A Hybrid Theory of Ethical Thought and Discourse

Drew Johnson, PhD

University of Connecticut, 2022

Abstract: What is it that we are doing when we make ethical claims and judgments, such as the claim that we morally ought to assist refugees? This dissertation introduces and defends a novel theory of ethical thought and discourse. I begin by identifying the surface features of ethical thought and discourse to be explained, including the realist and cognitivist (i.e. belief-like) appearance of ethical judgments, and the apparent close connection between making a sincere ethical judgment and being motivated to act on it. I examine prominent attempts to explain these features, with a focus on recent 'hybrid' theories combining elements of expressivism and cognitivism. Despite their initial promise, I argue that extant hybrid theories are nevertheless committed to problematic semantic, metasemantic, or pragmatic assumptions. I then discuss what I take to be the strongest existing option, ethical neo-expressivism (Bar-On and Chrisman, 2009; Bar-On, Chrisman, and Sias, 2014), and develop it into an explicitly hybrid theory proposing that ethical judgments incorporate both motivationally-charged affective states and moral beliefs. I then supplement this account with a theory of the proper function (following Millikan, 1984) of ethical claims and judgments, arguing that they function simultaneously to *track* the morally salient features of social situations, and to *coordinate* our behavior around these features. Finally, I defend one of the cognitivist commitments of the theory – namely, that objective moral knowledge is possible – by applying recent work on the epistemology of fundamental or 'core' intellectual commitments (Lynch, 2012; Pritchard, 2016) to the moral realm.

A Hybrid Illocutionary Act Theory of Ethical Thought and Discourse

Drew Johnson

B.A., Grinnell College, 2013

M.A., Northern Illinois University, 2015

M.A., University of Connecticut, 2017

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2022

A Hybrid Theory of Ethical Thought and Discourse

Doctor of Philosophy Dissertation

A Hybrid Theory of Ethical Thought and Discourse

Presented by

Drew Johnson, B.A., M.A.

Approved by:

Major Advisor: Dorit Bar-On

Associate Advisor: Michael P. Lynch

Associate Advisor: Paul Bloomfield

Associate Advisor: William Lycan

University of Connecticut
2022

Acknowledgements

Writing a dissertation, I have been warned, involves long periods of solitude and isolation. I have found this to be true. Paradoxically, however, I have found that writing a dissertation is also an essentially social enterprise. This dissertation could not have been completed without many kinds of help from many people.

I would first like to thank the communities and institutional spaces I have been fortunate enough to be part of in my time at UConn: the UConn philosophy department, the *Expression, Communication, and Origins of Meaning* research group, and the Humanities Institute. This dissertation work has benefitted from these groups in very tangible ways; comments received, questions raised, and conversations pursued have shaped this work for the better. It has also benefitted from other kinds of support by these groups. To all the members of the UConn philosophy community, thank you for providing a unique and rich intellectual environment that pushed me to grow as a philosopher. Thanks to Don Baxter and Lewis Gordon for their support of my research as department heads. There are many things in Aristotle's work that I disagree with, but it rings true that certain material conditions – financial, institutional, educational – must be met to be able to pursue the value of philosophical research. The department has gone above and beyond to provide these conditions for myself and other graduate students.

Thanks to the professors I've had while studying at UConn, who helped shape me as a philosopher, including Jc Beall, Lewis Gordon, Mitch Green, Hallie Liberto, Lionel Shapiro, and Keith Simmons. Thanks also to Duncan Pritchard, who encouraged me to submit my first paper for publication (it was a success!), and whose seminar on radical skepticism given at UConn has greatly influenced my thinking ever since. Special thanks are also due to Ruth Millikan, who has been an inspiration and guiding light, both as a philosopher and as a person.

Many thanks are owed as well to all the (former and current) UConn graduate students with whom I have had discussions over the years, including Teresa Allen, Nimra Asif, Eric Berg, Kristin Culbertson, Hanna Gunn, Michael Hegarty, Jared Henderson, Mengyu Hu, Ken Ito, Nathan Kellen, Junyeol Kim, Yuhan Liang, Heather Muraviov, Chris Rhawles, Jenelle Salisbury, Ryo Tanaka, Taylor Tate, Cody Turner, and Katrina Van Dyke. Special thanks to Ahmad Jabbar, who has been an ideal research partner on formal aspects of expressivism; I've learned so much from our discussions.

The *Expression, Communication, and Origins of Meaning* research group has supported my work in several ways: in providing funding support for research, for hosting reading groups that have informed this work, and for putting me in conversation with philosophers and academics from other disciplines at UConn and beyond. The Humanities Institute, as well, has supported my work; the Research Fellowship I have been fortunate to have has provided important time and space for writing, research, and maintaining some sanity. The community of Humanities fellows has also provided a rich intellectual environment. I have learned much from my fellow fellows, and appreciate your patience and feedback as I worked out ideas for this dissertation during our writing groups, research roundtables, and at my fellows talk. Special thanks to Erik Freeman, Carol Gray, Fiona Vernal, Shiloh Whitney, and Anna Ziering for bearing the frustrating conceptual messes I sent you all for our writing group meetings. It is amazing to watch you all transform the mess into something coherent (or at least telling me how to make it coherent). Thanks also to Prakash Kashwan, whose fierce intellect and bold questions challenged me to think about the broader implications of this project.

My journey in philosophy began during my second year as an undergraduate at Grinnell college, when I found myself in disagreement with, but immensely engaged by, Plato's views on censorship and education in the *Republic* – I suppose this is fitting given that moral and political disagreement plays such a large role in the dissertation. Thanks are owed to Tammy Nyden for sparking my passion for philosophy. I would also like to thank several other of my professors at Grinnell: Joe Cummins, Johanna Meehan, Joe Neissor, and Alan Schrift. Special thanks go to my undergraduate capstone project advisor, John Fennell, who got me started thinking about Wittgenstein's *On Certainty*. I haven't stopped thinking about it.

Thanks are owed to those who helped along the next step of my philosophical journey, at the master's program at Northern Illinois University. Thanks to David Buller, who got me started thinking about truth. Lenny Clapp instilled in me a great appreciation for philosophy of language and showed me what it looks like to have an absolute blast at the front of a classroom – I strive to imbue that energy in my own teaching. Mylan Engel taught me much of what I know about epistemology and spurred my moral commitment to veganism (along with Jason Farr and Rodrigo Narbona). Thanks also to Valia Allori, Steven Daskal, Alicia Finch, Jason Hanna, and Geoff Pynn. I would also like to thank the close community of grad students I found at NIU, where I first had a group of friends who wanted to talk (incessantly) about philosophy until 2a.m. in the 6-bedroom run-down house we rented. Thanks especially to Ali Aenehzodaee, Teresa Allen (again), Jason Farr, Griffin Klemick, Rodrigo Narbona, Cole Rathjen, and Tom Shea.

In my committee I have found a patient and insightful group of interlocutors. To Paul Bloomfield, thank you for your frank advice, for always asking the tough questions, and for encouraging intellectual perseverance. Thanks to Bill Lycan for careful (yet impressively fast) comments on chapter drafts, from which I always gain a new way of looking at the issues. Michael Lynch provided much-needed perspective and mentorship while on the job market.

Michael's comments have always forced me to think about the big philosophical issues, and not to lose the forest for the trees.

I quite simply couldn't have asked for a better dissertation advisor than Dorit Bar-On. I think Dorit has somehow seen more drafts of my papers and chapters than I have (at any rate, it is surely in the hundreds by now), and always provides prompt and detailed comments. Our hours of phone conversations and back-and-forth drafts on various projects have been the impetus for so much of my growth as a philosopher. I struggle to find an apt linguistic vehicle to a-express my gratitude.

Finally, I would like to thank my parents, Hannah and Glenn, and my brother Nick, for supporting my decision to pursue philosophy, especially when it all seemed overwhelmingly difficult. And to Emily: I could not have completed this project without your love and support. Thank you for helping me learn when to put away the books and turn to more important things.

A Hybrid Theory of Ethical Thought and Discourse

Table of Contents

Chapter 1

The Features of Ethical Thought and Discourse

1.1 Introduction

What is it that we are doing when we make ethical claims (for instance, if I were to utter “it is morally wrong to torture prisoners”)?[1] Is making an ethical claim a way of stating or describing a portion of objective reality – the moral part of the world (if there is such a thing)? Or do I instead describe an aspect of my subjective worldview, such as the fact that *I disapprove* of torturing prisoners? Or, alternatively, does such a moral claim function, not to describe or state anything at all, but rather to express a certain attitude (such as approval or disapproval) of mine? Answering questions like these is the main goal of this dissertation. The dissertation thus engages with one of the core topics in metaethics over the past century, concerning the meaning of moral sentences and claims, and the contents of moral judgment. I propose a novel account of ethical thought and discourse, according to which our ethical claims and judgments have the function of coordinating our behavior in a limited range of ways around features of moral concern. As I go on to discuss, this proposal incorporates insights from and improves upon several competing families of metaethical theory, including cognitivism, moral realism, expressivism, and prescriptivism.

This introductory chapter is devoted to three main purposes. The first is to lay out my methodology for delineating the target for analysis – i.e. what distinguishes ethical thought and discourse from other domains. The second is to identify the surface features of ethical thought and discourse, features which an adequate metaethical theory should either explain or explain

[1] In this dissertation, I shall use ‘ethics’ and ‘morality’ interchangeably, with a slight preference for ‘ethics’.

away as merely apparent. And third, the chapter will briefly introduce some of the main categories of metaethical theory and their relative successes and challenges, to set up the discussion in Chapter 2 of expressivist theories.

1.2 Ethical claims and judgments

The stated goal of this dissertation is to provide a theory of ethical thought and discourse. In order to accomplish this goal, we must first characterize the target phenomenon to be explained. We thus face a preliminary question about how to individuate domains of thought and discourse. What are the core features of ethical claims and judgments, that sets them aside from claims and judgments in other areas? A starting point will be to examine the claims and judgments we are pre-theoretically inclined to categorize as belonging to the ethical domain and identify any features they appear to share and that claims and judgments in other domains lack. This provides a criterion for assessing the plausibility of metaethical theories: It is a mark in favor of a theory if it accommodates and explains the surface features of ethical claims and judgments, and it is a mark against a theory if it does not.[2]

One way in which we might individuate different areas of thought and discourse is according to their subject matter. What makes a judgment a *scientific* one, rather than an aesthetic one, is what the judgment is *about*. In turn, what makes a judgment about a scientific matter, say, chemistry, the thought continues, is that it centrally employs concepts of chemical kinds; that is, concepts that refer to properties that are metaphysically alike in respects of interest

[2] In this respect, my methodological approach is the standard one used in metaethics. For some recent examples, se See Chrisman (2017) and Enoch (2011). Enoch construes metaethics as proceeding by tallying up the ‘plausibility points’ of various theories. Chrisman describes the discipline as proceeding by comparing the plausibility of the various commitments of each theory. Other introductory texts in metaethics (such as van Roojen, 2015, and Schroeder, 2010) proceed in similar fashion, articulating a set of core questions or puzzles in metaethics, and portraying competing theories as aiming to best address these puzzles while avoiding implausible commitments in other areas of philosophy.

to chemists. The assumption here, in accordance with a naturalist philosophical stance, is that the natural world is organized into discrete groupings of properties, which it is the job of our concepts to track.[3] I will call this a 'bottom-up' approach to domain individuation, because ultimately it is the natural contours of the ontology of the world that constrain domain membership for our judgments. The sorts of properties studied by chemists, we can assume, are metaphysically similar to each other in a number of respects and dissimilar to the sorts of properties studied by e.g., psychologists, and understanding these natural divisions – the joints of the world – is perspicuously accomplished by grouping our concepts into different domains.[4] Applied to the moral case, according to this approach, we identify the category of ethical judgment by a criterion admitting any judgment that substantially employs a moral concept, where moral concepts are such in virtue of picking out moral properties, i.e. properties such as *moral rightness*, that are naturally clustered together.

