestral account is that truth in a deflationist framework like Dreier's is not a substantial property, just a way to think and talk about our moral judgments. In that sense, truth is unlikely to be the end in evaluations based on reliability. However, it is unclear whether expressivists need to settle for that account. A different option here is the picture of truth we have been working with as the property of moral judgments that cannot be further improved or are stable through improvements of coherence, imagination, maturity, sensitivity to finer details, etc. This view is closer to minimalism than deflationism in allowing truth to be a substantial property, one that is designated by all our platitudes surrounding the concept —stability being the primary one. Stability thus works as a proxy for truth. In this sense, truth is a more substantial property capable of figuring in explanation relations. As many expressivists claim, stability has important connections with how they conceive of moral inquiry, which will be important here too. In moral inquiry, we aim to get correct responses to our practical moral questions of what to do, think or be. And correct responses, expressivists argue, are stable ones.

Expressivists also need an account of judgment-producing processes as means to form true moral judgments that at least meet conditions (1) to (3). I will focus on (2) non-accidentality since it seems to be the most problematic given the challenge. We need a non-accidental connection between our moral judgment-formation processes and moral truth. One way for expressivists to think about this connection is that these processes produce better moral judgments such that they are stable or at least closer to stability, where stability is not something for our judgments to track but to *achieve*. This is an explanatory and non-accidental connection in the way means are generally explanatorily and not accidentally connected with their ends —more on this in a moment. Proponents of reliability and debunking challenges are right that moral judgment-formation processes are not connected with moral facts but can still be connected with truth. Stability is our proxy for truth; thus, if there are methods that produce better judgments such that they can reach stability *as a result,* then that is the non-accidental connection we were looking for.[204]

In our initial case, *S* judged that it was all things considered good for their friend to lie. And S came to form this judgment by an unbiased process *P.* The process here has a non-accidental connection with truth because it makes the moral judgments *better*: less arbitrary, more consistent with *S*'s other values and beliefs, or sensitive to distinctions that the biases would otherwise obscure. The idea of tracking moral properties in any substantial sense does not need to play any role in this picture since we are not positing those properties to begin with. So these means are disanalogous in the way we expected. Let us then call these means to produce stable judgments improving-means. We can contrast this with forming beliefs based on perception. In line with the ancestral picture, processes of basing beliefs on perception are systematic and non-accidental means for true beliefs. Roughly speaking, in normal circumstances, the process can connect beliefs with the facts they are about. The means give us ways to keep our beliefs in line with the things we are getting them *from.* Thus, while in the case of perceptual beliefs, *S* has the belief that *p* because of the fact that p via a process *P* based on perception, in the case of moral judgments, we can say that *S* has the judgment that *p* because of the goal that *p* is stable, via

[204] A similar but not quite parallel view is Marcus Arvan's (2019) proposal that for some parts of moral inquiry moral truths are not strictly speaking discovered but negotiated. This would be parallel and thus support my point in that means to produce true moral judgments would be primarily focused on the result, not the source —like improvement means. However, Arvan argues that some of the commitments of quasi-realist expressivists makes their picture of inquiry better classified as a discovery model, which contrast with his negotiation model.

process *P**. Though disanalogous, we can see that they are both instances of the same pattern we get from the ancestral account.

One question about this approach is whether this is the right kind of connection to secure the sense of non-accidentality we need to block premise ER). After all, the connection in the premise is of explanation. However, expressivists can argue that there is an explanatory connection: the one we ordinarily have between means and ends. Means-ends talk here is not very precise; our processes do not have ends in any literal sense. But there are plausible ways for expressivists to flesh out this idea. One is to think of moral inquiry as an *activity* with certain goals like producing true (or stable) moral judgments.[205] Reliable processes then have ends in the derivative sense that they are elements of that activity meant to help us achieve the end. And these means-ends relations are ordinarily explanatory: think about instances of explanations like "-why did you take my bottle? -To refill it."[206] One may question whether this is the right kind of explanation. While critics like Korman and Locke get right that what we need to respond to the reliability challenge is an explanatory connection, pinning down what that relation amounts to is extremely difficult,[207] and it is unclear that elements incompatible with the relation just sketched are essential to it. As Carrie Jenkins explains on behalf of expressivists, "'explanation'-talk is multifaceted and ordinarily allows for many different kinds of things to count as 'explanations'" (2015: 74).[208] Thus, even if stability as a result or end is not strictly speaking a fact, it can figure in relations of explanation with our processes of producing moral judgments. Moreover, this relation seems non-accidental in the way needed for reliability; it does not seem like a coincidence or accident that we end up with mostly true beliefs given the end of our processes.[209] The burden of proof then seems to be on the critic to show why this is not the proper explanatory connection.

---

[205] For a similar picture of inquiry as an activity with ends see Kelp (2021: Ch. 1). Expressivists often think of moral inquiry as aiming for stable moral judgments. See Gamester (2021: 451-3) for more on this.

[206] These are sometimes called teleological explanations, but they should not be confused with biological teleological explanations —although the latter may be a subset of the former. See for example Woodfield (2017).

[207] For some of the difficulties see Faraci (2019) and Noonan (Forthcoming).

[208] The suggestion that expressivists could use a relation of explanation to solve their issues with debunking challenges was, to my knowledge, first raised by Jenkins (2015), and my solution here owes much to hers. However, her solution is about knowledge so does not rely on the means-ends (or any) understanding of reliability, and the explanans in her view is a sentence.

[209] One possible disanalogy could be that teleological explanations are sometimes thought to be causal. See for example, Mele (2010). And the kind of explanation needed to block ER) is unlikely to be causal. See Faraci (2019). However, that view is controversial. See Sehon (2010: 124-6) for alternatives.

With all the elements in place, we can see how expressivists, using this approach, can deny premise ER). Recall:

> ER) Given the nature of moral thought and language in an expressivist framework, moral truths and moral judgments are not connected by any process *P*.

As explained in Section II, the challenge was that expressivists faced problems accounting for this explanatory connection from both sides. On the one hand, given the function of moral judgments, our processes for producing them do not seem to have the appropriate explanatory connection with moral truths. On the other hand, truth, for expressivists is not something that can have that connection either. We can now see how both claims are false. There is an explanatory connection between moral judgments and truth. Moral judgments aim at stability and thereby aim at truth. And, under this understanding of truth, it *is* something that expressivists can claim is apt for an explanatory connection, too.

Like Golub's, this view also relies on the idea that the substantial questions about the reliability of our moral judgments are first-order moral questions about which moral judgment formation processes are the best ones. In the picture here, judgments formed without biases are reliable because that is a process we are likely to endorse in moral practice, and we presumably do so because we think the judgments it produces are epistemically better. I believe this is an element of reliability that no expressivist should deny. What I deny is that the explanation of that reliability should proceed by mimicking the reliability of perception as much as the nature of moral judgments allows. Instead, expressivists can offer their own account of the explanatory connection between moral judgments and truths, which does not entail the lack of explanatory connection in ER). This meets condition 2): non-accidentality. Moreover, both accounts, that of the reliability of perception and that of the reliability of moral judgments, are instances of the more general ancestral account. Thus, even if there is a disanalogy between them, it should not represent a limitation for expressivists.

Condition (1), systematicity, should follow naturally from our account of condition (2). Processes like forming our moral judgments without biases are the kind of thing that not only have a non-accidental connection with truth, but it is also a connection that we can expect to hold systematically across different instances and under the right conditions. Thus, it is no surprise that these methods we trust in moral practice

and know have historically produced good results.[210] As for (3), no defeaters, we already covered one sort of defeater: the absence of an explanatory connection. But this picture should make room for more ordinary defeaters. Some considerations undercut the connection between our judgment formation processes and stability. Forming our judgments based, for example, on impartiality may not make them any better when the situation calls for a partial response. For example, when the correct response to a moral question depends on having the kind of personal relation with the subject characteristic of partiality. With (1) and (3), we have secured a connection between processes and truth that fits well with our general ideas about reliability but also meets an important strand of reliability challenges by denying ER). There is still, however, the question of what expressivists can say about ED).

## V. Evolution

A complete response to the debunking and reliability challenges does not only explain why we are good in producing true moral judgments but is also either as good as the evolutionary story or compatible with it. Providing an account that is as convincing but incompatible with the evolutionary story is a very challenging task, one that expressivists, as metaethicists, may not even be in the best position to offer. However, offering a compatible account seems more plausible with the alternative understanding of reliability. My strategy will be to offer an example of a capacity that was selected for its evolutionary advantages but which, applied to moral thinking, counts as improvement-means to true moral judgments. The example is Hugo Mercier and Dan Sperber's (2011, 2017) account of *argumentation* as a capacity for reliable *reasoning*. The idea is that this example will help as a proof of concept for expressivists to discharge the pressure of debunking arguments. Of course, a fully general account would be preferable, one that can help expressivists draw a more substantial link between our best evolutionary and empirical accounts of moral thought and processes we can characterize as improvement-means. I cannot offer that account here, but the example will be illustrative enough for our purposes.

Mercier and Sperber's theory is meant to solve an apparent tension between our ideas of reasoning as a reliable capacity to gather accurate information and empirical tests that show we do not tend to be good reasoners. Gathering accurate information offers a clear evolutionary advantage; without well-attuned capacities for it, a species like us is unlikely to survive. And we would think that a capacity for correct reasoning is

[210] We may even take the systematicity of processess we do not really understand as an indication that there may be a non-accidental connection like (1). See Graham (2017: 43).

one of our best means to gather this accurate information. This is not to say this is our only or even primary capacity for that task, but often enough, it is essential for it. However, empirical tests have repeatedly shown that we are not good reasoners. We often fail to follow simple deduction rules, have confirmation biases, or are easily led astray by irrelevant details. It is then puzzling that given the evolutionary advantage of reasoning, we are so badly equipped for it.

Mercier and Sperber's response is that our capacities for reasoning were not selected to solve problems in isolation but to find the best response in social settings by arguing for our proposed solutions. In these settings, we can appreciate the reliability of our reasoning capacities. And their tests seem to confirm the hypothesis. They asked participants to solve the problems that people usually fail when trying to reason by themselves, but this time in a social setting where they need to use their argumentative capacities. The results were that the best or correct answers tend to win in these settings. Thus, our reasoning capacity seems to be indeed selected for its reliability, but when understood as a capacity for argumentation and when it is not used in isolation. These results have substantial and interesting consequences for the expressivists' problems with reliability. This argumentative reasoning seems reliable in the sense we are looking for: having a systematic and non-accidental connection with truth. Mercier and Sperber pointed out that their view applies to moral reasoning too (2011: 68, 72-3; Mercier 2011). Moreover, they stated the view pairs well with Gibbard's expressivist picture of moral thinking (2011: 60, 68), and although they did not elaborate on that claim, we can work out the details here.

To see how Mercier and Sperber's picture applies to moral judgments and our picture of reliability, we can divide it into the two parts of the ancestral account: means and ends. First, processes of producing moral judgments by argumentation are means to produce true moral judgments. As Mercier and Sperber (2011: 58) think of arguments, they are complex meta-representations composed of propositions, of which one is a conclusion, and the rest are reasons for that conclusion. The capacity for argumentation is a capacity to represent other people's arguments, assess them, and, in most cases, respond with our own reasons for a different or similar conclusion. This is generally compatible with the conclusion and some of the reasons having moral content. It seems uncontroversial that we can offer reasons for our judgments with moral content —and following the discussion from the last chapter, some of these may be evidence, but that is not essential here. What may be controversial for expressivists is that propositions with moral content are not primarily representational. Thus, representations of them cannot be, strictly

speaking, metarepresentations. However, on the one hand, Mercier and Sperber are clear that these representations are not restricted to representations of ordinary objects and events.[211] Moreover, as they clarify, these representations are just meant to be propositions. Most expressivists accept that there are moral propositions, even if they are not representational. And these, too, can be supported by reasons.[212]

Second, the end of argumentation is truth. Just claiming that we can offer reasons for our moral judgments or propositions does not seem very informative. What is substantial about Mercier and Sperber's view is that this capacity was selected for its reliability only when it is used in social contexts. In these contexts, like in the non-moral case, Mercier and Sperber (2011: 72; Mercier, 2011) claim the best answer tends to win. This claim may be controversial since testers would have to assume they have the correct answers to corroborate the participants' results. Mercier (2011: 140-3) offers ways to work around the issue by focusing on arguments between children and adults. For our issue here, we are assuming we can reliably produce true moral judgments to look for the right expressivist explanation of that element of moral practices. And so, we can assume the truth of some of our moral judgments to grant Mercier and Sperber's point. However, it is important to be clear on what we are granting about the view. That social, moral argumentation *just* produces moral truths seems like a controversial first-order moral claim. But, first, Mercier and Sperber are working under the assumption that we already think of *reasoning* as a reliable capacity, which is relatively uncontroversial. Ther claim is that the best explanation of the evidence is that our reasoning capacities were only selected for their reliability when used in a social argumentative setting. Second, the connection with truth we are after here is relatively weak. This capacity to offer and give reasons for our judgments has a non-accidental but *defeasible* connection with stability —and so with truth. While not infallible, moral judgments produced in these settings are epistemically *better* since they are backed by reasons and these reasons are continuously assessed, which is a less controversial moral claim.

Argumentation in the right social setting and as applied to moral judgments seems to be compatible with an expressivist interpretation of it as improving-means. We

[211] A reason why they may think their picture is still compatible with Gibbard's expressivism is the example they offer of metarepresentations that are not of ordinary objects and events is *verbal* representations, that is representations of linguistic items. These may include moral terms understood are purely linguistic items and so may not involve commitments problematic for expressivists. In any case, positing non-representational moral propositions seems to be a better option to establish this compatibility.
[212] See Schroeder (2013), Ridge (2014: 124), and Brown (2019).

can explain how argumentation makes our moral judgments without thinking that any of the reasons make us better at *tracking* moral truth. What we get from argumentation, expressivists can argue, are epistemically *better* moral judgments, even if they are desire-like —like Blackburn's examples, after examining and arguing about the reasons for them, these judgments can be more consistent, imaginative, mature, sensitive to finer details, etc. Argumentation as a reliable process for producing moral judgments should be effective in discharging some of the pressure from the debunking challenge.