There are two concerns for such an approach relevant here. The first concern is that the approach is especially contentious when we consider the moral domain, for it is a challenge to explain how moral facts and properties could exist from a naturalist philosophical perspective. If, as some have thought, there are no genuinely moral facts as part of the natural world, the moral domain would seem to lack a subject matter, frustrating the very attempt to delineate ethical thought and discourse from other domains. However, I set this concern aside for now, because even contemporary moral anti-realists concede that moral concepts at least appear on their surface to operate much like clearly representational concepts, and our current aim is just to characterize these appearances, not (yet) to assess their accuracy.

[3] This construal of the naturalist project draws inspiration from Millikan's description of how the natural world supports cognition, in Chapter 1 of *Beyond Concepts* (2017).

[4] This general approach to domain individuation is adapted from Michael Lynch (2009, pp. 78-82).

Second, more significantly to present purposes, even if we could identify a subject matter for ethics as sketched above, there is a worry that this would not be sufficient to characterize what is distinctive of the moral domain, for what distinguishes moral thought and talk from other areas is not – or not just – what such thought and talk is about. Additionally, we must appeal to the *function* of paradigmatic claims and judgments in the domain. This can be illustrated by considering thought and talk about humor: Imagine someone, like Data from *Star Trek*, perfectly reliable in telling whether a joke is funny, yet who is incapable of experiencing amusement from a joke (see Bar-On and Chrisman, 2009, footnote 8).[5] This person can perfectly well track whatever features make for a humorous joke – the properties in virtue of which something can be considered funny – yet never enjoys or experientially engages with humor the way we normally would. It seems that there would be something crucial missing in this person's understanding of humor. There is a sense in which, although they are perfectly capable of understanding jokes, they are not a full participant in the domain of humor, because they cannot properly engage with its purpose.

The key point here is just that identifying the referents of core concepts in a given domain does not, in itself and in general, suffice to characterize what is distinctive of that domain, i.e., in virtue of what a claim or judgment belongs to that domain. I do not claim that we should reject the bottom-up approach to domain-individuation altogether. Rather, the idea is that the bottom-up approach characterizes a diagnostic principle for domain-individuation. Another dimension along which we must contrast domains is in terms of the *function* that our thought and talk in each domain serves for us. This latter functional principle for domain individuation can be

[5] See If a more realistic case is wanted, consider that there is research suggesting that individuals with autism and Asperger's are impaired in humor appreciation (though there is some anecdotal evidence to the contrary). See Lyons and Fitzgerald (2004) for discussion.

considered a ‘top-down’ approach. It individuates domains according to their functional profile, what they do for us who engage in thought and talk in those domains, rather than according to the fundamental features of the natural world we inhabit. Of course, it may be that the concepts in a given domain accomplish their function by enabling us to track the natural contours of a particular bit of the world. But we should not assume that all domains will have such a tracking function. It can be that in some domains, the core concepts *construct* or *constitute* properties, rather than tracking an antecedently existing objective reality. (Humor may be an example; what makes something humorous maybe that it is apt to produce a certain kind of amusement in suitably placed subjects.)[6] And it may be that in some domains, the central concepts do not denote properties at all (whether constructed or tracked), but rather serve some other purpose, such as expressing attitudes – consider that some pejorative terms seem to lack clear descriptive content, such as ‘jerk’. I assign theoretical priority to the top-down principle for domain individuation, because when domains can be individuated according to their subject matter, this is *because* of the function of those domains; the bottom-up approach is better understood as diagnostic, rather than explanatory.[7] For instance, chemistry and psychology each arguably share a descriptive function, but differ in what aspect of the world they purport to describe.

The top-down principle can operate at a finer or coarser grain. The function of a domain may itself contribute to the functioning of an encompassing broader domain. So for instance,

[6] For this reason, some have proposed that truths about humor are not true in virtue of representing an independently existing realm of humor facts, but instead are true in virtue of the responses or dispositions of competent participants in humor discourse. See Wright (1992) for discussion of the case of humor and its significance for the theory of truth.

[7] This division between bottom-up and top-down principles of domain individuation and the priority assigned to top-down principles is thematically in the spirit of recent pragmatist approaches to representation in general. See Price (2013) for a representative example. Price distinguishes, for instance, between what he calls ‘e-representation’ (or external, ‘environment-tracking’ representation) and ‘i-representation’ (to do with the internal, inferential functional role of representation) (2013, p. 36). Price suggests that we can assign priority to i-representation and still find a place for e-representation: it may be that the internal inferential function of some class of representations is to track external states of affairs – i.e. it may be the i-representational function of some domain to e-represent.

assuming for illustrative purposes that the function of chemical thought and discourse as a whole is to enable us to better navigate the chemical part of the natural world, different areas of chemical thought and discourse may contribute to this function in different ways: whereas ‘pure’ academic lab research contributes to this function by contributing to our knowledge of chemical kinds, chemical engineering contributes to this function by applying this knowledge in order to manipulate chemical processes for a variety of uses, e.g. the development of medicines and other products. It is important to acknowledge the complexity and interconnectedness of the various areas of our thought and talk, while also acknowledging the important distinctions between them. We must remember, as Wittgenstein remarks, that “our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with regular straight streets and uniform houses” (1973, §18). When it comes to the moral realm, too, we should not assume that all ethical thought and discourse is entirely uniform and isolated from other domains. I shall spend time (in Chapter 5) distinguishing between the distal and proximal functions of ethical claims and judgments, for instance. My goal will be to identify the most distal function that ethical claims and judgments share, that sets them apart from other claims and judgments in other domains, while recognizing that different sorts of ethical claims and judgments can contribute to this function in different ways.

Having articulated some of the ambitions of this project, I now commence with the exposition and argument, beginning with a characterization of the surface features of ethical thought and discourse to be explained. I divide these into two categories; features of the ethical domain that are shared with other, paradigmatically factual domains, and features that distinguish ethics from these and other domains.

1.3 Continuities with other domains: the cognitivist appearances

Ethical sentences, claims, and judgments are like other ordinary factual sentences, claims, and judgments in a number of ways. I categorize these similarities into two groupings: the *semantic continuities*, and the *cognitive continuities*.

1.3.1 The semantic continuities

Consider the following sentences:

(1) "It is morally wrong to consume animal products."

(2) "You ought to apologize for having lied to your friend."

(3) "It is a moral duty to keep one's promises."

Each of these sentences is apt for use in making ethical claims. And they are semantically continuous with ordinary non-moral sentences in a variety of ways, including:

Truth-aptness: Ethical sentences are capable of being assessed as either true or false.

I take it as platitudinous that ethical sentences are truth-apt.[8] As far as the surface-level features of ethical thought and discourse are concerned, note how naturally the truth-predicate applies to (1)-(3) in ordinary English: "It is true that it is morally wrong to consume animal products"; "It is false that you ought to apologize for having lied to your friend"; etc., are perfectly grammatical pieces of English. Of course, merely admitting that ethical claims are truth-apt does not immediately lead to any more substantive metaethical commitments, because admitting a certain domain is truth-apt does not require any particular (substantive) theory of truth for that domain, nor does it say anything about what the truthmakers of (1)-(3) might be if

[8] While early emotivists (such as Ayer, 1936) held that ethical claims were not truth-apt, it is generally agreed upon now that ethical sentences must be truth-apt in some sense, though different theorists will explain their truth-aptness in different ways.)

any of them are true. We are, recall, only articulating the surface appearances of the ethical domain.

Related to the feature of truth-aptness, we have:

> *Embeddability*: Ethical sentences, like any ordinary declarative sentence, are capable of being the objects of propositional attitude ascriptions, can be significantly embedded in conditionals and other constructions, can be meaningfully translated, etc.

Again, I take it to be platitudinous that ethical sentences exhibit embeddability. Embedding (1)-(3) in conditionals and other constructions is clearly linguistically meaningful, as in: "if it is wrong to lie, then it is wrong to get one's little brother to lie". Ethical sentences also appear as the objects of that-clauses in propositional attitude reports, as in "John believes that torture is always morally impermissible". This suggests that ethical sentences, like other ordinary indicative sentences, express propositions, where propositions are understood as whatever it is that is preserved in good translation, are the objects of propositional attitudes, and are what sentences contribute to the meanings of constructions in which they are embedded.

More generally, ethical sentences appear on their surface to be semantically unremarkable. There is nothing in their surface appearance to motivate providing a completely separate semantic analysis specifically for ethical sentences. The plausibility of proposing a completely distinct analysis for ethical sentences, then, is contingent on the strength of the metaethical considerations for thinking the ethical domain must be somehow significantly semantically different from other domains.

1.3.2 The cognitive continuities

In addition to the continuities between ethical *sentences* and sentences in other domains, there are also several continuities between ethical *judgments* and beliefs in other domains. For instance, as we saw in the previous section, ascriptions of *moral beliefs* to others appear unproblematic. On the surface, then, ethical judgments seem to be beliefs. Moreover, when a person makes an ethical claim, it is natural to say that in doing so, she expresses a moral belief in the content of the claim. Call this:

> *Belief-expression*: When S makes a sincere ethical claim, S expresses her belief in the content of that claim.

Belief-expression highlights one of the ways in which ethical claims appear to be continuous with ordinary non-ethical claims. In this instance, the continuity is not only a semantic one, but an apparent *epistemic* continuity. Along these lines, note that it is also unremarkable for us to talk of moral *knowledge*. Given a justified-true-belief model for knowledge, this leads us to suppose (as also seems *prima facie* to be the case) that ethical beliefs can enjoy epistemic justification, and that we can exchange such justifications in joint moral deliberation. This leads us to:

> *Justification-aptness*: Moral beliefs can be (or fail to be) epistemically justified.
>
> When S makes an ethical assertion, S becomes accountable for defending her claim with her justification for it if challenged.

Ethical claims, like many ordinary claims, enter the game of giving and asking for reasons. We often do offer and request reasons for accepting certain ethical claims, and we do assess some reasons as better or worse, as supporting an ethical position or not. If one were to make an ethical claim in uttering “abortion is morally wrong”, then, *ceteris paribus*, one is appropriately subject to requests for reasons for thinking that abortion is morally wrong. This feature of ethical thought and discourse is more substantive than the previous points, as this does set ethical thought and discourse apart from some other domains that exhibit truth-aptness and embeddability. For instance, thought and discourse about matters of taste, though arguably satisfying truth-aptness, embeddability, and belief-expression, does not generally exhibit justification-aptness (at least, not to the same degree as ethical discourse) – there is not the same expectation in taste discourse that one be able to supply reasons for one's interlocutor to adopt one's own attitude. We are often content to agree to disagree on such matters. Similarly, ethical thought and discourse is unlike thought and discourse about taste in exhibiting:

> *Objectivity*: Whether a basic ethical claim M is true does not vary depending on what individuals or groups happen to think about M. And if M is true, it is true for anyone in relevantly similar circumstances.

Part of the explanation of why it is that ethical claims are subject to requests for reasons, I think, is that ethics is at least implicitly taken to be an objective matter. We expect each other to provide reasons for the claims we put forward in ethics because we take ethical matters to be ones about which our beliefs can be correct or incorrect, and the exchange of reasons is how we seek to get things right. I think objectivity also contributes to explaining ethical disagreement. : When it comes to ethical disagreement there is what I call a 'felt rational pressure' towards resolution to a substantive consensus, where this amounts to, roughly, a *pro tanto* commitment on the part of the disputants to there being a correct position to take on the matter, and that a satisfactory resolution of the disagreement requires both disputants to settle on what that correct position is.[9] Note, again, that this is a way in which moral disagreement is different from (some) disagreement about taste.

The notion of objectivity offered here, while marking a way in which ethical thought and discourse is different from some other areas of thought and discourse that exhibit truth-aptness,

[9] Johnson (2021, p. 4).

embeddability, belief-expression, and justification-aptness, is nevertheless a rather minimal condition. I wish to construe objectivity in a such a way that it does not imply moral realism. This is a surprisingly tricky matter, largely because there has been some unclarity about what both objectivity and realism amount to.[10] Since I need to contrast objectivity with *subjectivity* and *relativity*, rather than with anti-realism, I find it useful to understand objective truth as mind-independent, in the sense that the correct moral position to take in a given case does not vary depending on what anyone happens to think. This notion of objectivity is neutral on whether moral truths are mind-independent in the sense that they would obtain even in a universe lacking any moral agents. This leaves open the possibility that even if it is not up to us *what* the moral truths are, *that* there are moral truths is dependent on the existence of minds like ours.[11] Though the view I go on to defend is compatible with robust moral realism, we should not build robust moral realism into the very data to be explained by an adequate metaethical theory.

The features discussed so far – truth-aptness, embeddability, belief-expression, justification-aptness, and objectivity – are all characteristic features of genuinely cognitivist thought and discourse. Cognitivism about an area of thought and discourse is, roughly, the idea that that area of thought and discourse aims at representing and describing objective features of the world. Thus, taking scientific inquiry as a paradigmatic instance of cognitivist thought and discourse, we can say that scientific claims aim to represent the world as it actually is. Such

[10] See Dunaway (2019) for discussion of the difficulties in characterizing realism and objectivism in respect to each other. See also Hopster (2017).

[11] Discussions of objectivity in metaethics sometimes affiliate it with realism, apparently under the assumption that if moral truths are objective, this is because they report on an independently existing reality. However, my goal here is to characterize objectivity as a surface feature of ethical thought and discourse, part of the data to be explained, so it is important to construe this feature in a theoretically neutral way. I therefore avoiding construing objectivity in terms of what Russ Shafer-Landau calls 'stance-independence', which is the idea that the moral truths are not made true in virtue of "their ratification from within any given actual or hypothetical perspective" (2003, p. 15). Stance-independence serves better as a characterization of realism (as Shafer-Landau intends) than of objectivity. See Hopster (2017) for recent discussion of objectivity and realism.

claims then are (i) capable of being true or false, depending on whether they succeed in representing the world as it is or not; (ii) sentences of that domain (e.g. “water is H2O”) clearly have propositional content that is embeddable in other constructions (“if water is H2O, then it will boil at sea level at 212 degrees Fahrenheit”), and (iii) that can be used to express beliefs, where those beliefs aim to represent the world, and (iv) that are subject to requests for reasons, and (v) whose truth depends only on the world being as it is represented rather than on features of the minds doing the representing. We have seen that ethical thought and discourse appears to be like scientific thought and discourse in many of these respects.

I turn now to explore some further features of ethical thought and discourse, features which distinguish ethics from other clearly cognitivist domains. Cognitivism, which I understand as the view that ethical claims are just like ordinary descriptive ones, and that ethical claims express representational beliefs, straightforwardly accounts for the various continuities just discussed between ethics and various other factual domains. However, cognitivism, on its own, does not offer any particular account of the distinctive features of the ethical domain. Thus, unless these features can be exposed to be merely apparent, cognitivism is itself seemingly an incomplete view about ethical thought and discourse.

1.4 Distinctive features of ethical thought and discourse

Consider the following case, adapted from Smith (1994, p. 6): suppose I were to make an ethical claim by uttering: “Donating to charity is morally good.” At the very next moment, volunteers from Oxfam walk through the door, seeking donations to charitable causes. You watch as I turn to them and say: “Yes, I judge that it is morally good to donate to charity, but I am not at all motivated to do so; I couldn’t care less whether I donate to your cause or not.” There would be something very strange in my saying this – you might start to wonder whether I was being sincere when I said it was morally good to donate to charity. The strangeness of this situation highlights a distinctive feature of ethical thought and discourse:

1.4.1 The connection to motivation

> *Connection to motivation*: If an individual makes a sincere ethical claim or judgment, there is an expectation that they will be at least somewhat motivated to act in accordance with that claim or judgment.

This feature of ethical thought and discourse sets it apart from other paradigmatically cognitivist domains; if I were to claim "it is raining outside", for instance, there is no expectation that I am motivated to act in any particular way, unless one also assumes that I have some other desires (such as a desire to walk outside without getting wet). While metaethicists generally agree that there is *some* sort of connection between sincere moral judgment and motivation, the character of this connection is subject to intense debate in the literature. I have stated the *connection to motivation* feature in a generic way that is compatible with various readings of different strengths. My goal below is to specify a reading of the *connection* feature that captures ordinary intuitions about the relation of moral judgment to motivation, but that does not assume any particular explanation of this connection.

The 'expectation' here can be read in several ways, yielding stronger and weaker versions of the *Connection* feature. A *strong descriptivist* reading of the expectation would have it that as a matter of conceptual necessity, one cannot sincerely judge that M and lack any motivation whatsoever to act in accordance with M. This reading is, in effect, a statement of a strong version of the controversial thesis of *motivational internalism*.[12] A *weak descriptive* reading would have

[12] The relevant thesis here, also known as moral judgment internalism (Darwall, 1983, 1995) or appraiser internalism (Brink, 1989, p. 40, 1986, p. 27), is to be distinguished from internalism about reasons, which holds, roughly, that a moral obligation to phi provides reason to phi. For discussions of the varieties of internalism in ethics, see Audi (1998), Darwall (1983, 1995), Parfit (1998) and Shafer-Landau (2003, p. 144-145).

it that as a matter of contingent psychological fact, those who sincerely judge that M are statistically typically at least somewhat motivated to act in accordance with M. Such a reading is compatible with motivational externalism (understood as the denial of internalism), while still potentially characterizing what seems to be distinctive about the moral domain.[13] A *strong normative* reading would have it that if one were to sincerely judge that M yet lack any motivation to act in accordance with M, one would exhibit a *moral* failing. And a *weak normative* reading would have it that if one sincerely judged that M yet lacked any motivation to act in accordance with M, one would violate a default presumption of moral thought and discourse, a presumption with weaker normative force than ordinary moral obligation.[14, 15, 16] The 'presumption' here is intended to be alike in normative force to the status, in humor discourse, of the presumption that only laughs at what one finds humorous.

In *Connection to motivation*, I intend the weak normative reading. While other readings may also be true, it is the weak normative reading that captures a distinctive feature of ethical thought and discourse as such. The strong normative reading encodes a substantive moral principle (that one morally ought to be motivated to act on the moral judgments one makes),

[13] Those persuaded by internalism, however, will likely complain that such a contingent psychological connection does not fully capture the phenomenon to be explained here. I share this concern.

[14] This is not to say that a violation would be statistically atypical, though it may be. See below for further discussion of the presumptive status of the *connection to motivation*.

[15] This assumes, of course, that there are different kinds of normativity or normative force. I think this assumption has intuitive appeal. Violations of a moral principle seem more serious, normatively speaking, than violations of the norms of etiquette, for instance. One might argue that really all normativity is of a piece – perhaps etiquette norms are reducible to moral norms to do with respect, for instance – but an argument would be needed. And even if all normativity does reduce to a single kind, we still need conceptual resources for mapping differences in normative force – e.g. we still need a principled way to distinguish strong or overriding normative reasons from weak or defeasible ones. The contention here is just that discourse norms will, as a class, have weaker normative force than typical moral requirements.

[16] The weak normative reading is amenable to recent attempts to restate the thesis of motivational internalism in a plausible manner, out of recognition of the need to accommodate apparent failures of motivation in the depressed, the satanic, and amoralists (more on this below). Communal forms of internalism claim that a necessary condition for a community to engage in moral thought and discourse is that members of that community are at least somewhat motivated to act on their moral judgments in a critical proportion of cases. See Bedke (2009) and Tresan (2006, 2009) for instances of communal internalism.

rather than a principle *about* moral thought and discourse itself.[17] The weak descriptive reading might suffice to state a difference between moral claims and judgments and non-moral ones, and one that calls out for explanation if it is true. But even so, there remains something still to be explained; the apparent *normative* profile of the connection between ethical judgment and motivation. This can be illustrated by considering a structurally similar issue in the literature on basic self-knowledge. Part of what sets the agenda for theories of self-knowledge is the observation that basic self-knowledge of current mental states (such as my knowledge that I am feeling anxious right now) is especially epistemically secure.[18] The security seems not to be (simply) a matter of contingent psychological correlation between first-order M's and judgments that one is in M. Nor does the security seem to consist in the operation of an especially reliable epistemic method of introspection or inner scanning, for instance.[19] Additionally, what needs to be explained is how basic self-beliefs are intuitively immune from certain epistemic challenges and doubts; challenges that other empirical judgments (even very epistemically reliable ones) are subject to. For instance, even if I am very reliable in judging how to get from one point in the city to another, it is still epistemically appropriate for you to ask *how* I know (e.g., by consulting a map), whereas such a question ("How do you know?") would be epistemically inappropriate to raise concerning avowed self-knowledge, e.g. my knowledge that I currently have a headache.

[17] Also, as Bloomfield (2001, pp. 156-159) points out, that we morally ought to be motivated to act in accordance with our moral judgments is not what is under debate between internalists and externalists; it is open to both sides of the debate to accept this. Internalists make a different claim: that the connection between moral judgment and motivation is somehow conceptually necessary, not (just) deontically necessary.

[18] This observation needs some qualification; basic psychological self-knowledge concerns states such as pains, hopes, thoughts, etc., and their contents. It does not include Socratic self-knowledge or knowledge of one's character traits, etc. See Bar-On (2004, Chapter 1), Coliva (2016a, Chapter 3), and Gertler (2011, Chapter 3) for discussion of the apparent epistemic authority of basic self-belief and qualifications to such authority. Some are skeptical that there we have such special authority at all, for any of our mental states (see Snowdon, 2012; Schwitzgebel, 2006; Carruthers, 2013; Williamson, 2001) Replies generally involve qualifying the strength and necessary conditions for authority to obtain (see e.g., Wright, 2015).