Consider the problematic premise:

> ED) Assuming both quasi-realist expressivism and evolution, there are moral truths that do not counterfactually depend on us, but our moral judgment formation capacities were not selected to track them. Thus, moral truths and our moral judgments are not explanatory connected by a process *P*.

First, Mercier and Sperber's view is an evolutionary account of the capacity for argumentation and should be compatible with the evolutionary history of our moral judgment-forming capacities and tendencies. Since Mercier and Sperber back up their account with empirical evidence, critics cannot accuse expressivists of offering a just-so or armchair evolutionary story if they appeal to this view. And even with its evolutionary plausibility, argumentation provides a systematic and non-accidental connection with truth. It is an improvement-means, so it should enjoy the explanatory connection we developed in the last section.[213] Thus, we have an example of a process that produces true moral judgments, and we can explain how that truth was not accidentally connected to it. Second, this is perfectly compatible with the expressivist picture of moral thought and language. Argumentation can connect our moral judgments with the truth, even if they are desire-like. There is no conflict between these two pictures, as ED) suggests, so expressivists can deny the premise.[214]

---

[213] One question here, however, is whether the explanation of the means-ends structure in terms of the activity of inquiry is the correct one. When we say that argumentation was selected for its reliability we seem to be talking about its *function*. See Woodfield (2017). Since all we need is that this story is compatible with expressivism I will stick to the activity interpretation. But there may be a way for expressivists to earn the right to this talk of functions too. See, for example, Köhler (2022).

[214] A question here is whether this response is also available to realists given that I am assuming that processes like argumentation are reliable in the sense we would endorse them from a first-order perspective. And that is something they would agree with. While I do not want to rule that out, argumentation as a way to block premise ED) is framed specifically as

To complete our response to premise ED), I will deal with two possible objections. An initial worry may be that basing epistemic evaluations on a reliable method like argumentation can give moral judgments a good epistemic status, like justification, too easily. For example, act-utilitarian judgments could be justified based on argumentation, even when a form of Kantian deontology is the correct view. However, recall that the project here was to explain how evaluations based on reliability are possible in an expressivist framework, and those are plausibly not infallible. Think of evaluations of *justification* based on reliability; it seems at least possible to have justified but false beliefs. I can believe justifiably that it is raining, given the reliability of my perception that it is, even if there is a defeater that I am unaware of —for example, that there is a recording playing rain sounds in the background. Reliability for moral judgments plausibly works similarly. Even if a judgment that I should lie given that it maximizes utility may be justified based on the reliability of argumentation, that is still compatible with the belief being false given the presence of defeaters. The account here then would not be unlike other plausible views on reliability.

Second, some debunkers may argue that the explanation of the development of our capacities and tendencies given evolution was not connected with truth in general and, for all we know, may have pulled us in a different direction. Thus, specific processes like argumentation that work with the results of those other processes will not be able to correct them. In response, this argument assumes the tracking account, in which the connection relevant for reliability is with moral facts that were there to be discovered to begin with. From that picture, it makes sense that truth is something from which evolution can pull us apart. But with the picture of reliability just offered, one that makes no use of tracking-means, whether individual evolutionary processes are distorting is something we cannot determine before engaging in moral inquiry. We may judge that we should value generosity because of the evolutionary pressures, but whether these are improvement-means that we could base the judgment on depends on the stability of the judgment. And that stability is something we can only determine by engaging in first-order moral inquiry.

We now have an expressivist-friendly account of our reliability, which explains how some of our capacities draw a systematic and non-accidental connection between our moral judgments and moral truth. This account is also consistent with the debunker's evolutionary account of our capacities to form moral judgments. There are aspects

---

an improvement means. Whether realists can endorse that further picture of reliability is unclear.

of these challenges that this account does not address. For example, I only offered one way to discharge the pressure of making my response to the reliability challenge consistent with the debunker's story by offering a proof of concept. Moreover, a comprehensive response requires further discussion of whether expressivists can legitimately appeal to the truth of some of our moral judgments —or, more generally, to the assumption that we are reliable in producing true moral judgment. However, my main goal here is to make sense of reliability as an element of our ordinary epistemic evaluations, and I am only addressing these challenges insofar as they are relevant to that goal. And in this respect, I believe the account should be effective in showing how are these evaluations possible.

## VI. Conclusion

In conclusion, like evidence, reliability is central in our epistemic evaluations. This was already a good enough reason for expressivists to explore an account of it that fits their views of the nature of moral thought and talk. But the account is also helpful in addressing some of the problems expressivists already face with reliability and debunking challenges. Even if I did not cover all aspects of these challenges, I hope that the account I offered here can open up more directions to address their more complex aspects. As for our project here, I believe the account gives us what we need for an expressivist moral epistemology: with accounts of both evidence and reliability, expressivists can be optimistic about earning the right to the basic epistemic concepts other views can just take for granted, and so have a good precedent for other epistemic evaluations. We are now in a good position to explore more substantial issues in moral epistemology, like the a priori status of some of our moral knowledge; this will be the topic of the next and final chapter.

# Chapter 6: A Priori Moral Knowledge

In this second part of the dissertation, I have explored notions central to epistemic evaluations: evidence and reliability. The goal was to provide a case that expressivists are able to account for the variety of epistemic evaluations by addressing a couple of representative and important examples. My discussion thus has abstracted away from distinctions between different sources of moral knowledge. This approach made sense, given that our epistemic evaluations, like knowledge and justification, cut across different sources: we can be, for example, justified in believing something via testimony, perception, reasoning, etc. But even if expressivism can account for various general conditions of epistemic evaluation, it would be in trouble if it cannot account for some particular kinds or sources of moral knowledge. More so, if we have independent reasons to believe that the sources of some of our moral knowledge, like that of fundamental or basic principles, are *a priori*. In this chapter, I explore the options for expressivists to explain this a priori status of some of our moral knowledge in a way compatible with their picture of moral thought and language.

The structure of the chapter is as follows. In section I, I explain why expressivists should want to account for a priori moral knowledge. Then, section II introduces a common way to frame the problem of explaining a priori knowledge in general, which will help us divide the existing options for expressivists. The main options are rejecting a priori moral knowledge as an illusion or mimicking some of the elements of intuitionism, according to which a priori moral knowledge has no empirical input. In sections III to V, I explore the prospects and problems for expressivists who want to adopt these three options. While I think aspects of these accounts are on the right track, they also have important limitations. Thus, in section VI, I introduce an alternative for expressivists to make sense of our thinking and talking of epistemic sources compatible with some of them being a priori. Finally, in section VII, I discuss a general objection to expressivism based on a connection between a priori moral and philosophical knowledge.

## I. Introduction: Why The A Priori?

Thus far, I have argued that moral judgments, even if desire-like in nature, are epistemically evaluable, and expressivists can make sense of some central elements of epistemic evaluations: reliability and evidence. These proposals gave us a good case that expressivists should, in principle, have no problem accounting for the variety of epistemic evaluations like justification, warrant, knowledge, etc. With

explanations of both epistemic evaluations and evaluability, expressivists can secure the two most important elements of their moral epistemology. However, epistemic evaluations apply to beliefs and propositions with different subject matters and sources. Some of these differences are relevant to our project here. Consider, for example, the difference between our knowledge of physical and mathematical facts. As many agree, the sources of mathematical knowledge are *a priori*; we can know mathematical truths without relying on empirical considerations. The same seems to be the case for *some* moral truths. For example, many believe we can know basic or fundamental moral principles a priori, without the help of empirical experience.[215] Think of principles like that one ought to maximise utility or not treat people as mere means; they do not seem to need empirical evidence to be epistemically justified or known. Given the plausibility that some of the sources of moral knowledge are a priori, it seems then that expressivists interested in a complete moral epistemology would be in trouble if they cannot account for it. As Teemu Toppinen explains, "It is very plausible that we have some normative knowledge, and it is also very plausible that some of this knowledge is a priori. If expressivism is to be otherwise successful, expressivists must be able to offer an adequate account of this." (2015: 248). An account of a priori moral knowledge then would be an important addition to my project here.

Focusing on the possible a priori sources for this final chapter represents a change in focus. Our epistemic evaluations cut across different sources of knowledge. We, for example, have reliable ways to form attitudes about a priori and a posteriori truths. Thus, my goal for this chapter is not to keep building the case that expressivists can account for epistemic evaluations by adding more representative cases. Instead, I want to show that expressivists can make these evaluations work across the different sources from which we plausibly get our moral knowledge. And since we have independent reasons to think that our knowledge of fundamental moral principles can be a priori, the challenge of accounting for it will be representative of this different approach. It is also currently unclear how our best accounts of a priori moral knowledge can fit the expressivist story of our moral thought and talk if they rely on

[215] Fundamentality or basicness is often cited as what distinguishes principles that can be known a priori. See McPherson (2013, 2020: 40), Depaul & Hicks (2021) for overviews and Ross (2002: xiii, xlii-xlix) for a classic articulation of this point. However, it is important not to confuse this sense of *epistemic* fundamentality, that we do not need further considerations to know (or be in a position to know) the principle, with metaphysical or explanatory fundamentality or with maximal generality of the principle. Plausibly we can know non-fundamental principles in the latter sense a priori, like mid-level principles. And so these are fundamental in the former sense. This epistemic sense of fundamentality is sometimes explained as non-inferentiality, and I will have more to say about it in section IV.

rational intuitions or self-evident propositions. And even the negative approach that most naturalists take of denying there is *any* a priori knowledge has not been adequately developed for this issue in expressivist moral epistemology. It is, for example, unclear whether the appearance of a priori moral knowledge is something expressivists as quasi-realists want to vindicate or as naturalists want to explain away. Either way, expressivists have something to explain.

There are many ways to think about the a priori, and much of our discussion later will help clarify the sense we need for our project here. One natural starting point is that a priori knowledge and justification can be obtained independently of *experience*. One can be justified in knowing that M) one ought not to treat people as mere means, without relying on empirical considerations. This is not to say that we cannot come to judge it with justification in other a posteriori ways, like testimony. It is the *possibility* of knowing it a priori that seems epistemologically important. A difference between knowing something a priori and a posteriori is a difference in sources. Sources in epistemology are often very broadly defined. Audi (2009), for example, describes them as something in the life of the knower that can provide them with justification or knowledge. These can be elements internal to the agent's capacities, like reasoning, understanding, and perception, or things that may need external factors for proper characterisation, like testimony. I will mostly refer to sources as procedures to keep this general characterisation. A priori sources are those that do not need to rely on empirical considerations to provide knowledge or justification, like pure reasoning.[216] A priori knowledge and justification differ in some respects, like their interactions with defeaters. However, to simplify our discussion, I will speak mostly of a priori knowledge and mark the difference with justification when relevant.

Most things we can say about the a priori beyond this rough characterisation are controversial.[217] One of the reasons the a priori knowledge is difficult to characterise is that radically different motivations for it are usually mixed in the literature. Using

[216] We have other examples that further illustrate this point like a priori warrant (Goldman, 1999) or a priori rules of rationality (Wedgwood, 1999) but those will not be crucial for our discussion here.

[217] One issue I am not planning to discuss here is whether a priori knowledge assumes that particularism is false. But I will assume a generalist framework in which moral principles (fundamental or otherwise) can and do play important roles in our practices. For a way to make a priori moral knowledge compatible with particularism see Dancy (2021). See also McKeever. & Ridge (2006) for discussion.

Tristram McPherson's (2018) distinction,[218] some of these motivations are fundamentally theoretical: theorists give a priori knowledge a fundamental explanatory role in their theories and then use those theories to explain our a priori knowledge of things like fundamental principles. For example, a priori knowledge seems like a direct consequence of the metaethical commitments of particular views. Suppose one is a non-naturalist who holds that moral properties are non-natural and so causally inefficacious. In that case, one is probably also committed to thinking of our knowledge of them as non-empirical and thus a priori.[219] A priori knowledge of fundamental moral principles is then an instance of this more general picture. Others, like Peacocke (2003: 198), aim for a general theory of the a priori across domains. Since fundamental moral principles are among the well-known candidates for a priori knowledge, applying the broader theory to explain them seems like a good test.

These fundamentally theoretical motivations do not, in principle, fit well with expressivism. Expressivists, even as quasi-realists, do not share the commitments that make realists need a view on the a priori. As we know, expressivists do not think moral facts are something you can access in any literal sense; thus, we do not need a non-empirical way to access them. Of course, they may have their own theoretical motivations. Expressivists deny that moral judgments have a primarily representational function; they are desire-like practical mental states. Thus, there is one important sense in which we do not acquire moral knowledge by experience (empirical or otherwise) of moral facts.[220] In that sense, expressivism could entail some form of a priori moral knowledge. However, this thesis does not say much about how we *do* acquire moral knowledge without moral facts; it is a negative thesis. Yet what is important about the idea that we can acquire moral knowledge a priori seems positive, that there is something interesting about the nature of our knowledge of fundamental moral principles.

Fundamentally theoretical motivations are not the only ones for expressivists to explain a priori moral knowledge. As McPherson explains, independently of the role

[218] McPherson uses the distinction between fundamentally theoretical and non-fundamentally theoretical claims about the a priori to establish a different but parallel challenge to naturalist versions of moral realism.
[219] Shafer-Landau (2003), Huemer (2005) and Wedgwood (2007) all defend a form of a priori moral knowledge as a way to support their versions of moral realism.
[220] Of course, some expressivists might want to say that in a different sense, that is, from the point of view of moralizing, it may make sense to say that we do acquire moral knowledge like that. My point is that experience should not play any *explanatory* role in our expressivist accounts of moral knowledge.

we assign to the a priori in our theories, there seems to be a fundamental divide between how we conduct moral and empirical inquiry. He points to what is characteristic of moral disagreement as a concrete example of this difference. We cannot seem to solve fundamental moral disagreements by agreeing on the relevant non-moral facts. And so, in those cases, we could reasonably end non-moral inquiry while moral inquiry remains open.[221] The autonomy of ethics discussed in Chapter 4 is, I think, a further example. The idea is that there is no reasonable inference to a moral conclusion based solely on non-moral considerations. And this has clear ramifications for how we conduct moral inquiry. Even if we can offer evidence with non-moral content that we get from experience for a moral judgment, this support depends on the right moral claim being in place as background assumptions. That shrimps cannot feel pain can give epistemic support to the judgment that it is permissible to eat them only with background assumptions about the badness of pain in place.[222] Expressivists then have good reasons to accept this divide between moral and empirical inquiry as a central assumption of moral practice. In fact, as McPherson explains, "this non-fundamental epistemological claim is markedly more plausible than any attempt to refine it in fundamental epistemological terms could be." (2018: 192). Let us call this the rationalist assumption.

> *The rationalist assumption*: there is a fundamental difference between how we conduct moral and empirical inquiry.

The rationalist assumption has a few concrete consequences for how we conduct moral inquiry. Two of them we already listed:

> *The disagreement consequence*: we cannot solve fundamental moral disagreements by agreeing on the relevant non-moral facts.
>
> *The autonomy consequence*: there is no reasonable inference to a moral conclusion based solely on non-moral considerations.

Thus, that there are moral truths we can know a priori seems like a third (but not unrelated) consequence of the rationalist assumption. Perhaps part of what it means to conduct moral and empirical inquiry differently is that our moral knowledge of fundamental principles does not need an empirical source. In this sense, expressivists

[221] This may actually be a strong reason for expressivists to want to make sense of what is distinctive about moral disagreement like its *intransigence*. See Cosker-Rowland (2018). Since this is a different phenomenon, I will have nothing to say about it here.

[222] Since particularism may be more relevant for the discussion here, it is important to be clear that what I say here is compatible with the moral claim being a moral principle or just appliable to this particular instance.

have motivations to explain the a priori, as a consequence of the very plausible rationalist assumption. Let us call this the a prioricity consequence:

> *The a prioricity consequence*: we can know fundamental moral principles a priori.

The rationalist assumption and the three consequences are the kinds of elements we can plausibly attribute to our moral practices; the first is just a claim about how we conduct these practices when it comes to inquiry, and the latter is a consequence of the former. Expressivists, as quasi-realists, want to explain and vindicate these practices, so they have strong non-fundamentally theoretical reasons to explain a priori moral knowledge. With this better motivation in place, the task for expressivists is to explain both the rationalist assumption and the a prioricity consequence.

To be clear, the other two consequences, autonomy and disagreement, are closely related to both the rationalist assumption and the a prioricity consequence and so will be relevant to the discussion at different points. But my claim here is not that they are necessary to account for a priori moral knowledge.

## II. Dividing the Main Options

We drew the a prioricity consequence from the rationalist assumption because one way in which moral inquiry differs from empirical inquiry is that moral knowledge often has a priori sources. The best way for expressivists to approach their task of explaining the two theses is by explaining or explaining away this difference in sources. Carrie Jenkins (2008: 438) offers a useful way to divide the main options depending on how one responds to what she calls the basic problem. Take a toy model of perceptual knowledge, according to which we go through three steps in acquiring it. First, we receive an external input, then process it, and finally, we get knowledge about the external world by basing it on the perceptual input. What is puzzling about any kind of a priori knowledge is that the first step is missing. We can know fundamental moral principles, mathematical truths or the laws of logic without empirical inputs. The task then is to explain how knowledge without the first step is possible, and the responses are usually divided into three: either there is no input needed, there is no *empirical* input needed, or a priori knowledge is just an illusion. These three ideas guide the main approaches to a priori knowledge, and I will henceforth refer to them as No Input, Non-Empirical Input and Illusion for brevity.

Moral expressivists have usually located themselves either in some form of the Non-Empirical Input or Illusion camps. However, these views have never been developed in-depth, their fundamentally theoretical and non-fundamentally theoretical motivations are never distinguished, and so their limitations are not fully appreciated. I will work through these in the following two sections. Ultimately, I will argue that the best option for expressivists is neither of these two but a version of No Input, one that appeals to a form of epistemic expressivism about epistemic evaluations that incorporate the sources.

## III. A Priori Moral Knowledge is an Illusion: Naturalism

With a few exceptions, the main assumption in the literature is that, for expressivists, a priori moral knowledge is an illusion. There are a few ways to motivate this view.[223] However, once we understand why expressivists should want an account of the a priori to explain the a prioricity consequence, the best version of this view will not seem as promising. Still, it is important to understand how the Illusion approach fails and the lessons we should draw from it.

Perhaps the best way to motivate an Illusion approach is via one of the central commitments of expressivism: naturalism. One of the most central motivations for an expressivist approach in metaethics is to locate our capacities of moral thought and talk in a natural picture of the world and us.[224] A priori knowledge then seems mysterious, a kind of knowledge that is somehow not justified by any part of that natural world. Outside of metaethics, some of the main proponents of this approach have similar motivations.[225] Simon Blackburn (2009: 211) shows some affinities with this view in claiming that moral practice may be compatible with some understanding of moral principles as necessary but not with them being known a priori. An initial difficulty here is that the rationalist assumption is a non-fundamentally theoretical claim, so it does not get its plausibility from a theory that rivals naturalism but from a simple and convincing observation about moral practice and inquiry. Thus, expressivists could, in principle, deny it, but not without a cost.

[223] One articulation of this view we can discard right away is Depaul & Hicks's (2012) and Mares' (2011: 139). According to these views, expressivism is committed to moral knowledge being, in general, an illusion since moral judgments are desire-like and thus unfit to be considered knowledge. If moral knowledge is, in general, an illusion, then a priori moral knowledge is too, by implication. However, as we know, this is just based on a general misunderstanding of quasi-realist expressivism.

[224] For more on this motivation and its role in the development of expressivism see Gibbard (2003) and Blackburn (2015).

[225] The classic reference in this respect is Quine (1951). See Jenkins (2008: 439) for an overview.

The quasi-realist commitment of expressivists means that they not only want to give an account of moral practice but also vindicate all those aspects of moral practice that seem incompatible with their approach. And given that as we are framing it, the plausibility of the rationalist assumption is due in part to the fact it is an element of moral practice; it is the kind of thing expressivists would want to vindicate, not deny.

A more plausible option for expressivists would be to accept the rationalist assumption and explain how it does not conflict with their naturalist commitments but question its connection with the a prioricity consequence. Expressivists are unlikely to have problems defending the first element of this approach. They have offered a few ways to think of moral inquiry as fundamentally different from empirical inquiry and still compatible with a naturalistic picture of the world and us. We have been advocating for one throughout the dissertation. According to this view, a central goal of moral inquiry is to get correct responses to our questions of what to do, think, or be. And for expressivists, a mark of these correct responses is that they consist of moral judgments that cannot be further improved or are stable through the relevant epistemic improvements like sensibility to finer detail, information, imagination, maturity, or coherence. Our practices of gaining moral knowledge are not guided by the goal of an accurate picture of moral reality in any literal sense but by the goal of improving our moral sensibilities. This picture of inquiry is compatible with naturalism. Expressivists do not need to posit moral facts or a special capacity to track them: those do not play any substantive role in the account. They only need natural facts compatible with a scientific picture of the world and us. And so, this picture of moral inquiry should be compatible with both naturalism and the rationalist assumption.

What could make this a version of the Illusion approach is the claim that once we understand what the rationalist assumption amounts to, there is no need for an a prioricity consequence —and so no need for some of our principles to get their status as knowledge without input or without empirical inputs. We conduct moral inquiry differently from empirical inquiry, but that does not mean that some moral propositions are epistemically special. We may think that there are, given that there are some moral propositions that are very important for us such that we cannot conceive a better moral view that does not include them, or because we think moral inquiry should help us find fundamental moral principles that could epistemically ground the rest of our moral judgments. However, the proponents of an Illusion account would argue that those convictions have nothing to do with the *nature* of our moral knowledge about those propositions.

The main issue with this approach is that even if we do not think that our knowledge of propositions about fundamental moral principles is epistemically special, we can still legitimately ask what is the source of that knowledge. Plausibly, all of *our* knowledge has a source, something from which we, as knowers, get it. And moral principles should be no exception.[226] However, it is unclear what resources expressivists currently have to explain that. Presumably, they do not want to say that the sources are the typical a posteriori examples like perception. And the whole point of an Illusion approach is to rule out sources that are commonly associated with a priori knowledge, like understanding or pure reasoning. But if none of these is the right option, what alternative do expressivists have? Again, this is a legitimate question, given the plausibility that all of our knowledge has a source, but the Illusion approach offers no way to respond to it.

From this Illusion approach, we know that whatever expressivist account of a priori moral knowledge needs to be compatible with their naturalist commitments. Expressivists already have the resources to get halfway there; their views on moral inquiry are compatible with the rationalist assumption. What still needs to be shown is how they can explain the sources of moral knowledge and whether their explanation vindicates or explains away the a prioricity consequence.

## IV. Non-Empirical Input: Intuitionism

Though expressivists' primary motivation to account for the a priori is non-fundamentally theoretical, that does not mean that they cannot try to mimic some of the tools developed by other theories that do give a priori knowledge a more fundamental explanatory role in their accounts. This may help expressivists overcome the limitations of an Illusion approach by focusing more directly on the issue of the sources. And perhaps the most well-known theory of a priori knowledge in metaethics, *intuitionism*, goes in line with the Non-Empirical Input approach. In this section, I will explain how intuitionists explain a priori moral knowledge —and so the rationalist assumption and the a prioricity consequence. This will give us a good way to contrast and better understand how expressivists currently mimic intuitionist resources, but it will also show what those views are missing.

Intuitionists like W. D. Ross (2002 [1930]), Robert Audi (1999, 2005, 2018) and Russ Shafer-Landau (2003) hold that some moral propositions, like those about fundamental moral principles, are self-evident or can be known by rational

[226] This does not depend on any view about the structure of epistemic support. Presumably even for coherentists the sources of our knowledge is the overall coherence of our beliefs.

reflection.[227] Self-evidence is the property of some true propositions that, by adequately *understanding* them, we can come to know them.[228] As most intuitionists clarify, self-evidence does not mean the proposition is obviously true. Understanding a self-evident proposition may be a difficult and costly process, and understanding does not guarantee knowledge of the self-evidence itself. Moreover, though they play no role in grounding our knowledge of self-evident propositions, we may still need empirical inputs as a means to understand them. This notion of self-evidence explains why we do not need empirical input to know fundamental moral principles; the only input we need is the self-evident proposition, and our source of knowledge of it is understanding. Thus, intuitionism is helpful in explaining the nature of our moral knowledge, and with that, things like the rationalist assumption and the a prioricity consequence. This is also why intuitionism is interesting for metaethical views like realism and expressivism. However, intuitionism is not only a view about the *nature* of moral knowledge.

Intuitionists are often motivated by providing a foundationalist moral epistemology that responds to sceptical worries. One such worry is that we either get some of our moral knowledge non-inferentially or do not get it at all. We get most of our moral knowledge inferentially or based on further considerations. These further considerations need to be known themselves if they will provide us with knowledge in the first place. And given the autonomy of ethics, at some fundamental level, the considerations that provide the basis for the rest of our knowledge must be purely moral. Following this logic, there must be a point where we either have some purely moral knowledge that does not get this status inferentially or, as it turns out, we do not have any moral knowledge to begin with. Intuitionists argue that knowledge based on understanding a self-evident moral proposition can provide that non-inferential knowledge. What makes fundamental moral principles candidates for a priori knowledge is that they are fundamental in this non-inferential sense. Though, to be clear, which are the fundamental moral principles is a very contested matter

[227] Not every view labelled 'intuitionism' merits the classification as Non-Empirical Input — or should be classified as a view of a priori knowledge. Bedke (2008, 2013) divides intuitionism in two senses. One is in line with what I explain here, and refers to rationalist views that appeal to a sense of self-evidence or rational reflection to explain a priori moral knowledge. The other refers to views about distintive mental states called intuitions which are analogous to perceptual inputs on which we can base beliefs. Intuitions in that sense do not need to involve a commitment to independence of experience and so are not directly relevant for the present issue of a priori moral knowledge. Thus, here I will restrict my use to the former sense of intuitionism.

[228] For this definition and a very complete account of the conditions for understanding a self-evident proposition see Audi (1999, 2018).

and though we used 1), that one ought not to use people as mere means, as an example, it is by no means uncontroversial. Intuitionists agree with this and would argue that understanding a self-evident proposition is complex, and we cannot guarantee that we can identify which propositions have this property — understanding a self-evident proposition does not mean we understand it is self-evident.[229]

The fact that intuitionism has this anti-sceptical element reflects a view not often made explicit in discussions of a priori moral knowledge: a priori sources not only help us explain how moral knowledge is possible it also grounds it. Audi, for example, offers a way to make the notion of sources more precise by separating these two elements: "Understanding sources of knowledge, and as they are generally conceived in philosophical literature, they are not just where knowledge comes from; they also provide the knower with grounds of knowledge." (2009: 82) Taking my headphones off because my ear is itching may have caused my auditory perception of rain and also caused the justified belief that it is raining, but it is not the source. The *perception* that it is raining is both a cause and a source of my justified belief that it is —provided I base the belief on the perception and there are no defeaters. A priori sources work similarly. If we assume intuitionism, we may say that I know that 1) one ought not to treat people as mere means because I understand that proposition and that understanding may be both a cause and a ground for my knowledge. Thus, sources then play this double role of causes and grounds for our knowledge. A crucial question for expressivists is whether they can explain this dual nature of epistemic sources.