[19] For an overview and criticism of 'inner scanner' approaches to self-knowledge, see Gertler (2011, Chapter 5). See also Bar-On (2004, Chapter 4) for criticism of 'epistemic' theories of self-knowledge generally.

Likewise, it seems to be a default presumption of the moral realm that when one judges that it is morally required to phi, one is at least somewhat motivated to phi, *ceteris paribus*. Consider: if someone to ask, "Why are you phi-ing?", it seems to suffice as an answer to make the moral claim that phi-ing is morally required. By contrast, making an ordinary empirical claim, in general, does not seem to suffice to explain any of one's actions in the absence of further assumptions (however implicit) about one's goals or antecedent motivations. I maintain that part of what needs to be explained in the *Connection to motivation* is why such further assumptions are not needed in the moral case. The strong descriptive reading of the expectation to be explained accords with the normative readings in positing a connection that is stronger than could be explained just by a contingent psychological connection. However, the strong descriptive reading appears to be *too* strong, inviting easy counterexamples. Purported cases of amoralists, and real cases of depression and *akrasia* spring to mind: depressed individuals seem, *prima facie*, capable of making ethical judgments, even if they sometimes lack motivation to act in accordance with those judgments owing to their depression.

This is an apt point at which to discuss the metaethical significance of motivational internalism and externalism in more detail. (The preceding discussion was intended to dissociate motivational internalism as a thesis about moral judgment from the surface feature of ethical thought and discourse this thesis might be used to explain). Motivational internalism I understand to include any view according to which there is a more than merely contingent connection between sincere moral judgments and motivation to act. Externalism is the denial of internalism. Internalists typically maintain that this is a necessary conceptual connection; internalists who accept this hold to the strong descriptive reading of *Connection to motivation*. Externalists can point to apparent counterexamples to the strong descriptivist reading to contest the ‘data’ appealed to in motivating internalist views.[20] However, internalists can also try to assimilate these apparent counterexamples in various ways – e.g. qualifying and weakening the internalist thesis, or rejecting that purported counterexamples really involve sincere and competently issued ethical judgments.[21] Motivational internalism is an intensely debated thesis, and I cannot fully present the literature on the topic. But several points concerning the relation of motivational internalism to the topic of this dissertation are in order.

A) The debate over motivational internalism is significant in part because the internalist thesis has been thought to have implications for discussion about ethical cognitivism and non-cognitivism, with a focus on arguments for thinking that the truth of motivational internalism entails the falsity of ethical cognitivism. Here is a rough characterization of two arguments for this entailment.

First, in the Humean tradition in the philosophy of mind, beliefs are taken to be motivationally inert and so cannot give rise to motivations to act except in conjunction with desires (or desire-like states).[22] But given the internalist thesis that ethical judgments are capable of motivating action all on their own, it then seems that ethical judgments cannot be beliefs. It follows that cognitivism is false. Similarly, one might argue that beliefs essentially have a mind-to-world direction of fit; one’s beliefs are to fit the way the world is – while desire-like motivational states have a world-to-mind direction of fit; the world is to be changed to fit the

[20] For discussion of cases that pose a challenge to motivational internalism, see Brink (1986, pp. 29-31; 1989, Chapter 3); Mele (1996); Miller (2008); Roskies (2003); Stocker (1979); and Svavarsdottir (1999).

[21] For discussion of qualified versions of internalism, see Bjornsson (2002); Blackburn (1998, pp. 59-68); Eriksson (2006, pp. 172-87); Gibbard (2003, p. 154); Timmons (1999, p. 140); Bedke (2009); Korsgaard (1996); Smith (1994); Wallace (2006); McDowell (1978, 1979); McNaughton (1988, Chapter 8); Tollhurst (1995); Wiggins (1991); and Tresan (2006, 2009).

[22] This is derived from Hume (1739/1888, p. 457): “Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason”.

content of one's desire.[23] Since ethical judgments necessarily involve desire-like motivational states (according to internalism), they have the wrong direction of fit to be beliefs. It follows that cognitivism is false.

Second, there is the 'queerness' argument, originally due to Mackie (1977) (I also discuss this in Chapter 6).[24, 25] The argument here is that if there were objective moral principles (as the cognitivist thinks our ethical thought and discourse aims to describe), they would have to be inadmissibly strange if motivational internalism were true – in order to explain the necessary connection between ethical judgments and motivation, the objects of such moral judgments would have to have 'to-be-pursuedness' built into them. But no objects have 'to-be-pursuedness' built into them. We apparently must either reject motivational internalism, posit an error theory, or deny that ethical thought and discourse is cognitivist in the first place.

B) Recent developments in thinking about motivational internalism cast doubt on the thought that internalism entails the denial of cognitivism. For instance, according to non-constitutional motivational internalism (Tresan, 2009), in order for a mental state to count as a moral judgment, it must be *accompanied* by a motivational state – hence internalism – but the moral judgment itself is not *constituted* by that motivational state, and so it is open to think the judgment is an ordinary, motivationally inert, belief. Additionally, the Humean approach in the philosophy of mind and direction-of-fit considerations have not gone without challenge. For instance, Millikan (2005a) identifies a category of representations – Pusmi-Pullyu

[23] The notion of 'direction of fit' is introduced in Anscombe (1963) and provides additional ammunition to the Humean psychological theory considered above.

[24] Thought Mackie intends the argument to support his moral error theory, the argument is perhaps more significant as providing indirect support for expressivism.

[25] I will sometimes refer to this as the 'Mackie-inspired argument'. I find the label ("the queerness argument") problematic, as it seems to imply that there is something wrong or inadmissible about being queer.

Representations – that have both directions of fit at once.[26] And these representations are not particularly complex or unusual – Millikan identifies this sort of representation in animal communication, including, for instance, a hen's food-call to its chicks. Likewise, the view I shall go on to defend holds that in the ethical case, moral judgments exhibit two directions of fit at once. If such an account of moral judgment can be defended, the tension between internalism and cognitivism arising from considerations of direction of fit would be defused.

Regarding the Mackie-inspired argument: it is not clear to me that moral properties must be inherently motivational if motivational internalism is true, as the argument supposes. The Mackie-inspired argument derives a conclusion about the metaphysics of moral properties from a premise positing a necessary connection between ethical judgment and motivation. This inference assumes that for a moral judgment to be motivating independently of the antecedent desires of the judger, it must be what is *represented* (i.e., moral properties) by that judgment that is inherently motivating. We should deny this assumption. It would indeed be strange for a property to have 'to-be-pursuedness' built into it, but this assumption is not a necessary corollary of motivational internalism. I shall go on to defend an account on which the connection between moral judgment and motivation is borne out in the norms placed on the *act* of making the relevant moral judgment in speech or in thought, rather than in terms of what such judgments represent. Compare: part of what it is to properly play a game of chess is that one aims at checkmating the opposing king – this norm is partly constitutive of the rules of chess – but we should not conclude from this that 'to-be-checkmatedness' must be an inherent feature of the opposing king piece (nor should we be error theorists about chess pieces). Likewise, even if part

[26] See also McDowell's (1998) notion of 'besire'. Little (1997, p. 64) also points out that a single mental state could coherently have two directions of fit, so long as it has these with respect to different propositional contents. My own proposal in Chapter 5 will follow in this tradition, with particular emphasis on Millikan's notion of a Pushmi-Pullyu Representation.

of what it is to engage in ethical thought and discourse is that one is motivated to act in accordance with the ethical claims one accepts, this does not require there to be something magnetic about the things picked out by moral predicates such as ‘is good’.

C) Setting aside the purported conflict between internalism and cognitivism, motivational internalism about moral judgment is sometimes associated with *expressivism* as a thesis about moral discourse. If one accepted the argument considered above concerning direction of fit and Humean psychology one might then wonder what it is we are up to when we make ethical claims, if not expressing belief. Expressivism answers this question: according to traditional expressivism, in making an ethical claim, we do not purport to describe some feature of the world, or to express our beliefs; instead, we are expressing certain motivationally-charged non-cognitive states.[27] The expressivist thesis clearly fits neatly with internalism, and offers a straightforward explanation of the connection to motivation. Even if we reject the Humean picture of psychology, I think expressivism still offers a promising start on explaining the connection to motivation. I critically discuss expressivism in Chapter 2.

1.4.2 Action-guidingness and disagreement

Providing an adequate explanation of the connection to motivation will be the main focus of this dissertation. The challenge will be to propose an account that captures the special closeness there intuitively seems to be between sincere ethical judgment and motivation, while also accounting for everyday (and more radical) failures of motivation, wherein it seems intuitive that one has made a genuine moral judgment yet lacks the relevant motivation. There are, however, two other features seemingly distinctive of the ethical domain, each of which I shall also account for over the following chapters, especially Chapters 5 and 6.

[27] For a clear articulation of the support that motivational internalism is supposed to provide for non-cognitivist theories (including expressivism), see Shafer-Landau (2003, p. 119-121).

The connection to motivation captures a relation between agents' sincere ethical judgments and their motivations. But many ethical judgments also seem to have an *outward-*facing connection to the actions of others. It seems uncontroversial that discourse about moral topics matters to us in part because it matters for what we do and what we expect each other to do. Prescriptivists (e.g., Hare, 1952) even see it as the *primary* function of ethical claims to prescribe courses of action. At any rate, a surface feature of ethical claims is that they are supposed to have some sort of bearing on the actions of others:

> *Action-guidingness:* Ethical claims are supposed to have direct bearing on the actions of others.

Expressivism, we saw, offered one non-cognitivist view of the ethical domain, since expressivists deny that ethical claims express beliefs, at least insofar as such beliefs are supposed to represent some moral reality. Prescriptivism represents another form of non-cognitivism, since it holds that the purpose of ethical claims is to *prescribe*, rather than *describe*, and that accordingly ethical judgments are desire-like states, rather than representational beliefs. So each of expressivism and prescriptivism is naturally paired with motivational internalism, and the denial of cognitivism. Accordingly, they share many of the same advantages and disadvantages, focusing on different aspects of the features of ethical thought and discourse. In the next chapter, I shall focus on the prospects and problems for expressivism, but much of that discussion carries over to prescriptivism as well. In Chapter 5, I propose a hybrid view of ethical judgments that explains both the connection to motivation, as well as vindicating an element of the prescriptivist idea about the function of ethical claims.

Finally, ethical thought and discourse seems to be distinctive in the range and depth of moral disagreement. Moral disagreement, including inter- and intra-cultural and disagreement across time, has seemed to be both more widespread and more intractable than disagreement in other areas. As purported examples of wide-ranging and intractable moral differences, Michael Ridge suggests the following (2014, p. 65):

> Cannibalism was quite commonplace in pre-agrarian human societies and routinely seen as morally permissible. Culturally sanctioned violence as sport has also been commonplace historically - for example, the Roman Colosseum. Slavery is another obvious example. The systematic subordination of women is another. . . . Closer to home, we find deep moral disagreement within modern Western societies on issues of life and death, the role of freedom and equality in a just state, social justice more generally, criminal justice, the sanctity of life versus the quality of life, and so on.

Ridge contends not only that moral disagreement is rampant, but that it is also different in kind from other kinds of ordinary disagreement in that it is particularly intractable, such that it does not admit of rational resolution.[28] Thus, we have, as a proposed further distinctive feature of ethical thought and discourse, the following:

> *Disagreement*: Disagreement over ethical matters is more widespread, and more intractable, than disagreement in other cognitive areas of thought and discourse.