The intuitionist way of explaining sources is unlikely to attract expressivists in the first instance because it is commonly associated with a form of cognitivist moral realism. Realists who think of moral facts as non-natural, like Russ Shafer-Landau (2005), are attracted to the intuitionist approach.[230] If they think that moral facts are non-natural, then it makes sense that their picture of moral inquiry should be one of discovering moral reality via non-empirical methods. With this picture of moral inquiry, it makes sense to think of intuitionism as a Non-Empirical Input approach

[229] For more on this argument and the anti-sceptical goals of intuitionism see Audi (1999, 2005: Ch 4) and Sturgeon (2002).

[230] Likewise, Huemer (2005) and Wedgwood (2007) argue that their versions of non-naturalist realism requires a form of intuitionism. Although they make no use of the notion of self-evidence so I will not discuss them here.

since our a priori procedures, however we characterise them, are meant to correctly represent the inputs we can get from existing moral properties.

Even if some forms of moral realism may need intuitionism, it is unclear whether intuitionism entails moral realism. The latter depends on a few assumptions about what a self-evident proposition is and what understanding it amounts to. For example, Audi (2019: 375) claims that self-evidence is at its most plausible when the concepts that figure in the proposition stand for abstract entities that ground its truth. A different way to think about this is that some moral propositions are self-evident, but not all. And so there is a question of which are self-evident and which are not. Realists have a readily available response: self-evident propositions represent moral facts. However, it is unclear what expressivists can say to explain the difference. Again, this does not make self-evidence or understanding as a source incompatible with expressivism. Whether there is substantial incompatibility is not as straightforward, so expressivists could take advantage of this to recapture some of the elements of the view.

Take the two main elements of the view: self-evidence as a property of propositions and understanding them as a source of knowledge. One way for expressivists to approach this view is to accept the latter element and explain it in their own terms but reject or substantially revise the former. More specifically, they should reject that self-evident propositions can be a non-empirical input in their own right. There must be a different explanation of why these are the propositions we can know via understanding or other a priori sources. And this explanation should not involve the representation of moral facts. I believe this is a fruitful way to frame existing expressivist options that appeal to epistemic expressivism. This framing will help us draw a more direct connection with the issue of explaining the sources of moral knowledge and better understand what expressivists need to say about it.

## V. No Input Needed: Two Forms of Epistemic Expressivism

The best approach for expressivists to offer a parallel account to intuitionism is via a form of epistemic expressivism. This is Allan Gibbard's (2002, 2003: 33-7; 221) approach. According to Gibbard, there are only two elements of an intuitionist view that expressivists want to reappropriate: a view about the mental state we have *about* the moral proposition and a view about what it is to think of that judgment as knowledge. The existence of moral propositions is relatively uncontroversial among expressivists nowadays. What is controversial is how to best characterise them beyond the fact that their function is *not* representing moral properties and facts.

However, regardless of how we decide to characterise them, evaluations of a priori knowledge target the judgment with the moral proposition as content. And that judgment can be a desire-like moral judgment.[231] Think again of 1) one ought not to use people as mere means. For expressivists, a moral judgment with that content is best understood as a desire-like disapproval of using people as mere means. The second element is how we think of this judgment as knowledge. Gibbard's proposal, then, is to appeal to a form of epistemic expressivism.

We covered epistemic expressivism in other chapters, but to recapitulate, according to epistemic expressivism, since our epistemic talk has an essential normative element, we can offer an expressivist account of it. For example, Gibbard (2003: 199, 2012) argues that our knowledge talk expresses practical states like plans to rely on a specific judgment. Take a specific epistemic evaluation, like knowing that 1) one ought not to use people as mere means. Given epistemic expressivism, we should understand this as the expression of a state of planning to rely on that judgment. This aligns with how expressivists approach these issues in general. As we know, expressivists would not start their accounts by asking what a priori moral knowledge is, and instead would ask what it is to *think* and *talk* about knowing a moral proposition a priori.[232] Thus, we can apply that framework to our issue here.

Evaluations of a priori justification are different from ordinary evaluations of justification. As Gibbard (2002: 227, 2003: 34-5, 2008: 20-3) explains the difference, in his picture of epistemic justification for moral judgments, there must be a place for *intuitions*.[233] According to him, it is just a psychological fact that we sometimes judge things like that pain is bad on no further grounds than that we think so, and we would still think it upon challenge. And in the case of *some* judgments thinking so is enough to get a form of defeasible epistemic justification. There is a sense in which this justification would be a priori, since it rests in no further consideration, empirical or otherwise. Thus, this psychological fact has some normative epistemic import:

[231] Gibbard is not alone in trying to capture what intuitionists say by focusing on intuitions as desire-like states. See also for example Lenman (2018) for a different example. And they may have some independent support that some mental states can rightly call intuitions are desire-like states. See Loev (2022).

[232] I explain this expressivist manoeuvre in more depth in chapter 2.

[233] At many points Gibbard seems to suggest he understands intuitions in the more psychological sense that, as I explained in footnote 214, may not be relevant for a priori knowledge. However, as I explain below, he thinks of the term intuition in a psychological and normative sense. It is in the normative sense that I believe his views are relevant for a priori moral knowledge and rival intuitionist views on self-evidence.

> "An intuition, we could say, is a state of mind of accepting something, not on the basis of further reasoning even upon challenge, that we ought to place some trust in. To think something an intuition in this sense is to plan to rely on it. I'll call intuitions in the non-normative sense in which they figure in psychology "de facto" intuitions. These are judgments made confidently, on no further grounds, with no felt need for further grounds even upon challenge. Intuitions in the normative sense I'll call intuitions "de jure." These are de facto intuitions to rely on. It's a normative claim, then, that de facto intuitions are genuine intuitions—and one we need, I have been claiming, for coherent planning." (2008: 23).

While Gibbard does not say much more about this dual character of intuitions, there is something right about this distinction that we could expand on and connect with the issue of sources. We can think about it this way: once we understand what it is to evaluate something epistemically at all, we can see that the question of which are the sources of moral knowledge may not be for expressivists to solve. We conduct moral inquiry differently from empirical inquiry because, at some points, we cannot rely on empirical considerations or any considerations anymore, just on intuitions (in Gibbard's sense). This is just a fact about moral inquiry —perhaps one essentially connected with the rationalist assumption. And nothing about expressivism is incompatible with that fact. Intuitions, as mental states with moral contents, can be understood as desire-like states. However, whether we can sometimes evaluate some of our judgments as justified on no further grounds is something for us to decide when we engage in moral inquiry; it is, as Gibbard explains, a normative issue, not for expressivists to decide. Thus, unlike intuitionists, expressivists may not need a way to distinguish between which propositions are self-evident because our justification to believe them is not grounded on a property of the proposition itself but on how we think of the judgment with the proposition as its content.

While I agree with this general picture, I think it is still missing a crucial piece. Even if no property distinguishes self-evident propositions, our *thinking* of a judgment as justified seems to be different from thinking of it as justified on no further ground. Of course, expressivists could say that these are just two forms of thinking and there is nothing further about their difference that they need to explain. However, the difference is directly connected to how epistemic expressivists explain our epistemic *thought*. For example, planning to rely on someone's judgment or endorsing certain moral judgment-producing procedures are states that are responsive to reasons or at least to the truth —as we saw in Chapter 2. However, it is unclear how that

connection with the truth works when evaluating an intuition in Gibbard's sense, when there are no further grounds.[234] Thus, given the state we would be expressing, expressivists do owe us an explanation of the difference.

Hartry Field's (2001, 2018) epistemic expressivism about a priori knowledge of logic[235] offers a different way to flesh out a view like Gibbard's. Field argues that propositions with a claim to be known a priori are those that we are entitled to believe by default, or our justification for them has what he calls default reasonableness. We can divide his explanation into two parts that correspond roughly to the concepts of reasonableness and defaultness. Field explains reasonableness similarly to Gibbard, using epistemic expressivism. To call a proposition one we could reasonably accept is to express a mental state like a plan to rely on it. What is specific to Field's view is that when evaluating something a priori, our options are to accept the evaluation without any further support or not accept it at all. But given that propositions about logic are so central to our theories, not accepting them at all would be a big cost. Thus, the best option is to accept them without further support: this is defaultness.

Field claims that the best way to locate judgments that can be known a priori is via a chain of epistemic support —sort of in the way intuitionists explain non-inferential knowledge. If we keep inquiring about the epistemic support of our claims, we find judgments that either can be held without further support or not held at all. In the case of logical truths, we have a very good way to mark how we reached the end of the chain: to justify them, we need to assume them. Say, for example, that we need to justify the rules that govern the conjunction of two propositions: *p* & *q*. To do this, we most likely need non-atomic sentences, just like *p* & *q*. For example, we need to be able to claim that a conjunction like *p* & *q* is true when both *p* is true, *and q* is true. But claims like that need to follow the rules of conjunction, too. So, we would have to assume something like the rules of conjunction if we are to justify the rules

[234] To be sure, this challenge is for the epistemic expressivist element of the view. It is the state we express when we think of the intuition as knowledge that is supposed to be responsive to the truth. Other senses of intuition that do not have this element do not have this problem. Intuitions as quasi-perceptual states, for example, could be unresponsive to truth by persisting even after realizing upon reflection that they are false.

[235] Here I need to make two clarifications. Field extends this account to our knowledge of some methods of epistemic justification like induction and deduction. I will restrict my presentation to logic to keep things simple but nothing I say here should be incompatible with the other parts of his view. Second, Jenkins (2008: 440) believes Field's view requires a different category all together, and the No Input label should be reserved for views that think of a priori knowledge as trivial. I do not think we need to worry about that distinction here. Most if not all views of a priori moral knowledge reject the claim that it is trivial (or analytic). Moreover, as it will be clear Field's view counts as a No Input approach for the purposes for which we have been using the distinction here.

of conjunction. Justification of logical laws seems, then, inescapably circular. Even if Field does not emphasise this part of his view, I believe it is crucial: it explains what is distinctive about our thinking of a priori justification —at least when it comes to logic. Put in a different way, even if we do not understand a priori justification as a property, like self-evidence, we still need something that marks judgments that are evaluable in this way. And for the case of logic, the inescapable circularity of their justification is a very good candidate. We can call this defaultness by circularity.

Field's view improves epistemic expressivism of evaluations of a priori knowledge by adding defaultness by circularity. To distinguish propositions we can know a priori we do not need a property of the proposition itself, like self-evidence. Rather, circularity is a matter of how we can approach the justification or knowledge of the proposition given, on the one hand, how central it is for our theory, but, on the other, how limited our options are: either we accept the justification without further support or we give up on the proposition. The problem is that there seems to be nothing like defaultness by circularity in ethics. Even granting that, like the laws of logic, we get to our fundamental moral principles by a chain of epistemic support —which is itself controversial— we cannot know that we reached them by an inescapable sense of circularity. Think again of 1) one ought not to use people as mere means. Even assuming we cannot offer further epistemic support for it, it seems implausible that this is because we need to appeal to it in order to justify it. Unlike the rules of conjunction, we do not need to assume its truth to even articulate it. But 1) is the kind of moral principle we seem to be able to know a priori. Thus, even if the way we approach the evaluation is important, the parallels between a priori evaluations of logical and moral knowledge seem thin in this respect. Perhaps we need a different example. Some may object that 1) is not as fundamental as we should expect principles we can know a priori to be. For example, under some views, to justify the proposition that pain is bad, we need to appeal to the nature of our experience of pain itself, and so we reach a kind of circular bottoming out.[236] Thus, the proposition would be fundamental in the right way. However, even granting that is a form of circularity, it is not analogous to the defaultness by circularity Field is talking about. It seems, for example, possible to articulate the proposition without assuming its truth. But if the idea is just that we seem to be able to just appeal to our thinking of pain as bad to justify the proposition, then that seems closer to Gibbard's idea of an intuition, and so it shares its limitations.

[236] Parfit (2011: 54) for example, seems to suggest this.

## VI. Epistemic Sources for Expressivists

Expressivists have good reasons to explain the rationalist assumption because of what it seems to tell us about moral inquiry and the a prioricity consequence because of what it seems to tell us about the sources of some of our moral knowledge. Both the No-Input and the Illusion approaches have some important limitations. Most centrally, they seem to leave some crucial questions of the sources of our knowledge of fundamental moral principles unanswered. While the Illusion approach is right that expressivists want their accounts of moral knowledge and the rationalist assumption to be compatible with naturalism, it leaves us with no answer to the question of where our moral knowledge comes from. Existing versions of epistemic expressivism offer good general explanations of what it is to think of something as justified in general, so there are elements we could draw to think of a priori justification. But these views either cannot explain the difference between thinking of something as justified and thinking of something as justified a priori, or are not the right fit for *moral* a priori knowledge.

Sources are both causes and grounds for knowledge. One way we could approach the limitations of views like Gibbard's is that when asked why think that 1) one ought not treat people as mere means, all we could say is in the causal sense, things seem that way even upon challenge. But it is hard to think of that psychological fact as a source of knowledge in the grounding sense. In this latter sense, the source is rather a capacity that can connect the subject *S* with the truth of 1). In that second sense, expressivists like Gibbard could say that a priori knowledge consists of taking for granted the truth of certain propositions provisionally while we get to some ideal form of coherent planning. This would not be unlike our picture of reliability from Chapter 5, in which our moral judgment-producing capacities are non-accidentally connected with the truth via the goal of stable moral judgments. Like in that picture, stability through the relevant improvements can work as a proxy for truth. A judgment formed with a process meant to produce stable moral judgments would have a secure connection with the truth we seek. What is distinctive about the source in this case is that the connection with truth is secured a priori via taking the judgment to be true provisionally, without further support, with the aim to, at one point, get stable moral judgments out of it. Using our terminology from before, we can call this defaultness by provisionality.