Moral disagreement has metaethical significance for several reasons. First, some take the range of moral disagreement, together with certain assumptions about semantic and conceptual competence with moral terms, to undermine the idea that ethical terms are like natural kind terms in denoting naturalistic moral properties. Second, some take the intractability of moral disagreement to support the thought there are no objective moral facts that could settle such disagreements. And third, in the epistemology of disagreement, it has seemed to many that the fact of disagreement itself provides some reason for one to lower confidence in one's view.[29] This, combined with the observation that moral disagreement is widespread and intractable,

[28] See also Rowland (2016) for a discussion of the epistemological significance of moral disagreement.

[29] This is roughly the contention of conciliatory views in the epistemology of disagreement. See Feldman (2006); Christensen (2007); Elga (2007) for defenses of conciliatory positions.

seems to lead to the conclusion that we are never warranted enough in our ethical judgments to rationally claim moral knowledge (Rowland, 2016). I discuss each of these arguments in greater length in Chapter 6, where my goal is to defend the possibility of moral knowledge (one of the cognitive continuities) against the purported skeptical import of moral disagreement.

One might wonder whether ethical disagreement really is more widespread or more intractable than disagreement in other purportedly factual domains. For instance, what appears at first to be a deep moral disagreement may ultimately turn out to be based on a further disagreement concerning a non-moral matter, such that settling the non-moral matter would allow for a resolution of the moral disagreement. (For example, an apparent deep disagreement over the permissibility of the death penalty may sometimes be due to a difference in opinion over the effectiveness of the death penalty at deterring future crime). While I think there is some truth to those criticisms, and that the extent and depth of moral disagreement has been sometimes exaggerated, I do still think that serious moral disagreements occur more frequently than serious non-moral disagreements, and I think that moral disagreements are at least sometimes intractable – but I also think there are intractable disagreements concerning non-moral matters. See Chapter 6 for further discussion, where I argue, as part of a defense against moral skepticism that (i) there are intractable disagreements in moral matters, but (ii) intractable disagreement generally does not have the skeptical consequences some take it to. Any difference between disagreement in the moral and the non-moral realm, I suspect, is a matter of degree rather than of kind, and so I see *Disagreement* as having a dubious status as a feature of ethical thought and discourse to be explained.

Moral disagreement, surprisingly, figures *both* in arguments against and for moral realism. Typically, those who take moral disagreement to be more widespread or more intractable than ordinary disagreement use moral disagreement in arguing against moral realism. But objectivist moral realists might also see the *existence* of genuine moral disagreement as speaking against anti-realism. For if the anti-realist is right, and there are no moral facts or properties, there is nothing for us to be disagreeing *about*, and so we do not genuinely disagree. An additional challenge for moral anti-realism, loosely related to disagreement, concerns the possibility of moral error. In the next chapter, I explicate what I take to be one of the most pressing problems for traditional expressivism (an anti-realist view), namely, that traditional expressivism cannot explain the intuition that one could regard oneself as possibly in fundamental moral error (Egan, 2007).

1.5 A survey of the landscape

I have now articulated the features of ethical thought and discourse to be explained, and along the way I have given some indications of how various metaethical positions propose to account for some of these features. In this section I provide a very short discussion of some of the main families of metaethical theories and indicate their strengths and weaknesses when it comes to accounting for the features above, in order to situate the expressivist and hybrid theories that will be the focus of the remainder of the dissertation. The discussion to follow is brief, introductory, and follows standard moves in the literature, so readers already familiar with these ideas can skim or skip the following subsections. As we shall see, I find expressivist theories to be of particular interest, because I find them to offer important insights for explaining the distinctive features of ethical thought and discourse. However, traditional expressivism has some serious objections when it comes to accounting for the continuities between ethics and other domains. The goal of the next chapter will be to assess the prospects for pure expressivism when it comes to accounting for these continuities.

1.5.1 Non-naturalist moral realism

Moore, in the first chapter of his *Principia Ethica*, defended the view now known as non-naturalist moral realism.[30] The core commitments of this view are that (i) ethical judgments purport to represent objective moral facts and properties, (ii) such judgments are at least sometimes true, in virtue of correctly representing the moral facts and properties, and (iii) moral facts and properties are fundamentally unlike any natural facts and properties. Commitments (i) and (ii) amount to a non-skeptical moral realism, and afford a straightforward account of the semantic and cognitive continuities between ethics and other factual domains. The distinctive claim of non-naturalism is (iii), which sets ethics apart from other domains. And if one thought that one of the peculiar features of moral properties was that they were somehow inherently motivating, the resulting non-naturalist view promises to explain the connection to motivation and the action-guiding nature of ethical judgment.

Moore's motivation for (iii) derives from his Open Question Argument against naturalistic attempts to reductively analyze moral predicates in terms of natural properties. Very briefly, the idea is that moral properties could not be natural properties, because it is not possible for us to reductively define moral predicates in terms of natural predicates. For instance, consider the very simple proposal that the moral predicate 'is good' can be analyzed as 'causes pleasure'. Now, when it comes to analytic definitions, certain questions appear to be 'closed'. For instance, given the standard definition of 'bachelor' as 'unmarried male', it is not conceptually open for us to ask: "I know that John is an unmarried male, but is John a bachelor?". Asking such a question would show that one does not understand the concepts involved. By contrast, Moore suggests, analogous questions using moral predicates always seem 'open' to competent speakers, in that it

[30] Moore (1903). More recent defenses of non-naturalist moral realism can be found in Enoch (2011) and Shafer-Landau (2003).

always seems at least intelligible to ask “I know that X causes pleasure, but is X good?” (and similarly, for any other proposed naturalistic reductive definition). Asking such a question does not seem to betray any misunderstanding of the concepts involved. Accordingly, one version of the Open Question Argument contends that the best explanation of the apparent openness of these questions is that they are indeed open. From this, Mooreanism concludes that moral predicates are undefinable, and that the properties they picked out therefore have to be fundamentally different from other kinds of properties.[31]

Non-naturalist realism faces some serious objections. To my mind, the most pressing issues for non-naturalism are (i) that the view requires rejection of a naturalistic worldview, and (ii) that moral knowledge appears mysterious on the non-naturalist account. A full defense of a naturalistic worldview goes beyond the scope of this dissertation, but given parsimony considerations, the burden of proof is surely on the non-naturalist to provide strong independent reasons for thinking we must countenance non-naturalist entities in our ontology. The non-naturalist’s problem with moral knowledge is this: if non-naturalism is correct, then it seems that moral properties are unlike natural properties in that moral properties are causally inert. But ordinarily, our knowledge of the world is causally mediated – having knowledge of ordinary matters of fact seems to require that our beliefs be causally related to the facts of the matter in the right way. Otherwise, our beliefs, even when true, would seemingly only be *accidentally* true. And accidentally true beliefs intuitively are not instances of knowledge. Non-naturalism

[31] Another route to the non-naturalist position combines the Open Question phenomenon with the observation that reference to moral properties and truths is indispensable to our deliberation over what to do. Just as some take the *explanatory* indispensability of positing the existence of certain scientific or mathematical entities to provide justification for thinking those entities really exist, one might also take the *deliberative* indispensability of moral facts and truths to provide justification for thinking moral properties really exist. Combined with the Open Question Argument, it would seem that such entities would have to be fundamentally unlike natural properties (see Enoch, 2011, Chapter 3).

thus owes us an account of what could justify moral beliefs so as to make them knowledgeable when true – that is, it owes an account of justification-aptness, one of the cognitive continuities discussed above.[32] While non-naturalists have made some proposals (such as intuitionism about moral knowledge), the view I shall go on to defend in this dissertation supports an account of moral knowledge according to which it is continuous with knowledge in other domains.

1.5.2 Naturalist moral realism

Naturalist moral realism sees the moral domain as broadly continuous in significant respects to other factual domains.[33] This family of views shares with non-naturalism the commitments that (i) ethical judgments purport to represent objective moral facts and properties, and that (ii) such judgments are at least sometimes true, in virtue of correctly representing the moral facts and properties. But it holds that moral facts just are natural facts, and accordingly avoids the two challenges for non-naturalism above. Now, in light of Moorean Open Question considerations, and given certain developments in the philosophy of language and mind over the past 50 years, contemporary naturalist realists tend not to offer analytic reductions of moral predicates in naturalistic terms – a temptation to do so may seem these days to reflect an inclination towards an outdated verificationist stance about meaning. There are several varieties of non-reductionist naturalist moral realism, and I cannot consider them all here. Below, I briefly discuss just one prominent way of defending naturalist realism. (And later, in Chapter 5, I propose a biosemantic account of ethical judgment that is amenable to naturalist moral realism).

[32] There is some license for optimism for a non-naturalist account of moral knowledge however. The assumption that all knowledge is causally mediated can be questioned – consider mathematical knowledge for instance. It is hard to deny that we can ordinarily know that 2+2=4, even though it is hard to say what ‘causal contact’ with mathematical facts or objects would look like. See Balaguer (1998) for discussion of epistemological problem in mathematics.

[33] See, among others, Bloomfield (2001); Boyd (1988); Brink (1986); Foot (2001); Hursthouse (1999); Jackson (1998); Sturgeon (1985); and Railton (1986).

One influential trend has been to adapt semantic externalism and a causal-historical picture of reference (owing to Burge, 1979; Kripke, 1980; Putnam, 1975) to the moral domain to explain how moral properties could be identical to certain complex natural properties, even though moral predicates and concepts are not analytically reducible in naturalistic terms. The basic idea is that moral predicates (such as 'is good') refer to certain naturalistic properties in virtue of the fact that those properties have historically causally regulated our use of 'good' and its translations in other languages, despite the fact that ordinary competent users of 'good' may not have a clear conception of which property, exactly, that term refers to. In this respect, moral predicates are supposed to function just like natural kind terms on the semantic externalist view. 'Water', as Putnam famously proposed, refers to H2O because it is that chemical substance that causally regulates our use of the term, even for competent users who may have no idea about the chemical composition of water.

The resulting view appears to improve upon non-naturalist realism in that it does not require a special epistemology for the moral domain, and it does not posit ontologically strange non-natural properties. Accordingly, naturalist moral realism appears to account for all of the semantic and cognitive continuities between ethics and other domains highlighted above. However, naturalist moral realism, on its own, does not seem to offer any particular explanation of the distinctive features of the moral domain, including the connection to motivation, the action-guiding nature of ethical claims, and the extent of moral disagreement. So at the very least, it seems to me that the burden lies with the naturalist moral realist to either propose some explanation of these features compatible with their view, or to provide strong justification for thinking that the apparent distinctive features of the ethical domain are misleading, and that ethics is in all respects like other paradigmatically factual domains.