This outline of a view would get us closer to an expressivist picture that would address the issue of sources directly. To get a plausible grounding element for a priori

sources beyond what epistemic expressivism tells us, expressivists needed to make better sense of how that grounding was connected with truth. Moreover, since this picture relies on stability as a proxy for truth, it should be compatible with the rationalist assumption like the initial Illusion approach. However, I do not think expressivists need to be limited to defaultness by provisionality. What makes the view plausible is more general. Consider this structure of epistemic evaluations grounded in sources:

> S) The moral judgment that *p* counts as knowledge given a procedure *P* that causes and would be conducive to the stability of *p*, and so grounds the evaluation.

Some instances of *P* could be a posteriori sources that incorporate non-moral evidence —like the ones we considered in Chapter 4. But apart from those, expressivists are free to incorporate any sources conducive to stability but do not need further support. Defaultness by provisionality would be one instance of this more general pattern. But crucially, which sources can ground our evaluations and which cannot is a normative matter for us to resolve in first-order moral inquiry. Expressivists do not need substantial commitments in that respect, so they can have a relatively broad view of a priori sources:

> S*) The moral judgment that *p* counts as knowledge given a procedure *P** that causes and would be conducive to the stability of *p*, and so grounds the evaluation without requiring any further support.

Expressivists then could argue that this is all they need to say to explain the a prioricity consequence. Part of what is involved in the rationalist assumption, that we conduct moral and empirical inquiry differently, is that some of the sources of moral knowledge are a priori. What that means for expressivists is that, from the perspective of moral practice, there are plausible instances of evaluations like S*). However, which sources could figure here as *P** is for us to decide as a matter of first order-moral inquiry about which procedures we trust will produce and ground true moral judgments.

A remaining issue for expressivists to address would be to show that their view is *compatible* with our best candidates for a priori sources that meet condition S*). However, now that we understand what sources amount to, it is unclear if there are especially problematic candidates, even *understanding*. Consider judgments like M2) discrimination would still be wrong even if we think it was not. As we know, expressivists would interpret this as a judgment about the conditions under which

one should make the judgment that discrimination is wrong, namely even those in which one does not think it is. But on the one hand, this seems to be the kind of truth one can know a priori, without any empirical or further consideration. On the other hand, as many expressivists would argue, there is a sense that understanding the proposition would be enough to be in a position to know it. For example, knowing that it is just part of the correct application of the concept of wrongness that we should apply it like this is part of what being competent its use depends (Sinclair, 2018; 2021: 202-16) or part of what categorizing actions as wrong involves (Brown, 2019).[237] Understanding some moral propositions then would be enough to appreciate that they would not be ruled out after improvements in coherence, sensitivity of finer details, imagination, etc.[238]

An important question is whether this explanation of understanding is not incompatible with the version of evidential holism defended in Chapter 4. If it is, then expressivists seemed forced to choose between an appropriate account of evidence and accounting for the possibility of a priori knowledge. In response, the core idea of holism is that no consideration can evidentially support anything by itself. This may still be the case for propositions known a priori. Similarly to how intuitionists respond to holistic worries, we may still need background considerations as a means to understand the proposition. These would not make the evaluation non-a priori since they do not play a substantial role in indicating stability.[239] Of course, this is just a sketch, and there are likely further important issues to address.[240] My point, however, is not to offer a detailed explanation of understanding for expressivists but to offer an example of how expressivists may not need to rule out paradigmatic and plausible examples of a priori sources like understanding. Whether we should rely on it as a source is a question for first-order moral inquiry and not for expressivists to resolve. This is also why expressivists do not need to mimic the foundationalist element of intuitionism.

---

[237] As indicated in section V, which is the right view of moral propositions for expressivists is a very complex matter. However, it seems relatively uncontroversial that whatever the right account is it should not give any role to the representation of moral properties, and so moral concepts should not have the role of picking them out. And the role we do assign to moral concepts is based on or at least related to the general practical function of moral thought and talk. The two views suggested here depend only on different articulations of these two points.

[238] This is also perfectly compatible with other procedures *P* that could produce and ground the same judgment, but are a posteriori, like testimony.

[239] See Sturgeon (2002) for a similar response.

[240] Intuitionists like Audi (2019: 364-74) offer detailed conditions understanding a proposition should meet. For example, recognitional range, a sense of rejectability, logical comprehension, confirmational sensitivity, etc. It is then a good question how these work on an expressivist framework. That, however, is beyond the scope of this chapter.

As we know, intuitionists who rely on self-evidence think of a priori sources not only as a view about the nature and sources of our moral knowledge but as a way to respond to sceptical worries with the possibility of non-inferential moral knowledge. The view offered here says nothing about that, and this may be especially problematic for expressivists. Recall from Chapter 1 that expressivism is often wrongly paired with scepticism in the contemporary landscape. We already addressed some of the motivations for that view. Expressivists, as quasi-realists, are not interested in denying moral knowledge; their views on moral thought and language are not incompatible with epistemic evaluability and evaluations. But perhaps the fact that their view offers no way to address sceptical worries could be a different reason why expressivism seems to fit awkwardly with other views in moral epistemology —like intuitionism. However, expressivists are likely to argue that though they want to vindicate our moral convictions that there is moral knowledge and explain how that is possible, their job is not to convince a moral sceptic.[241] They can perhaps separate two kinds of questions with the framework just offered. One possible question is what the source of our knowledge of *p* is for a moral *p*. Expressivists can explain that, when moralizing, we can offer responses of the form S) and S*), which is completely compatible with their framework. A different question is which instances of S) and S*) are plausible in light of sceptical doubts. That, they can argue, is not a question we can answer by explaining the nature and sources of our moral knowledge. In that sense, foundationalist intuitionism, for expressivists, seems like a first-order view on how best to conduct moral inquiry in those circumstances. The expressivist explanatory project and the intuitionist's anti-sceptical goals are orthogonal.

## VII. A Priori Normative Knowledge and Self-Defeasibility

Teemu Toppinen (2015) claims that expressivists need to make sense of a priori moral knowledge because it would dispel a worry about self-defeasibility. As stated by Crispin Wright (2002b), the worry comes from the plausibility of normativism about mental states —which many expressivists accept and we used in Chapters 1-3 to establish and respond to the challenge of epistemic evaluability. According to this view, the correct account of mental states like desires and beliefs must include the norm or norms that regulate them. For example, the norm is to aim at goodness/reasons for desires and truth for beliefs. In this sense, correct views about these mental states are normative views. By being committed to a non-cognitivist

[241] Two especially clear articulations of this view are Blackburn (2017) and Lenman (2018).

view of moral judgments, expressivism wants to account for these mental states, but by their own admission, such a view will be normative. In this sense, expressivism itself is a normative view; it entails normative claims about mental states like moral judgments. Expressivism is also a philosophical view, so if it can be known at all, it has to be a priori. Wright argues that if things can be known a priori, rejecting them must involve some kind of cognitive error. But if non-cognitivism is true, mistakes in normative knowledge should not involve primarily cognitive errors (or successes, for that matter). However, since expressivism is a normative philosophical view that we can know a priori, there is no way expressivists can make sense of the conditions to know whether their own view is incorrect. And so it seems either expressivism is self-defeating or it cannot account for all of our normative knowledge. Either option seems problematic.

Our account of a priori moral knowledge represents good and bad news for expressivists who want to respond to Wright's objection. Toppinen argues that by making sense of a priori moral knowledge, expressivists can make sense of a priori knowledge of normative views in general. Given that expressivism is a normative view, they would be vindicating themselves by vindicating their views about a priori knowledge of normativity. The problem with this argument, as we can now see, is that the sense of a priori *moral* knowledge expressivists want is much more restricted and not straightforwardly transferable to all areas of normative inquiry. It employs elements distinctive of moral practice and inquiry, and whether they apply philosophical inquiry, normative or otherwise, in any straightforward way is unclear. For example, the pressure to aim at moral judgments that cannot be further improved comes from the dynamics of moralizing. When moralizing, we want the correct responses to our questions of what to do, think or be. Whether the dynamics of our practice of philosophical inquiry are the same is difficult to tell —consider, for example, what role correct responses to metaethical questions play in the dynamic of philosophical inquiry. To be sure, I am not saying that there is no parallel. The point is that the translation from one domain to another requires substantial assumptions about the epistemology of philosophy that expressivists should establish first. Still, expressivists are in a good position to put some pressure on Wright's argument. The crucial premise is that if a view is a priori justified, we would display a cognitive failure by rejecting it. And we know now that not every view on a priori knowledge must involve this element. Moreover, the view here about a priori moral knowledge is just an instance of the general pattern of employing epistemic expressivism. Field's view on a priori knowledge of logical truths may very well be

a different instance. Thus, even if we cannot translate the account directly to philosophical knowledge, it would at least not be ad hoc to think that a similar option may be available.

## VIII. Conclusion

In the first five chapters of the dissertation, we focused on the possibility of epistemic evaluations for desire-like moral judgments. A big driving motivation for that task was to show that moral judgments were sufficiently similar to ordinary beliefs non-moral to make sense of similar epistemic evaluability and evaluations. However, moral epistemology is distinctive in many respects, one central one being that it seems to assume the possibility of a priori moral knowledge of fundamental moral principles. I argued that some aspects of that commitment are not always distinguished, so even if there was a sense that expressivists needed to account for that possibility and have some available options, many crucial aspects of that task remained unclear. By working through these issues, I opened the door for a different option that captures what was right about the existing ones. As a result, we can now appreciate what an expressivist commitment to a priori moral knowledge amounts to and what we gain by appealing to a form of epistemic expressivism. Epistemic expressivism was often appealed to as a key to developing an expressivist moral epistemology, but it is not often developed as a view about different epistemic evaluations. The account I offer here about a priori knowledge and justification should be a helpful start.

## Concluding Remarks

We saw in Chapter 1 expressivism is generally excluded from the landscape of contemporary moral epistemology. We are now better positioned to see why that is a mistake. However, as I explained in Chapter 6, the idea should not be that expressivists have an account of the sources of knowledge, which can provide a secure foundation for our moral knowledge and dispel all sceptical worries. Unlike other views, expressivism is not in the business of responding to the sceptic. Still, whether we are dealing with sceptics or ordinary people, what is central for us in moral practice when we use epistemic concepts is to present and evaluate our moral views. And expressivists can argue that they provide as much vindication of those practices as we can reasonably ask from them. We can think of this in terms of the three steps that Bex-Priestley and Gamester (2023) explain quasi-realists need to go through to earn the right to and so vindicate aspects of our moral practice.

First, expressivists need our epistemic evaluations to *make sense*; to show that they are meaningful. Expressivists already did well in this respect since most of them endorsed a form of epistemic expressivism according to which the *meaning* of epistemic evaluations is explained in terms of conventionally associated non-representational mental states. Thus, given epistemic expressivism, it is easy to see how our epistemic evaluations are meaningful in an expressivist framework. Moreover, expressivists may not even need radically different conceptions of justification, knowledge, rationality and the rest to make good on this part of their project. As shown in the second part of the thesis, they can earn the right to some of the most central concepts used in those evaluations, like evidence, reliability and even a prioricity. The second step is to prove that there is nothing *incoherent* in making our ordinary epistemic evaluations for moral judgments and assuming expressivism is true. I believe this is the best place to locate the challenge of epistemic evaluability. The problem is that there seems to be something inconsistent in thinking of our moral judgements as desire-like and epistemically evaluable. And even if we understand those evaluations expressively, that tells us nothing about their proper objects. We can think about my discussion in the first part of the thesis as a way to understand and respond to that worry. In that way, expressivists should be confident that they can take the second step in vindicating our ordinary practices of epistemic evaluations.

Lastly, the question is whether the practice is justified, given that expressivism is true. As Bex-Priestly and Gamester explain, this is, strictly speaking, a first-order

moral matter. But we can now easily see how that is the case. In moral practice, we often need correct responses to our moral questions of what to do, think or be. And our best approach is finding responses that are reliable, backed by evidence, justified, warranted, rational, count as knowledge, etc. Thus, we have excellent reasons to evaluate our moral judgments epistemically, even assuming expressivism is true. With these three steps, expressivists can vindicate our ordinary practices of epistemic evaluation, so there is no reason to exclude them from the landscape of contemporary moral epistemology.

Over ten years ago, Mark Schroeder wrote:

> "[E]pistemological notions promise to be some of the hardest to get straight regarding what the right non-cognitivist account of them should be —not least because these epistemological notions are controversial in their own right. This is because, if someone knows something only if they truly believe it, then providing an expressivist account of knowledge will be at least as hard as providing an expressivist account of belief and providing an expressivist account of truth, put together —but non-cognitivist accounts of belief and of truth are themselves complicated topics, each in its own right. And that still leaves out whatever further condition is required to turn true belief into knowledge, which epistemologists have had enough trouble understanding, even on the assumption that noncognitivism is false. So, all told, there are excellent reasons to expect the epistemological issues facing expressivism to be very difficult, as well as being particularly difficult to resolve without first resolving general issues about logic, truth, and belief." (2010: 139)

I was unaware of that passage until I almost finished writing this dissertation. However, it is striking how much I came to agree with him about the fundamental problems of an expressivist moral epistemology and how difficult it would be to get the responses right. I hope the problems I presented give a fair articulation of that difficulty. But I also hope the accounts of epistemic evaluability and evaluations I offer go some way into addressing them.