1.5.3 Error theory

Error theorists agree with moral realism when it comes to the semantic continuities, and some of the cognitive continuities, between ethics and other factual domains.[34] In particular, error theorists agree that ethical sentences are truth-apt and embeddable, and that ethical claims and judgments purport to represent objective moral features of the world. However, error theorists deny that there are any moral facts or properties, and so they think that no basic ethical sentence, claim, or judgment is ever literally true. So, error theorists must deny some of the most obvious surface features of ethical thought and discourse, including the possibility of moral knowledge, which requires at least that some of our moral judgments are true. I take this to be a serious cost of error theory; in my view, we should only accept error theory as a last resort, in case all other views prove to have fatal flaws.[35]

But it is worth considering arguments for error theory, since some of them employ the distinctive features of the ethical domain discussed above as premises. A main goal of this dissertation is to present an account that preserves the continuities with other cognitive domains, but that also explains the distinctive features of the ethical domain; accordingly, I shall need to respond to some of the arguments for error theory. Two important such arguments are Mackie's (1977) argument from relativity and argument from 'queerness' – since I discussed the 'queerness' argument in 1.5.1, I will only describe the argument from relativity here. The argument from relativity starts from the premise that there is a greater degree in variation in moral belief both within a society over time and between different societies than there is in other factual domains. (This premise is closely related to the point about disagreement discussed

[34] See Garner (2007); Joyce (2001b); Kalderon (2005); Mackie (1977); Olson (2011).
[35] But see Lutz (2014) for arguments that error theory does not have objectionable moral and practical implications that make it unappealing.

above, namely, that moral disagreement seems to be more widespread and intractable than disagreement in other factual domains.) The argument then continues by proposing that the best explanation of the comparatively large degree of variation or disagreement in morality is that our moral beliefs reflect our various ways of life, rather than answering to some objective moral reality. As we shall see, adequately responding to this argument will require showing either that moral disagreement is less widespread and less intractable than it has seemed (but again, this risks denying what was supposed to be explained), or that there is an alternative explanation of moral disagreement that is superior to the error theorists. I address these issues in Chapter 6.

1.5.4 Cultural relativism and subjectivism

The defining commitment of relativism and subjectivism about ethics is the claim that moral truth is not objective, but instead varies according to the moral perspective that is predominant within a culture (cultural relativism), or the moral perspective of individual agents (subjectivism). (I shall limit my discussion to subjectivism, since the arguments for and against cultural relativism closely parallel the arguments for and against subjectivism, except on a cultural rather than individual scale).[36] In one sense, subjectivism is a form of realism, since it holds that there are moral truths, facts, and properties. Still, subjectivism is often considered as an anti-realist view, since the term ‘realism’ tends to be associated with the idea that there are *objective* moral truths that are in some significant sense mind-independent. More specifically, subjectivism can be understood to endorse the semantic thesis that ethical claims and judgments describe the moral perspective of the agent making those claims or judgments. Thus, on a simple subjectivist view, when S asserts “torture is morally wrong”, the content of S’s assertion is something like: that S disapproves of torture.

[36] For more sophisticated versions of views along these lines, see Dreier (1990) and Harman (1975).

Simple subjectivism at least accounts for the truth-aptness, embeddability, belief-aptness, and justification-aptness of ethical judgments. (However, one standard complaint is that the subjectivist accounts for justification-aptness and moral knowledge *too* well, by implausibly making moral knowledge as easy to come by as self-knowledge of one's current mental states).[37] Moreover, since the contents of ethical judgments, according to simple subjectivism, concern our own motivationally-charged attitudes such as moral approval/disapproval, subjectivism can provide a straightforward explanation of the connection to motivation. However, simple subjectivism has difficulties in accounting for moral disagreement. For simple subjectivism predicts that when S asserts "Torture is wrong", the propositional content of S's assertion is that S disapproves of torture, but when T then asserts "Torture is not wrong", the propositional content of T's assertion is that T does not disapprove of torture.[38] But those propositions can each be true at once. So wherein does the disagreement between S and T lie? Thus, far from accommodating the appearance that moral disagreement is widespread and intractable, simple subjectivism seems to predict that there is hardly any moral disagreement at all, unless we happen to disagree about what our interlocutor's actual moral attitudes are. I set aside simple subjectivism here, but in Chapter 3 I consider some recent hybrid views that take inspiration from subjectivism while aiming to account for the apparent objectivity of ethics and moral disagreement better than simple subjectivism.

1.5.5 Non-cognitivism: emotivism, prescriptivism, expressivism

[37] This standard objection is clearly laid out in van Roojen (2015, pp. 106-108), alongside other objections to simple subjectivism.

[38] For a statement of this standard objection, see van Roojen (2015, pp. 109-111). It is not obvious that similar objections will successfully apply to more sophisticated versions of subjectivism, but the objections set the agenda for a refinements of the subjectivist idea.

Non-cognitivist theories of ethical thought and discourse depart from the families of theory discussed so far in that they begin by considering, not what moral judgments are *about*, but rather, what is it that we are *doing* when we make moral judgments.[39] One might for this reason construe non-cognitivism as a pragmatist approach; it begins inquiry in the ethical domain by considering the function of ethical judgment – what such judgment does for the creatures that make them – without assuming from the outset that such judgment must have a descriptive, representational function. Indeed, traditional non-cognitivist theories deny that ethical claims have describing features of the world as their primary function. Accordingly, non-cognitivism incurs a steep explanatory cost, in that it appears to reject the various semantic and cognitive continuities between ethics and other domains. The main goal of the next chapter is to assess the extent to which one non-cognitive view – expressivism – is able to successfully accommodate those continuities.

Non-cognitivism takes the distinctive features of the ethical domain very seriously in constructing an account of the function of ethical judgment. Emotivists, for instance, see ethical claims as expressing certain affective, motivationally-laden states, such as moral approval or disapproval, and influencing others to take up similar attitudes (Ayer, 1936; Stevenson, 1937, 1944). Expressivists, too, view ethical claims as expressing motivationally-charged non-cognitive (that is, non-doxastic) states, but contemporary expressivists have a much richer conception than emotivists of the structure of such states, as we shall see (Gibbard, 1990, 2003; Blackburn, 1984, 1993). Each of these views provides a simple explanation of the apparent close connection between ethical judgment and motivation to act (as discussed above in 1.4.1), since

[39] The classic emotivist position is articulated in Ayer (1936), and Stevenson (1937, 1944). The most prominent proponent of prescriptivism is Hare (1952). Prominent versions of expressivism are found in Blackburn (1984, 1993; 1998); Gibbard (1990, 2003).

they hold that ethical judgments *express* motivationally-charged states. And prescriptivism emphasizes the action-guiding nature of ethical claims, holding that such claims have an imperatival illocutionary force, rather than an assertive force.

Superficially, non-cognitivist theories appear similar to simple subjectivism, insofar as they each posit a close connection between ethical claims and the motivationally-charged attitudes of the claimers. The crucial difference is that, whereas simple subjectivism maintains that to say that torture is wrong basically amounts to saying that one disapproves of torture, non-cognitivist views emphasize that *expressing* one's mental state is very different from *saying that* one is in it (Ayer, 1936, p. 111). Accordingly, non-cognitivism does not run into the problem of lost disagreement that faces simple subjectivism. Instead, non-cognitivists view moral disagreement as a kind of disagreement in *attitude* (Stevenson, 1937). However, early proponents of non-cognitivism, in holding that it is the semantic function of ethical sentences to express non-cognitive states, endorsed the much more radical idea that ethical sentences are not even truth-apt, and that ethical judgments are not beliefs. The main question for the next chapter is: To what extent do nuanced, sophisticated versions of expressivism avoid commitment to a radical denial of the semantic and cognitive continuities that seem to hold between ethics and other domains? Despite the important challenges and questions for the non-cognitivist approach, I think that the non-cognitivist idea of explaining the connection to motivation in terms of the expressive character of ethical claims is an important insight, and it is this feature of the view that I aim to preserve – even though the hybrid view I defend ends up sharing little else with its expressivist forebears.

1.6 Conclusion

The features of ethical thought and discourse calling out for explanation are truth-aptness, embeddability, belief-expression, justification-aptness, objectivity, motivation, action-guidingness, and disagreement. I have stated these features in as theoretically neutral a way as possible, so as to avoid confounding explanandum with explanans. The existence or extent of some of the features listed are nevertheless themselves the subject of debate.

As mentioned above, I view cognitivism as a family of theories with fairly minimal commitments, requiring just that ethical sentences are semantically like any other ordinary descriptive sentence, and that ethical claims express beliefs. Accordingly, cognitivism all on its own does not provide an explanation of the distinctive features of ethical thought and discourse. *Pure* cognitivism, we can say, outright denies that such an explanation is needed, by denying that ethical thought and discourse really is different in any significant way from other factual domains that do not exhibit motivation, action-guidingness, or disagreement. Since I think the distinctive features of the ethical domain do call out for explanation, I shall set aside pure cognitivism here. In the next chapter, I consider the prospects for purely *expressivist* views, which are anti-realist, and hold that ethical claims serve to express motivationally-charged non-cognitive attitudes. I argue that extant versions of pure expressivism are unable to account for all the various continuities identified in this chapter. Moreover, I point out that a certain mentalist metasemantic thesis is a necessary commitment of pure expressivism. Insofar as we would like to avoid commitment to this controversial thesis, and insofar as we want a more satisfactory explanation of the continuities between ethics and other domains, we have reason to consider alternatives to pure expressivism.

In Chapter 3, I consider the prospects and limitations of more recent *hybrid* theories. Hybrid theories integrate elements of cognitivism and non-cognitivism to yield a complete account of ethical thought and discourse that successfully explains each of the features discussed in this chapter. I argue, however, that extant hybrid theories are problematically limited because they elide the important distinction (identified in Bar-On's neo-expressivist theory) between expression in the semantic sense, and expression in the act sense (Bar-On, 2004, *inter alia*). In Chapter 4, I defend a hybrid version of ethical neo-expressivism, and supplement this with a hybrid account of ethical assertion. In Chapter 5, I continue to develop my hybrid view by situating it in a Millikanian biosemantic framework. While the neo-expressivist and biosemantic frameworks employed at crucial stages of this dissertation are unconventional in the standard metaethical literature, I argue that this is no disadvantage, since adopting these frameworks puts some metaethical issues in a new light, offering new paths forward. The view I defend in Chapters 4 and 5 is a hybrid version of cognitivism that is consistent with realism, and that views ethical claims as having a particular kind of social coordination function. This view accounts for truth-aptness, embeddability, belief-expression, justification-aptness, motivation, and action-guidingness. Chapter 6 completes the theory of ethical thought and discourse by providing an account of moral disagreement, and the possibility of moral knowledge.

Chapter 2

Traditional Expressivism

2.1 Introduction

In the previous chapter, I outlined various features of ethical thought and discourse to be explained, and I briefly described some prominent families of metaethical views and how they purport to account for these various features. The purpose of this chapter is to examine one family of views, those falling under the label ‘expressivism’, at greater length. I am interested in the expressivist position because I think the expressivist’s account of the distinctive features of ethical thought and discourse – in particular, the expressivist explanation of the connection to motivation – provides important insights worth preserving. A theory that does not recognize the expressive dimension of ethical claims and how it relates to motivation misses something important about the purpose of ethical thought and talk.

However, as we shall soon see, various attempts to develop this insight into an adequate theory have difficulty in accounting for the semantic and cognitive continuities. This can create the impression that we must either give up on the expressivist insight about the connection to motivation or else deny the cognitivist appearances of the ethical domain. In Chapters 3, 4, and 5, I will be arguing that we are not forced to choose, because we can retain an expressivist (or at least an expressivist-style) explanation of the connection to motivation while rejecting some problematic commitments of traditional expressivism. So, while I am sympathetic to some of the traditional expressivist’s ideas, the goal of the present chapter is to critically discuss the traditional expressivist position and articulate what aspects of the view I find problematic.