Some expressivists are likely to be familiar with some of the elements of the accounts presented; I hope the framing here is helpful to show how they can think more systematically about these issues and how different parts of their views connect with moral epistemology. While expressivists have been concerned with truth and correctness for moral judgments, the connection with epistemic evaluability has not

been properly established or explored. After all, the explicit articulation of the challenge of epistemic evaluability is very recent. The version of the challenge and responses I develop here should make the connection clear. Another example is the recent attention expressivists have devoted to moral explanations. The epistemic implications of those views have not been drawn, so I hope my accounts of evidence and reliability provide a helpful starting point.

I am sure expressivists need to and want to say more about these issues, and these do not exhaust everything that needs to be addressed for a complete moral epistemology —for example, what are the consequences of this picture for things like moral testimony or moral learning? But I hope that with what I offer here, it is clear that the idea of an expressivist moral epistemology is not based on a category mistake.

# Bibliography


- Ahlstrom-Vij, Kristoffer & Dunn, Jeff (2018). *Epistemic Consequentialism*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra. (2004). *Evidentiality*. Oxford University Press.
- Alston, William P. (1995). “How to Think about Reliability.” *Philosophical Topics* 23 (1):1-29.
- ---------- (2005). *Beyond “Justification”: Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.
- Anderson, Elizabeth, (2004). “Uses of value judgments in feminist social science: A case study of research on divorce.” *Hypatia* 19 (1):1-24.
- ---------- (2020) “Feminist Epistemology and Philosophy of Science”, *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/feminism-epistemology/>.
- Armour-Garb, Bradley, Daniel Stoljar, and James Woodbridge. (2021) "Deflationism About Truth", *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2023/entries/truth-deflationary/>.
- Arvan, Marcus (2019). “The Dark Side of Morality: Group Polarization and Moral Epistemology.” *Philosophical Forum* 50 (1):87-115.
- Audi, Robert (1999). “Self-evidence.” *Philosophical Perspectives* 13:205-228.
- ---------- (2002). “The sources of knowledge.” In Paul K. Moser (ed.), *The Oxford Handbook of Epistemology*. Oxford University Press. pp. 71--94.
- ---------- (2003) “Contemporary Modest Foundationalism” in Louis J. Pojman (ed.) *The Theory of Knowledge: Classical and Contemporary Readings*. Wadsworth.
- ---------- (2005). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton University Press.
- ---------- (2011). "Moral Knowledge," in *The Routledge Companion to Epistemology*, Bernecker & Pritchard (eds.). Routledge.

- ---------- (2018). “Understanding, Self-Evidence, and Justification.” *Philosophy and Phenomenological Research* 99 (2):358-381.
- Ayer, Alfred Jules (1936). *Language, Truth, and Logic*. Dover Publications.
- Baehr, J. (2019) “Virtue Epistemology, Virtue Ethics, and the Structure of Virtue” in Battaly, H. (Ed.), *Routledge Handbook of Virtue Epistemology*. Routledge.
- Bar-On, Dorit & Chrisman, Matthew (2009). “Ethical neo-expressivism.” In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics: Volume 4*. Oxford University Press.
- Beddor, Bob (2019). “Noncognitivism and Epistemic Evaluations.” *Philosophers’ Imprint* 19.
- ---------- (2020). “A solution to the many attitudes problem.” *Philosophical Studies* 177 (9):2789-2813.
- ---------- (Forthcoming). “Noncognitivism Without Expressivism.” *Philosophy and Phenomenological Research.*
- Bedke, Matt (2009). “The Iffiest Oughts: A Guise of Reasons Account of End-Given Conditionals.” *Ethics* 119.4: 672-698.
- ---------- (2008). “Ethical Intuitions: What They Are, What They Are Not, and How They Justify.” *American Philosophical Quarterly* 45 (3):253-270.
- ---------- (2013). “Ethics makes strange bedfellows: intuitions and quasi-realism.” In Matthew C. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?*_ Routledge.
- ---------- (2018). “Non-Descriptive Relativism: Adding Options to the Expressivist Marketplace.” Oxford Studies in Metaethics 13:48-70.
- Bergström, Lars (1990). *Grundbok i värdeteori.* Stockholm: Thales.
- Berker, Selim (2013). “Epistemic Teleology and the Separateness of Propositions.” *Philosophical Review* 122 (3):337-393.
- Bex-Priestley, Graham (2018). “Error and the Limits of Quasi-Realism.” *Ethical Theory and*
- *Moral Practice*, 21(5), 1051–63.
- Bex-Priestley, Graham & Gamester, Will (2023). “Sidestepping the Frege-Geach Problem.” Philosophical Quarterly.
- Björnsson, Gunnar (2001). “Why emotivists love inconsistency.” *Philosophical Studies* 104 (1):81 - 108.

- Björnsson, Gunnar & McPherson, Tristram (2014). “Moral Attitudes for Non-Cognitivists: Solving the Specification Problem.” *Mind* 123 (489):1-38.
- Blackburn, Simon (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford University Press.
- ---------- (1993) *Essays in Quasi-Realism,* New York: Oxford University Press.
- ---------- (1996). “Securing the nots: moral epistemology for the quasi-realist”. In Walter Sinnott-Armstrong Mark Timmons (ed.), *Moral Knowledge? New Readings in Moral Epistemology*. Oxford University Press.
- ---------- (1998a). *Ruling Passions: A Theory of Practical Reasoning*. Oxford University Press.
- ---------- (1998b). “Wittgenstein, Wright, Rorty and minimalism.” Mind 107 (425):157-181.
- ---------- (1999). “Is objective moral justification possible on a quasi-realist foundation?” Inquiry: An Interdisciplinary Journal of Philosophy 42 (2):213 – 227.
- ---------- (2001). “Reason, virtue, and knowledge.” In Abrol Fairweather & Linda Trinkaus Zagzebski (eds.), Virtue Epistemology: Essays on Epistemic Virtue and Responsibility. Oxford University Press.
- ---------- (2002). Précis of Ruling Passions. Philosophy and Phenomenological Research 65 (1):122-135.
- ---------- (2005). “Quasi-Realism no Fictionalism.” In Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford University Press.
- ---------- (2009). “Truth and A Priori Possibility: Egan’s Charge Against Quasi-Realism.” *Australasian Journal of Philosophy* 87 (2):201-213.
- ---------- (2017) “Moral epistemology for sentimentalists” in Stueber, Karsten & Debes, Remy (eds.) *Ethical Sentimentalism: New Perspectives*. Cambridge University Press.
- ---------- (Ms) “Sharon Street on The Independent Normative Truth as such.” <http://www2.phil.cam.ac.uk/~swb24/PAPERS/Meanstreet.htm> Accessed June 21, 2021.
- Boyd, Richard (1988). “How to be a Moral Realist.” In G. Sayre-McCord (ed.), *Essays on Moral Realism*. Cornell University Press.

- Broome, John (2004). “Reasons.” In R. Jay Wallace (ed.), *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*. Oxford University Press.
- ---------- (2013). *Rationality Through Reasoning*. Malden, MA: Wiley-Blackwell.
- Brown, James L. D. (2019). “Expressivism and Cognitive Propositions.” *Journal of the American Philosophical Association* 5 (3):371-387.
- ---------- (2021). “Interpretative expressivism: A theory of normative belief.” *Philosophical Studies* 179 (1):1-20.
- Brown, Jessica (2015). “Evidence and Epistemic Evaluation.” *Oxford Studies in Epistemology 5*. Oxford University Press.
- ---------- (2018). “Pragmatic Approaches to Belief”, in Conor McHugh, Jonathan Way, and Daniel Whiting (eds), *Normativity: Epistemic and Practical*. Oxford University Press.
- Bykvist, K. & Hattiangadi, A. (2013) “Belief, Truth and Blindspots.” In Chan, Timmothy (ed.), *The Aim of Belief*. Oxford University Press.
- Bykvist, Krister & Olson, Jonas (2009). “Expressivism and moral certitude.” *Philosophical Quarterly* 59 (235):202-215.
- ---------- (2012). “Against the Being For Account of Normative Certitude.” *Journal of Ethics and Social Philosophy* 6 (2):1-8.
- Campbell, Richmond, (1998). *Illusions of Paradox: A Feminist Epistemology Naturalized.* Rowman & Littlefield Publishers.
- ---------- (2015). "Moral Epistemology", *in The Stanford Encyclopedia of Philosophy* , Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/moral-epistemology/>.
- Carter, J. Adam & Chrisman, Matthew (2012). “Is epistemic expressivism incompatible with inquiry?” *Philosophical Studies* 159 (3):323-339.
- Chan, Timothy (ed.) (2013). *The Aim of Belief.* Oxford University Press.
- Chisholm, Roderick M. (1991). “Firth and the ethics of belief.” *Philosophy and Phenomenological Research* 51 (1):119-128.
- Chrisman, Matthew (2010). “Constructivism, Expressivism and Ethical Knowledge.” *International Journal of Philosophical Studies* 18 (3):331-353.
- ---------- (2011). “Ethical Expressivism”. In Christian Miller (ed.), *Continuum Companion to Ethics*. Continuum.

- ---------- (2012). “Epistemic Expressivism.” *Philosophy Compass* 7 (2):118-126.
- Clarke-Doane, Justin (2020). *Morality and Mathematics*. Oxford University Press.
- Conee, Earl and Richard Feldman, (1998). “The Generality Problem for Reliabilism”, *Philosophical Studies*, 89(1): 1–29. doi:10.1023/A:1004243308503
- ---------- (2004). *Evidentialism*. Oxford University Press.
- ---------- (2008). “Evidence.” In Q. Smith (ed.), *Epistemology: New Essays*. Oxford University Press, pp. 83–104.
- Cosker-Rowland, Rach (2017). “The Significance of Significant Fundamental Moral Disagreement.” Noûs, 51(4), 802–31.
- ---------- (2018). “The intelligibility of moral intransigence: A dilemma for cognitivism about moral judgment.” *Analysis* 78 (2):266-275.
- ---------- (2019a). *The Normative and the Evaluative: The Buck-Passing Account of Value*. Oxford University Press.
- ---------- (2019b). “Local Evolutionary Debunking Arguments.” *Philosophical Perspectives* 33 (1):170-199.
- Cowie, Christopher (2019). *Morality and Epistemic Judgement*. Oxford University Press.
- Cowie, Christopher & Cosker-Rowland, Rach (2019). *Companions in Guilt: Arguments in Metaethics.* Routledge.
- Crisp, Roger (2018). "Against Partiality" A Lindley Lecture at The University of Kansas. Recovered from: https://kuscholarworks.ku.edu/handle/1808/26747
- Cuneo, Terence (2020). “Can expressivism have it all?” *Philosophical Studies* 177 (1):219-241.
- Dancy, Jonathan (2021). “Necessity, Universality and the a priori in ethics” in *Practical Thought*. Oxford University Press.
- Dasgupta, Shamik (2017). “Constitutive Explanation.” *Philosophical Issues* 27 (1):74-97.
- David, Marian (2001). “Truth as the Epistemic Goal.” In M. Steup (ed.), *Knowledge, Truth, and Duty*. Oxford University Press.
- ---------- (2014). “Truth as the Primary Epistemic Goal: A Working Hypothesis.” In Matthias Steup, John Turri & Ernest Sosa (eds.),

*Contemporary Debates in Epistemology* (Second Edition). Wiley-Blackwell.

- DePaul, Michael and Amelia Hicks (2021). "A Priorism in Moral Epistemology", in *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/moral-epistemology-a-priori/>.
- DiFate, Victor (2007). "Evidence." *Internet Encyclopedia of Philosophy*.
- Diggin, Steven (2022). "Ethical Evidence." *Synthese* 200 (4):1-24.
- Dorr, Cian (2002). "Non-cognitivism and Wishful Thinking." *Noûs* 36 (1):97–103.
- Dreier, James (1996). 'Expressivist Embeddings and Minimalist Truth.' *Philosophical Studies* 83(1): 29-51.
- ---------- (2002). "The Expressivist Circle: Invoking Norms in the Explanation of Normative Judgment," Philosophy and Phenomenological Research 65, pp. 136–143.
- ---------- (2004). "Meta-ethics and the problem of creeping minimalism." *Philosophical Perspectives* 18 (1):23–44.
- ---------- (2006). "Negation for expressivists: A collection of problems with a suggestion for their solution." In Russ Shafer-Landau, editor, *Oxford Studies in Metaethics*. Oxford University Press.
- ---------- (2012). "Quasi-Realism and the Problem of Unexplained Coincidence." *Analytic Philosophy* 53 (3):269-287.
- ---------- (2015). "Explaining the Quasi-real." In R. Shafer-Landau (Ed.), Oxford Studies in Metaethics Vol. 10. Oxford: Oxford University Press.
- ---------- (2022) "The Normative Explanation of Normativity" in *Meaning, Decision and Norms: Themes from the Work of Allan Gibbard*, William Dunaway and David Plunkett (Eds.). Maize Books. University of Michigan.
- Dworkin, Ronald (1996). "Objectivity and Truth: You'd Better Believe it." *Philosophy and Public Affairs* 25 (2):87-139.
- Egan, Andy (2007). "Quasi-realism and fundamental moral error." *Australasian Journal of Philosophy* 85 (2):205 – 219.