In order to make good on the strategy in this dissertation – i.e. retaining the expressivist explanation of the connection to motivation while rejecting certain problematic expressivist commitments in the philosophy of language – we must first decompose the traditional expressivist theory into its logically independent commitments. In what follows, I introduce the traditional expressivist position by articulating its component commitments. I also review some prominent criticisms of traditional expressivism in the literature, including the well-known Frege-Geach problem, with a focus on the negation problem (Unwin, 1999, 2001), and the problem of fundamental moral error (Egan, 2007). The goal is to get the traditional expressivist position into focus and to identify the most pressing difficulties for it.

A challenge in characterizing the expressivist position is that expressivism's defenders (and some of its critics)[1] have continued to refine and develop the view in response to various problems, so that it is not entirely clear exactly how expressivism should be understood. For instance, at one point, it was common to characterize expressivism as denying that ethical claims are truth-apt at all, and yet contemporary expressivists, under pressure to accommodate the surface cognitivist appearances of the moral realm, now readily grant that ethical claims are truth-apt. Expressivists have also not always been entirely clear on some core aspects of their view; for instance, concerning how the notion of 'expression' is to be understood, or whether expressivism is best understood as a semantic or a metasemantic view about moral discourse.[2] Below I lay out the core commitments what I call the Traditional Expressivist view (TE), of the sort defended by Blackburn and Gibbard.

[1] I have in mind Schroeder (2008a), who develops a detailed expressivist semantic program, for the purpose of showing the controversial commitments it must undertake.

[2] Regarding the issue of 'expression' for expressivists, see Schroeder (2008b, and 2008a, Chapter 2). See Ridge (2014, pp. 102-107) for discussion of expressivism as a semantic vs. a metasemantic view, and see Schroeter and Schroeter (2019) for discussion of the relation of metasemantics to metaethics.

2.2 Traditional expressivism: Claims and commitments

TE is the view that the meanings of ethical claims and sentences are in some way determined by the non-cognitive mental states that such claims and sentences characteristically express. The relevant type of non-cognitive state may be akin to attitudes of approval or disapproval (as proposed by emotivists), or planning states (Gibbard 2003), or states of being for (Schroeder, 2008a), depending on the particular expressivist view – at any rate, such states are what Hume would have called a 'passion', as opposed to an exercise of reason. Along with this non-cognitivist view about the mental states expressed by ethical claims, TE also takes an anti-realist stance about moral metaphysics, either denying that there are any genuine moral facts or properties, or else holding that they have a different, less 'substantive' metaphysical status from the facts and properties found in realist domains.[3] The traditional expressivist picture, then, can be decomposed into four logically independent theses:[4]

1. A positive expressivist thesis (expression): ethical claims characteristically express motivationally-charged non-cognitive mental states or attitudes.
2. A positive constitutivist thesis (internalism): To count as sincerely making a claim or judgment in the domain of ethics, an agent must be appropriately related to a motivationally-charged non-cognitive state.
3. A negative metaphysical thesis (anti-realism): Either there are no moral facts or properties for ethical claims to report on or describe, or else such moral facts or

[3] Early emotivist versions of expressivism held to the strict denial that there are any moral facts or properties. Contemporary expressivists now acknowledge that there may be moral facts or properties, but continue to insist on the rejection of moral realism. See Gibbard (2003, 2006, 2012, 2015).

[4] This decomposition of the expressivist view is modeled on the decomposition offered in Bar-On and James Sias (2013, p. 702). I have adapted Bar-On and Sias's decomposition to apply specifically to *ethical* expressivism, and to reflect innovations adopted by the strongest contemporary versions of ethical expressivism.

properties are in some way metaphysically less substantive than the entities reported upon or described by paradigmatically cognitivist thought and discourse.

4. A negative semantic thesis (anti-representationalism): claims in the domain of ethics do not have a descriptive representationalist semantic function.

The first two commitments, of course, lead to the expressivist explanation of the connection to motivation that I find promising. These commitments together with commitments (3) and (4) constitute the traditional expressivist position TE. Note, however, that (3) and (4) are not themselves entailed by the first two commitments of traditional expressivism. This reveals the logical coherence of views accepting the expressivist account of the connection to motivation while rejecting the expressivist's anti-realist and anti-representationalist commitments. Again, I go on to discuss such 'hybrid' views starting in the next chapter. Here, I consider the prospects for expressivist views that accept each of 1-4.

As I've said, claims (1) and (2) above do not strictly entail (3) and (4). Still, 1-4 seem to come together as a package given certain further assumptions in the philosophy of language and of mind. In particular, if we add to (1) and (2) the following:

5. A mentalist semantic assumption: the semantic content of a sentence is determined by the mental states that sentence characteristically expresses.

and

6. A Humean psychological picture: beliefs and belief-like states that represent the world as being a certain way are essentially distinct from desires and desire-like states that represent some way the world is to be made. No single mental state can have both belief-like and desire-like elements at once.

then (3) and (4) are natural additions, if not strictly entailed. The positive expressivist thesis (1) together with the mentalist semantic assumption (5) entails that the semantic content of an ethical sentence is the distinctive non-cognitive state it is used to express. And given that the expressivist proposal that these non-cognitive states fall on the 'desire' side of the Humean distinction in (6), it seems ethical sentences could not have the semantic function of describing or representing some moral way the world is. This leads to the negative semantic thesis (4). Given that ethical claims and judgments do not even purport to represent some moral way the world might be, the anti-realist thesis (3) is a natural further commitment. It is still not strictly entailed, but it is unclear why anyone would endorse a purely non-cognitivist view of ethical thought and discourse together with robust moral realism – that would be to say that there are genuine moral facts, but that our moral thought and talk puzzlingly do not even purport to be about such facts.

The Humean psychological picture is generally assumed to be true in discussions of expressivism: after all, it figures in prominent arguments against ethical cognitivism, on the way to motivating expressivist theories in the first place (see discussion in 1.4.1). Without the Humean assumption, it would remain open for cognitivists to account for the connection to motivation and action-guidance by appeal to a motivationally-charged non-cognitive *aspect* to moral judgment, compatibly with also holding that moral judgments represent moral features of the world. The Humean psychological picture is an important piece of the initial motivation for traditional expressivism.

The mentalist assumption (5) turns out to be highly significant in the context of assessing the expressivist project. As a matter of historical fact, expressivists have certainly seemed happy to endorse the mentalist assumption. For instance, Gibbard writes that an important element of the 'norm-expressivist' analysis he offers is "its claim that the meaning of normative terms is to be given by saying what judgments normative statements express – what states of mind they express" (1990, p. 84). And it may appear that expressivists have good grounds for endorsing the mentalist assumption. First, the mentalist assumption, though controversial, is arguably motivated by considerations in the philosophy of language independent of the metaethical concerns driving ethical expressivism: thus, at the very least, appealing to the mentalist assumption is not simply an *ad hoc* move by expressivists.[5] Second, the mentalist assumption seems to play an important role in enabling expressivism to capture the motivating and action-guiding character of ethical thought and discourse while avoiding devastating objections to superficially similar subjectivist alternatives to expressivism.

Simple subjectivism (introduced in 1.5.4) is like traditional expressivism in at least two respects: First, it takes very seriously the need to account for the motivating and action-guiding features of ethical thought and discourse. Simple subjectivism accounts for this by proposing that ethical claims *report on* the speakers motivationally-charged states (e.g. of disapproval): When S utters "Torture is morally wrong" in the course of making an ethical claim, simple subjectivism maintains that the propositional content of S's claim is that S disapproves of murder. Second, simple subjectivism avoids commitment to a purportedly problematic moral metaphysics, since it proposes to fully account for the content of ethical claims just in terms of speaker's motivationally-charged attitudes, without appeal to a stance-independent realm of moral facts or properties to serve as the semantic values of moral sentences and terms. However, simple subjectivism faces two major challenges that expressivism avoids by adopting the mentalist assumption.

[5] For a classic articulation and defense of a mentalist approach to linguistic 'nonnatural' meaning, see Grice (1957). See also Davis (2002) for a sustained recent development of a mentalist approach to meaning.

First, there is the problem of lost disagreement (see 1.5.4). Intuitively, it seems possible to *disagree* about moral matters: for instance, where S asserts “Torture is always morally wrong” and T asserts “Torture is not always morally wrong”. However, if simple subjectivism is correct, the content of S’s assertion would be: that S always disapproves of torture. And the content of T’s assertion would be: that T does not always disapprove of torture. But there is no semantic contradiction between these contents – so wherein lies the intuitive disagreement between S and T?

Subjectivists might appeal to some expressivist resources – in particular, Stevenson’s (1937) notion of ‘disagreement in attitude’ – to recapture the disagreement. The idea would be that S and T disagree, not by putting forward as true propositions with logically incompatible content, but rather by expressing conflicting attitudes. However, this does not really solve the problem for the subjectivist. For another way in which T can intuitively disagree with S’s assertion is by claiming “What S said is false”. If subjectivism is correct, S’s assertion is true just in case S always disapproves of torture, and so T’s denial (“What S said is false”) is true just in case S does not always disapprove of torture. But this is certainly not what T would take herself to be disagreeing about: T is not plausibly understood as denying *that S always disapproves of torture*. Simple subjectivism runs into this problem because it conflates the truth conditions of “Torture is wrong” with those of “I disapprove of torture”. Expressivism improves on simple subjectivism by avoiding this conflation: an utterance of “Torture is wrong” *expresses* one’s attitude of disapproval towards torture, but expressing this attitude is very different from *saying that* one has it.[6]

[6] This is the *deep* disagreement problem for simple subjectivism. See Schroeder (2008a, p. 17).

Second, there is the problem of making moral knowledge too easy (1.4.1). If ethical claims simply reported on the speaker's attitudes, then a speaker could acquire moral knowledge – i.e., knowledge that some moral claim or judgment is *true* – just by knowing her own attitudes. Again, because expressivism does not conflate the truth conditions of "Torture is wrong" with "I disapprove of torture", mere knowledge of one's own attitudes does not put one in a position to acquire moral knowledge.

The expressivist solution to these problems for simple speaker subjectivism consists in the contention that the meaning of the sentence "Torture is morally wrong" is given by the non-cognitive mental state an utterance of that sentence would serve to express, without being truth-conditionally equivalent to a self-ascription of such a state. Thus, the mentalist assumption plays a role in establishing an advantage for expressivism over its simple subjectivist competitor. However, it is important to note that the expressivist's endorsement of the mentalist assumption, though *sufficient* for avoiding the pitfalls of simple subjectivism, is not *necessary* for it. One can recognize that ethical claims have an important expressive function and connection to motivation, and avoid conflating moral claims with self-ascriptions of motivationally-charged attitudes, *without* assuming that the literal meaning of moral sentences are determined by the mental states they can be used to express. To avoid the problems for simple subjectivism, we must reject thinking that ethical claims are really psychological self-ascriptions in disguise. We do not also need to suppose that ethical claims get their literal semantic contents from the non-cognitive states of mind they express. This is so even if it is the case that an important feature of ethical claims is that they do express non-cognitive states. I discuss this point in depth in Chapter 4; it is key to the neo-expressivist approach (due to Bar-On, 2004; Bar-On and Chrisman, 2009; Bar-On, Chrisman, and Sias, 2014) I endorse for retaining an expressivist-style explanation of the connection to motivation and action-guidingness while avoiding the Frege-Geach and related problems. For now, my concern is to critically examine expressivist proposals that do rely upon the mentalist assumption to avoid the challenges to simple subjectivism.