- ---------- (2008). “Imagination, Delusion, and Self-Deception”, in *Delusion and Self-Deception: Affective Influences on Belief-formation*, Tim Bayne and John Fernandez (eds.). New York: Psychology Press.
- Elstein, Daniel Y. & Williams, Robert (Ms) “Suppositions, Revisions and Decisions.” Recovered from: https://philpapers.org/archive/ELSSAD-2.pdf
- Engel, P. (2005). ‘Truth and the aim of belief’, in D. Gillies (ed.), *Laws and Models in Science*. London: King’s College Publications.
- ---------- (2007). “Belief and normativity.” *Disputatio* 2 (23):179-203.
- Enoch, David (2009). “The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it.” *Philosophical Studies* 148 (3):413-438.
- Ezcurdia, Maite. (2004) "Conocimiento del lenguaje y actitudes de aceptación" [Knowledge of Language and Acceptance Attitudes], *Acta Poética*, Instituto de Investigaciones Filológicas, UNAM, México, XXV, no. 2.
- Fantl, Jeremy & McGrath, Matthew (2002). “Evidence, pragmatics, and justification.” *Philosophical Review* 111 (1):67-94.
- Faraci, David (2019). “Groundwork for an Explanationist Account of Epistemic Coincidence.” *Philosophers’ Imprint* 19.
- Feldman, R. (2002). “Epistemological duties.” In P. Moser (Ed.), The Oxford handbook of epistemology (pp. 362–384). Oxford: Oxford University Press.
- Field, Hartry H. (2001). *Truth and the absence of fact*. Oxford University Press.
- ---------- (2018). “Epistemology from an Evaluativist Perspective.” *Philosophers' Imprint* 18.
- Fletcher, Guy (2016). “Moral Testimony: Once More with Feeling.” *Oxford Studies in Metaethics* 11:45-73. Oxford University Press.
- Fletcher, Guy & Ridge, Michael (eds.) (2014). *Having It Both Ways: Hybrid Theories and Modern Metaethics*. New York: Oxford University Press.
- Flores Magon, Ricardo (1918). “Por la justicia” [For Justice]. Recovered from: https://archivomagon.net/en/obras-completas/discursos/d1918/
- Foley, Richard (1991). “Evidence and Reasons for Belief” *Analysis* 51 (2):98-102.
- Frápolli, María José & Villanueva, Neftalí (2012). “Minimal Expressivism.” Dialectica 66 (4):471-487.

- Friedman, Jane (2019) "Teleological epistemology" Philosophical Studies volume 176, pages673–691
- Fumerton, Richard A. (1995). *Metaepistemology and Skepticism*. Rowman & Littlefield.
- ---------- (2005). *Epistemology*. Blackwell.
- ---------- (2018). "Evidentialism and Explanatory Fit." In McCain Kevin (ed.), *Believing in Accordance with the Evidence*. Springer Verlag.
- Gamester, Will (2022). "Fallibility without Facts." *Ergo 8*: 40. doi: https://doi.org/10.3998/ergo.2241
- García-Ramírez, Eduardo (2018). "Towards Cognitive Moral Quasi-Realism." Philosophies 3 (1):5-0.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.
- ---------- (2002) "Knowing What to Do, Seeing What to Do." In Philip Stratton-Lake (ed.), *Ethical Intuitionism: Re-Evaluations*. Clarendon Press.
- ---------- (2003). *Thinking How to Live*. Harvard University Press.
- ---------- (2005), 'Truth and correct belief' Philosophical Issues, 15 Normativity, 338–50.
- ---------- (2008). *Reconciling our aims: in search of bases for ethics*. New York: Oxford University Press.
- ---------- (2011). "How Much Realism? Evolved Thinkers and Normative Concepts." Oxford Studies in Metaethics 6:33.
- ---------- (2012). *Meaning and Normativity*. Oxford University Press.
- ---------- (2015). "Improving Sensibilities" In R. Johnson and M. Smith (ed.s) Passions and Projections, Oxford University Press.
- Gilmore, Jonathan (2020). *Apt Imaginings: Feelings for Fictions and Other Creatures of the Mind*. Oxford University Press.
- Glüer, Kathrin & Wikforss, Åsa (2013). "Aiming at Truth: On The Role of Belief." *Teorema: International Journal of Philosophy* 32 (3):137-162.
- Glüer, Kathrin, Åsa Wikforss, and Marianna Ganapini. (2022) "The Normativity of Meaning and Content", *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), forthcoming URL = <https://plato.stanford.edu/archives/fall2023/entries/meaning-normativity/>.

- Goldman, Alvin I. (1979). “What Is Justified Belief?” in George S. Pappas (ed.), Justification and Knowledge: New Studies in Epistemology, Dordrecht: Reidel, pp. 1–25.
- ---------- (1993). Epistemic Folkways and Scientific Epistemology. Philosophical Issues 3:271-285.
- ---------- (1999). “A priori warrant and naturalistic epistemology” *Philosophical Perspectives* 13:1-28.
- ---------- (2009). “Internalism, Externalism, and the Architecture of Justification.” Journal of Philosophy 106: 309–38.
- ---------- (2012). *Reliabilism and Contemporary Epistemology: Essays.* Oxford: Oxford University Press.
- Goldman, Alvin & Beddor, Bob (2021) “Reliabilist Epistemology”, The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2021/entries/reliabilism/>.
- Golub, Camil (2017). “Expressivism and the Reliability Challenge.” Ethical Theory and Moral Practice 20 (4):797-811.
- Graham, Peter J. (2017). “Normal Circumstances Reliabilism: Goldman on Reliability and Justified Belief.” *Philosophical Topics* 45 (1):33-61.
- Greco, John & Reibsamen, Jonathan (2018). Reliabilist Virtue Epistemology. In Nancy Snow (ed.), The Oxford Handbook of Virtue. Oxford University Press.
- Gregory, Alex (2012). “Changing Direction on Direction of Fit.” *Ethical Theory and Moral Practice* 15 (5):603-614.
- ---------- (2013). “The Guise of Reasons.” *American Philosophical Quarterly* 51.1.
- ---------- (2021). *Desire as Belief*. Oxford University Press.
- Hayward, Max Khan (2017). “Practical Reason, Sympathy and Reactive Attitudes.” *Noûs*:51-75.
- Hills, Alison (2013). “Moral Testimony.” *Philosophy Compass* 8 (6):552-559.
- Horgan, Terry & Timmons, Mark (2006). “Cognitivist expressivism.” In Terry Horgan & Mark Timmons (eds.), *Metaethics After Moore*. Oxford University Press.

- ---------- (2015). “Modest Quasi-Realism and the Problem of Deep Moral Error” In Robert Johnson and Michael Smith (Eds.), Passions and Projections Oxford University Press.
- Horwich, Paul (2010). *Truth – Meaning – Reality*. Oxford University Press.
- Howard, Chris & Cosker-Rowland, Rach (2022). “Fittingness: A User’s Guide.” In Chris Howard & Rach Cosker-Rowland (eds.), *Fittingness*. Oxford University Press.
- Huemer, M. (2005). *Ethical Intuitionism.* Palgrave Macmillan.
- Jenkins, C. I. (2006). “Knowledge and Explanation.” *Canadian Journal of Philosophy* 36 (2):137-164.
- ---------- (2008). “A priori knowledge: Debates and developments.” Philosophy Compass 3 (3):436–450.
- ---------- (2015). “What Quasi-Realists Can Say About Knowledge.” In R. Johnson and M. Smith (eds.) *Passions and Projections*, Oxford University Press.
- Joyce, Richard (2018). "The Denial of Moral Knowledge" in Zimmerman, Aaron ; Jones, Karen & Timmons, Mark (eds.)Routledge Handbook on Moral Epistemology. New York: Routledge.
- Kearns, Stephen & Star, Daniel (2008). “Reasons: Explanations or Evidence.” *Ethics* 119 (1):31-56.
- ---------- (2009). Reasons as Evidence. Oxford Studies in Metaethics 4:215-42.
- Kelly, Thomas (2013) “Evidence Can Be Permissive.” In Contemporary Debates in Epistemology. 2d ed. Edited by Matthias Steup, John Turri, and Ernest Sosa, 298–312. Malden, MA: Wiley-Blackwell.
- ---------- (2020) “Evidence”, The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/evidence/>.
- ---------- (2022). *Bias: A Philosophical Study*. Oxford University Press.
- Kelp, Christoph (2021). *Inquiry, Knowledge, and Understanding*. Oxford University Press.
- King, Jeffrey C. (2014). “What role do propositions play in our theories?” In Jeffrey C. King, Scott Soames & Jeff Speaks (eds.), *New Thinking About Propositions*. Oxford University Press.
- Kitcher, P. (1990). ‘The Division of Cognitive Labor’. *The Journal of Philosophy* 87: 5–22.

- Klenk, Michael (ed.) (2020). Higher Order Evidence and Moral Epistemology. New York: Routledge.
- Köhler, Sebastian (2015). “What Is the Problem with Fundamental Moral Error?” *Australasian*
- *Journal of Philosophy*, 93(1), 161–65.
- ---------- (2017). “Expressivism, Belief, and All That.” *Journal of Philosophy* 114 (4):189-207.
- ---------- (2022). “What is (Neo-)Pragmatists’ Function?” *Australasian Journal of Philosophy*:1-17.
- Korman, Daniel Z. (2019). “Debunking arguments.” *Philosophy Compass* 14 (12):e12638.
- Korman, Daniel Z. & Locke, Dustin (2020). “Against Minimalist Responses to Moral Debunking Arguments.” *Oxford Studies in Metaethics* 15:309-332.
- Kung, Peter (2016), “Imagination and Modal Knowledge”, in Kind, Amy and Peter Kung (eds.), 2016a, *Knowledge Through Imagination*. Oxford University Press.
- Kvanving, Jonathan. (2003). *The Value of Knowledge and the Pursuit of Understanding.* Cambridge: Cambridge University Press.
- ---------- (2014). “Epistemic Appraisal” in *Rationality and Reflection*. Oxford University Press.
- Kyriacou, Christos. 2017. “Expressivism, question substitution and evolutionary debunking.” *Philosophical Psychology*, 30(8): 1019-1042.
- Lauria, Federico & Deonna, Julien (eds.) (2017). *The Nature of Desire*. Oxford University Press.
- Lenman, James (2003a). “Disciplined syntacticism and moral expressivism.” *Philosophy and Phenomenological Research* 66 (1):32–57..
- ---------- (2003b) “Noncognitivism and the Dimensions of Evaluative Judgement: A Review of Michael Smith: “Evaluation, Uncertainty and Motivation”” in BEARS: Brown Electronic Article Review Service
- ---------- (2007). Expressivism and epistemology: What is moral inquiry? Aristotelian Society Supplementary Volume 81 (1):63–81.
- ---------- (2018) "The Primacy of the Passions" in Anna Bergqvist and Robert Cowan (eds.): *Evaluative Perception*. Oxford University Press.
- Lewis, David K. (1974). “Radical interpretation.” *Synthese* 23 (July-August):331-344.

- ---------- (1976). “Probabilities of conditionals and conditional probabilities.” *Philosophical Review* 85 (3):297-315.
- ---------- (1980). “A subjectivist’s guide to objective chance.” In Richard C. Jeffrey (ed.), *Studies in Inductive Logic and Probability, Volume II*. Berkeley: University of California Press. pp. 263-293.
- ---------- (1986). “Causal explanation.” In Philosophical Papers Vol. II. Oxford University Press.
- Lee, Marcus; Robson, Jon & Sinclair, Neil (2021). “Moral Testimony as Higher Order Evidence.” In Michael Klenk (ed.), *Higher-Order Evidence and Moral Epistemology*. Routledge.
- Liao, Shen-yi and Tamar Gendler, (2019). "Imagination", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/imagination/>.
- Littlejohn, Clayton (2018). “The Right in the Good: A Defense of Teleological Non-Consequentialism in Epistemology.” In Kristoffer Ahlstrom-Vij Jeff Dunn (ed.), *Epistemic Consequentialism*. Oxford University Press.
- ---------- (2018). “Evidence and Its Limits” in Normativity: Epistemic and Practical, McHugh, Conor; Way, Jonathan & Whiting, Daniel (eds.). Oxford University Press.
- Loev, S. (2022) “Affectivism about intuitions.” Synthese 200, 274.
- Longino, Helen E. (ed.) (1990). Science as Social Knowledge: Values and Objectivity in Scientific Inquiry. Princeton University Press.
- Lord, Errol (2018). Epistemic Reasons, Evidence, and Defeaters. In Daniel Star (ed.), The Oxford Handbook of Reasons and Normativity. Oxford University Press.
- ---------- (Forthcoming). “Evidence and Epistemic Reasons” in The Routledge Handbook of Evidence, edited by Maria Lasonen-Aarnio and Clayton Littlejohn. Routledge.
- Lynch, Michael P. (2009). *Truth as One and Many*. Oxford University Press.
- Lyons, J. (2013). “Should Reliabilists Be Worried about Demon Worlds?” Philosophy and Phenomenological Research 86: 1–40.
- Machuca, Diego E. (2018). “Moral Skepticism: An Introduction and Overview.” In Diego E. Machuca (ed.), Moral Skepticism: New Essays. New York: Routledge.

- MacAskill, William ; Bykvist, Krister & Ord, Toby (2020). *Moral Uncertainty*. Oxford University Press.
- Maguire, Barry (2017). "The Autonomy of Ethics." In Tristram McPherson & David Plunkett (eds.), *The Routledge Handbook of Metaethics*. Routledge.
- Maguire, Barry & Woods, Jack (2020). "The Game of Belief." *Philosophical Review* 129 (2):211-249.
- Mares, Edwin (2011). *A Priori*. Routledge.
- McCain, Kevin (2014). *Evidentialism and Epistemic Justification*. New York: Routledge.
- McGrath, Sarah (2019) *Moral Knowledge*. Oxford University Press.
- McHugh, Conor. (2011). "What do we aim at when we believe?" *Dialectica*, 65(3), 369–392.
- McHugh, Conor & Whiting, Daniel (2014). "The Normativity of Belief" *Analysis* 74 (4):698-713.
- McKeever, Sean D. & Ridge, Michael (2006). "Particularism and the contingent a priori." *Acta Analytica* 21 (2):3-11.
- McPherson, Tristram (2013) "A priori ethical knowledge" in International Encyclopedia of Ethics, Hugh LaFollette (Ed). John Wiley & Sons, Ltd.
- ---------- (2018) "Naturalistic Moral Realism, Moral Rationalism, and Non-Fundamental Epistemology" in The Many Moral Rationalisms. Karen Jones and Francois Schroeter (Eds.). Oxford University Press
- ---------- (2020). *Epistemology and Methodology in Ethics*. Cambridge University Press.
- ---------- (2022). "Expressivism without Minimalism" In Billy Dunaway & David Plunkett (eds.), Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard. Maize Books.
- Mele, Alfred (2010). "Teleological Explanations of Actions: Anticausalism vs. Causalism." In J. Aguilar & A. Buckareff (eds.), Causing Human Actions: New Perspectives on the Causal Theory of Action. MIT Press.
- Melis, G. (2014). "Understanding Undermining Defeat." *Philosophical Studies*, 170(3), 433-442.
- Mercier, Hugo (2011). "What good is moral reasoning?" Mind and Society 10 (2):131-148.

- Mercier, Hugo & Sperber, Dan (2011) Why do humans reason? Arguments for an argumentative theory. Behavioral and Brain Sciences. Cambridge University Press
- ---------- (2017) The Enigma of Reason. Harvard University Press.
- Milona, Michael & Schroeder, Mark (2019). "Desiring under the Proper Guise." *Oxford Studies in Metaethics* 14:121-143.
- Moretti, Luca & Piazza, Tommaso (2018). Defeaters in current epistemology: introduction to the special issue. Synthese 195 (7):2845-2854.
- Morrison, Joe (2017). "Evidential holism." *Philosophy Compass* 12 (6):e12417.
- Moss, Sarah (2016). *Probabilistic Knowledge*. Oxford University Press.
- Nagel, Jennifer (2017). "Factive and nonfactive mental state attribution." *Mind and Language* 32 (5):525-544.
- Nolfi, Kate (2015). "Which Mental States Are Rationally Evaluable, And Why?" Philosophical Issues 25 (1):41-63.
- Noonan, Christopher (Forthcoming). "Evolutionary Debunking Arguments, Explanationism and Counterexamples to Modal Security." *Erkenntnis*:1-23.
- Orsi, Francesco (2015). "The Guise of the Good." *Philosophy Compass* 10 (10):714-724.
- Ortiz-Millán, Gustavo (2018) "Wrong Direction" in Mind, Language and Morality. Essays in Honor of Mark Platts, edited by G. Ortiz-Millán and J.A. Cruz Parcero, London-New York: Routledge.
- Parfit, Derek (2011). *On What Matters: Volume One.* Oxford University Press.
- Peacocke, Christopher (2004). *The Realm of Reason*. Oxford University Press.
- Pettigrew, Richard (2016). Accuracy and the Laws of Credence. Oxford University Press.
- Pettit, Philip & Smith, Michael (1990). Backgrounding desire. Philosophical Review 99 (4):565-592.
- Piazza, Tommaso (Forthcoming). Epistemic Defeaters. Routledge Encyclopedia of Philosophy.
- Pollock, John (1986). *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.
- Pritchard, Duncan (2018). "Moral Knowledge" in *What is this thing called Knowledge?* Routledge.

- Pryor, James (2000). “The skeptic and the dogmatist.” *Noûs* 34 (4):517–549.
- Quine, W. V. O. (1953). ‘Two Dogmas of Empiricism’ in From a Logical Point of View. Cambridge, MA: Harvard UP.
- Raskoff, Sarah Zoe (2018). “Getting Expressivism Out of the Woods.” *Ergo: An Open Access Journal of Philosophy* 5.
- Rayo, Agustín (2014) “An Anti-Reductionist’s Guide to Evidential Support”. Retrieved from: http://arayo.scripts.mit.edu/home/portfolio-archive/an-anti-reductionists-guide-to-evidential-support/
- Ridge, Michael (2007). “Epistemology for Ecumenical Expressivists.” *Aristotelian Society Supplementary Volume* 81 (1):83-108.
- ---------- (2009). “Moral assertion for expressivists.” In Ernest Sosa & Enrique Villanueva (eds.), *Metaethics*. Wiley Periodicals.
- ---------- (2014). *Impassioned Belief*. Oxford University Press.
- ---------- (2015). “I Might Be Fundamentally Mistaken.” *Journal of Ethics and Social Philosophy*, 9(3), 1–22. https
- ---------- (2018) “How to Be an Epistemic Expressivist” in Metaepistemology, McHugh, Conor; Way, Jonathan & Whiting, Daniel. Oxford University Press.
- ---------- (2020). “Normative certitude for expressivists.” *Synthese* 197 (8):3325-3347.
- Rinard, Susanna (2018). “Believing for Practical Reasons.” *Noûs* (4):763-784.
- ---------- (2019). “Equal treatment for belief.” *Philosophical Studies* 176 (7):1923-1950.
- Rosen, Gideon (2010). “Metaphysical Dependence: Grounding and Reduction.” In Bob Hale & Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford University Press.
- Ross, David (2002). *The Right and the Good*. Oxford University Press.
- Sayre-McCord (1996). “Coherentist Epistemology and Moral Theory.” In Walter Sinnott-Armstrong & Mark Timmons (eds.), Moral Knowledge? New Readings in Moral Epistemology. Oxford University Press.
- ---------- (2013) "Moral Epistemology,” in *The International Encyclopedia of Ethics*, Hugh La Follette (ed.). Blackwell.
- Scanlon, Thomas (2014). Being Realistic About Reasons. Oxford: Oxford University Press.

- Schellenberg, Susanna (2013). “Belief and Desire in Imagination and Immersion.” *Journal of Philosophy* 110 (9):497-517.
- Schroeder, Mark (2007). *Slaves of the passions*. Oxford University Press.
- ---------- (2008a). *Being for: evaluating the semantic program of expressivism*. Oxford University Press.
- ---------- (2008b). “What is the Frege-Geach problem?” *Philosophy Compass* 3 (4):703-720.
- ---------- (2010). How not to avoid wishful thinking. In Michael Brady (ed.), New Waves in Metaethics. Palgrave-Macmillan.
- ---------- (2013) ‘Two Roles for Propositions: Cause for Divorce?’ Noûs 47(3): 409-430
- ---------- (2014). “Does expressivism have subjectivist consequences?” *Philosophical Perspectives* 28 (1):278-290.
- ---------- (2018). “The Moral Truth.” In Michael Glanzberg (ed.), *Oxford Handbook of Truth*. Oxford University Press.
- Schueler, G. F. (2013). “Direction of Fit.” In Huge LaFollette (ed.), *The International Encyclopedia of Ethics*. Blackwell.
- Schwitzgebel, Eric, (2019). “Belief” in *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/belief/>.
- Sehon, Scott (2010). “Teleological Explanation.” In Timothy O'Connor & Constantine Sandis (eds.), A Companion to the Philosophy of Action. Oxford, UK: Wiley-Blackwell. pp. 121–128.
- Sepielli, Andrew (2012). “Normative uncertainty for non-cognitivists.” *Philosophical Studies* 160 (2):191-207.
- Setiya, Kieran (2012). *Knowing Right from Wrong*. Oxford University Press.
- Shafer-Landau, R., (2005). *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shah, Nishi (2006). “A new argument for evidentialism.” *Philosophical Quarterly* 56 (225):481–498.
- Shah, Nishi and David Velleman. (2005). ‘Doxastic Deliberation.’ *Philosophical Review* 114: 497–534.
- Shieh, Sanford (2018). “Truth, Objectivity, and Realism”. In Michael Glanzberg (ed.), Oxford Handbook of Truth. Oxford University Press
- Silk, Alex (2015). “How to be an ethical expressivist.” *Philosophy and Phenomenological Research,* 91(1): 47–81.

- Sinclair, Neil (2017). “Belief pills and the possibility of moral epistemology.” In Oxford Studies in Metaethics. Oxford, UK: Oxford University Press.
- ---------- (2018). “Conceptual Role Semantics and the Reference of Moral Concepts.” *European Journal of Philosophy* 26 (1):95-121.
- ---------- (2021). *Practical Expressivism*. Oxford University Press.
- Sinclair, Neil & Chamberlain, James (Forthcoming). “The Evolutionary Debunking of Quasi-Realism.” In Diego E. Machuca (ed.), *Evolutionary Debunking Arguments: Ethics, Philosophy of Religion, Philosophy of Mathematics, Metaphysics, and Epistemology*. London: Routledge.
- Sinhababu, Neil (2013). “Distinguishing Belief and Imagination.” *Pacific Philosophical Quarterly* 94 (2):152-165.
- ---------- (2015). “Advantages of Propositionalism.” *Pacific Philosophical Quarterly* 96 (1):165-180.
- ---------- (2017). *Humean Nature*. Oxford: Oxford University Press.
- ---------- (2022). “The reliable route from nonmoral evidence to moral conclusions.” *Erkenntnis: 1-21*.
- Smith, Michael (1994). *The Moral Problem*. Blackwell.
- ---------- (2002). “Evaluation, uncertainty, and motivation.” *Ethical Theory and Moral Practice*, 5, 305–320.
- Smithies, Declan (2022). “Moral Knowledge By Deduction.” *Philosophy and Phenomenological Research.*
- Soames, Scott (2015). *Rethinking Language, Mind, and Meaning*. Princeton University Press.
- Sobel, David & Copp, (2001). “Against direction of fit accounts of belief and desire.” *Analysis* 61 (1):44-53.
- Sosa, Ernest (2009). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford University Press.
- ---------- (2021). *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*. Oxford: Oxford University Press.
- Stalnaker, Robert (1984). *Inquiry*. MIT Press.
- Sterelny, Kim (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Wiley-Blackwell.
- Steup, Matthias (2018). “Is Evidential Fit Grounded in Explanatory Relations?” In McCain Kevin (ed.), *Believing in Accordance with the Evidence*. Springer Verlag.

- Stevenson, Charles. (1944). Ethics and Language. Yale University Press.
- Street, Sharon (2011). "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have it Both Ways." In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* 6. Oxford University Press.
- Sturgeon, Nicholas (2002). "Ethical Intuitionism and Ethical Naturalism." In Philip Stratton-Lake (ed.), *Ethical Intuitionism: Re-evaluations*. Oxford University Press.
- Sturgeon, Scott. (2014). "Pollock on defeasible reasons." *Philosophical Studies*, 169(1), 105-118.
- Tenenbaum, Sergio (2007). *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge University Press.
- ---------- (2008). "Appearing Good: A Reply to Schroeder." *Social Theory and Practice* 34: 131–138.
- ---------- (2020) "The guise of the good" in *The Routledge Handbook of Practical Reason*, Ruth Chang & Kurt Sylvan (Eds.). Routledge.
- Thomasson, Amie. (2020). *Norms and Necessity*. Oxford University Press.
- Toppinen, Teemu (2015). "Expressivism and the Normativity of Attitudes." Pacific Philosophical Quarterly 96 (2):233-255.
- ---------- (2017). "Enkrasia for Non-Cognitivists." *Ethical Theory and Moral Practice* 20 (5):943-955.
- Väyrynen, Pekka (Forthcoming). "Varieties of Normative Explanation." In David Copp & Connie Rosati (eds.), *The Oxford Handbook of Metaethics*. Oxford University Press.
- Velleman, David (2000). "On the Aim of Belief" in *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Vogel, Jonathan (2012). "The Enduring Trouble with Tracking" in Kelly Becker and Tim Black (eds) *The Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press.
- Wedgwood, Ralph (1999). "The a priori rules of rationality." *Philosophy and Phenomenological Research* 59 (1):113-131.
- ---------- (2002), "The Aim of Belief", Philosophical Perspectives, 16 , 267–97.
- ---------- (2007). *The Nature of Normativity*. Oxford University Press.
- ---------- (2023). "The Pitfalls of 'Evidence'" in *Rationality and Belief*. Oxford University Press.

- Williams, J. Robert G. (2019). *The Metapthysics of Representation*. Oxford: Oxford University Press.
- Williamson, T. (1996). “Knowing and Asserting.” The Philosophical Review 105: 489–523.
- ---------- (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- ---------- 2016, “Knowing by Imagining”, in Kind, Amy and Peter Kung (eds.), 2016a, *Knowledge Through Imagination*, New York: Oxford University Press
- Woodfield, Andrew (2017). Teleological Explanation. In W. H. Newton-Smith (ed.), A Companion to the Philosophy of Science. Oxford, UK: Blackwell. pp. 492–494.
- Woods, Jack (2017). “The Frege-Geach Problem”. In Tristram McPherson & David Plunkett (eds.), *The Routledge Handbook of Metaethics*. Routledge.
- Wright, Crispin (1992). *Truth and Objectivity*. Harvard University Press.
- ---------- (2002a). “(Anti-)sceptics simple and subtle: G. E. Moore and John McDowell.” *Philosophy and Phenomenological Research*, 65(2), 330–348.
- ---------- (2002b). “What Could Antirealism about Ordinary Psychology Possibly Be?” The Philosophical Review 111, pp. 205–233.
- ---------- (2021) “Minimalism, Deflationism, Pragmatism, Pluralism” in Lynch, Wyatt, Kim, & Kellen (Eds.) *The Nature of Truth (Second Edition)*. The MIT Press.
- Yalcin, Seth (2011). “Nonfactualism about epistemic modality.” In Andy Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.
- ---------- (2019). “Expressivism by force.” In D. Fogal, D. Harris & M. Moss (eds.), *New Work on Speech Acts*. Oxford University Press
- Zimmerman, Aaron (2010). *Moral epistemology*. New York: Routledge.
- Zimmerman, Aaron ; Jones, Karen & Timmons, Mark (eds.) (2018). *Routledge Handbook on Moral Epistemology*. New York: Routledge.