The mentalist assumption proposes to handle questions about the literal semantic meanings of sentences in terms of the states of mind those sentences can be used to express. Given the expressivist endorsement of the mentalist assumption, this would appear to make providing an account of the expression relation a primary concern for expressivist accounts of the meaning of ethical claims. However, some expressivists – such as Gibbard (2003) – have avoided providing such an account in detail. As Schroeder explains (2008a, pp. 17-18):

> The correct view, according to expressivism, is that 'murder is wrong' *stands to* disapproval of murder in the *same way as* 'grass is green' stands to the belief that grass is green. Notice that the beauty of this solution [to the challenges for simple subjectivism] is that it requires taking no view at all about what this relationship is. Whatever this relationship turns out to be, the expressivist will say, it must be adequate to explain why there is no modal or disagreement problem for 'grass is green'

So, the thought is, expressivists do not need to provide an account of the expression relation that holds between moral claims and the states of mind they are supposed to express, since they can rely on whatever account turns out to be correct for how ordinary descriptive sentences can be used to express belief. Ethical sentences are not supposed to differ from ordinary non-moral ones in *having* their meanings determined by the states of mind they can be used to express, but rather they differ in the *nature* of the mental states they express.

Unfortunately for traditional expressivism, as Schroeder (2008a, b) has compellingly argued, traditional expressivists do need to commit to a view about the expression relation. This is no trivial requirement, either: as Schroeder goes on to argue, no matter which of a variety of candidate accounts of the expression relation the expressivist endorses, she must "take on an extraordinary range of quite radical commitments" (Schroeder 2008a, p. 15), some of which I discuss just below.

Let us grant the mentalist assumption and treat it as a simple first-order semantic thesis: the content of a given sentence just is the mental state it is used to express. Since the semantic assumption is a general claim about language – i.e. it does not apply *only* to normative discourse – we have it that the semantic content of the sentence "Snow is white" (an ordinary descriptive sentence, amenable to cognitivist treatment if anything is) just is the mental state it expresses: presumably, the belief that snow is white. Ethical expressivists then add that for moral sentences, e.g. "torture is morally wrong", the mental state expressed is not a belief, but something like *disapproval of torture*: so this is the content of the moral sentence. It does not seem to matter, then, what account of expression we adopt; the hope is that whatever account explains how "snow is white" expresses the belief that snow is white will do equally well for moral sentences.

Difficulties arise when we begin to consider the meanings of simple ethical sentences when they appear embedded in complex constructions – e.g., "If torture is always wrong, then what Cheney did was wrong".[7] For an embedded atomic descriptive (non-moral) sentence we want to say that the meaning of the complex sentence is a function of the meaning of its parts. Given mentalism, this amounts to thinking that the mental state expressed by the complex sentence is a function of the mental states expressed by its parts. So the mental state of believing that snow is white is supposed to figure in determining the content of the mental state of believing that snow is white or grass is green. This appears to pose no real difficulties, because the belief that snow is white has propositional content (namely, *that snow is white*), and we know how functions on propositional contents yield other propositional contents. The trouble is that

[7] What follows is my interpretation of Schroeder (2008a, Chapter 2; 2008b). See also Camp (2019).

this explanation relies upon the point that beliefs derive their contents from the propositions they are about. This reverses the order of explanation offered by the mentalist assumption as the expressivist must understand it. For expressivists do not have recourse to an antecedent account of the contents of ethical judgment. If expressivists wanted to appeal to an antecedent understanding of the propositional contents of ethical judgment at this stage, the detour through mentalism about semantic meaning would begin to look like a distraction, and we might wonder what real advantage is gained over cognitivism (see Schroeder 2008a, pp. 22-25; 2008b, pp. 95-97).

What this illustrates is that the traditional expressivist's commitment to mentalism constrains what account of the expression relation she can adopt. We have seen that expressivists *cannot* say that the propositional content expressed by a sentence is derived from the propositional content of the mental states expressed in uttering that sentence. In the next section, I review some candidate accounts of the expression relation, and summarize some of the core difficulties for each such account. In the end, I agree with Schroeder that traditional expressivism should say something about the expression relation, and that some existing accounts of the expression relation are not suitable for the expressivists needs. Schroeder thinks that, given the constraints that the mentalist assumption places on how the expressivist can understand the expression relation, the expressivist's best bet is to endorse Schroeder's assertability account of the expression relation, and an associated assertability semantics. Perhaps this is the best bet for traditional expressivism. However, I shall be interested, in Chapter 4, in what happens when we drop the assumption that there is a *single* expression relation of interest to expressivists.

2.3 'The' expression relation

In this section, I articulate some candidate accounts of the expression relation that expressivists have endorsed or could endorse, and then indicate what commitments these accounts require given prior commitment to the mentalist assumption. Candidate accounts of the expression relation include (i) the 'same-content' account, (ii) the causal account, (iii) the intention-based account, (iv) the implicature account, (v) the assertability account, and (vi) the accountability account.[8]

> *The same-content account*: a sentence expresses a mental state when they have the same content (Schroeder, 2008a, p. 24; see also 2008b, §3.1).
>
> *The causal account*: a sentence expresses a mental state when that mental state causes the production of the sentence (Schroeder, 2008a, p. 25; 2008b, §3.2).
>
> *The intention-based account*: a sentence expresses a mental state when the speaker intends, by means of uttering the sentence, to indicate to her audience that she is in that mental state (Gibbard, 1990, p. 85; Schroeder, 2008a, p. 26; 2008b, §4.1).
>
> *The implicature account*: a sentence expresses a mental state when linguistic conventions governing the use of that sentence are such that an utterance of that sentence carries the implicature that the speaker is in that mental state (Copp, 2001, 2009, 2014; Barker, 2000; Schroeder, 2008b, §3.3).
>
> *The assertability account*: a sentence expresses a mental state when being in that mental state is a condition on the proper assertion of that sentence (Schroeder, 2008a, pp. 28-34; 2008b, §5).

[8] Notice that each account presupposes that the relata of 'the' expression relation are mental states, on the one hand, and sentences, on the other. This presupposition will also be challenged by the neo-expressivist account discussed in Chapter 4.

> *The accountability account*: a sentence expresses a mental state when linguistic conventions governing the use of that sentence are such that a speaker is *liable* for being in that mental state when asserting that sentence (Ridge, 2014, p. 109).[9]

There is reason to be skeptical of each of these accounts, either because they are independently implausible accounts of the expression relation, or because they cannot serve the needs of ethical expressivists. In what follows, I briefly indicate some challenges for each account, following Schroeder (2008b).

Against the same-content account: Expressivists cannot rely upon this account given their endorsement of the mentalist assumption. According to the mentalist assumption, sentences *acquire* their contents from the contents of the mental states they express – but the same-content then has the wrong order of explanation, for it requires an antecedent account the contents of sentences and of mental states (this is related to the problem canvassed at the end of 2.2).[10]

Against the causal account: Combining the causal account of the expression relation with mentalism is problematic insofar as the causal account treats 'expression' as a factive notion, such that one does not count as genuinely expressing one's mental state M unless one is in M. To illustrate: suppose that Alex sincerely expresses her belief that Joe is in Barcelona by uttering the sentence "Joe is in Barcelona". Now suppose that Billie lies, asserting "Joe is in Barcelona" when Billie does not in fact believe that Joe is in Barcelona. The sentence "Joe is in Barcelona" should have the same content, regardless of whether it is used to genuinely express a belief or if it is used to lie. But given that expression is a factive relation and given the mentalist

[9] I have simplified Ridge's statement of the accountability account. Ridge states his account as follows: "A declarative sentence 'p' in sense S in a natural language N used with assertive force in a context of utterance C expresses a state of mind M if and only if conventions which partially constitute N dictate that someone who says 'p' in sense S in C with assertive force is thereby *liable* for being in state M" (2014: 109).

[10] Jackson and Pettit (1997, pp. 244-245); Schroeder (2008a, pp. 24-25; 2008b, §3.1).

assumption, it turns out that the content of Alex's utterance is *that Joe is in Barcelona* (for that is the content of the belief expressed), but Billie's utterance cannot have that same content, because Billie does not believe that Joe is in Barcelona, and so cannot express that belief by uttering "Joe is in Barcelona". As Schroeder argues, this places an important constraint on how traditional expressivists can understand the expression relation, namely that "it must be possible to bear [the expression relation] to a mental state that the speaker is not in" (2008a, p. 26).[11]

Against the intention-based account: Schroeder argues that taking the linguistically articulate expression of a mental state to be dependent upon a speaker's intention to express that mental state, together with the mentalist assumption, yields implausible predictions about the contents of certain utterances. To illustrate the problem, suppose I utter the sentence "I believe that it is raining", with the intention of thereby indicating to my audience that I believe that it is raining. The intention-based account predicts that my utterance expresses my belief that it is raining, since that is what I intend to convey in making that utterance. The mentalist assumption then tells us that my utterance gains its content from the content of the mental state it is used to express: so we seem to have it that the content of my utterance is *that it is raining*. But, Schroeder suggests, this intuitively seems wrong: we would expect the content of my utterance of "I believe it is raining" to be *that I believe it is raining,* not *that it is raining*.[12]

I must confess that I do not find Schroeder's criticism here compelling, at least if it is supposed to be a problem for the intention-based account in its own right. It does not seem strange to me that ordinary utterances of "I believe it is raining" could simultaneously express

[11] It is also worth pointing out that even given mentalism, expressivists may find resources to avoid this problem. For instance, Bar-On distinguishes between expressing *one's* M and expressing *M* (2004, pp. 280-281); the former is a factive notion, while the latter is not. This could be a useful innovation for traditional expressivists to adopt, given that the problem outlined in this paragraph rests on the assumption of a purely factive expression relation.

[12] This objection takes inspiration from van Roojen (1996), Schroeder (2008a, pp. 27-28; 2008b, §4.1).

one's belief that it is raining, as well as one's belief that one believes that it is raining.[13] Indeed, expressivists in other areas readily grant this: for example, Bar-On's neo-expressivist theory of avowals explicitly proposes the Dual-Expression Thesis, according to which avowals simultaneously express the avowed mental state, as well as the avower's belief that she is in that state (2004, pp. 307-310). What *is* problematic is combining this observation with a straightforward application of the mentalist assumption; for that combination would indeed make implausible predictions about the contents of ordinary claims, such as that an ordinary utterance of "I believe that it is raining" would somehow have, as two distinct literal contents, *that it is raining*, and *that the speaker believes it is raining*.

In addition, the intention-based account seems implausible as an account of expression, independently of any awkward consequences it may have in combination with the mentalist assumption. This is because it seems implausible to assume that in any case that one uses a sentence to express one's mental state one must have the intention to indicate to an audience that one is in that mental state. The simplest sort of problem case will be those where our utterances are not directed at an audience at all, as when I talk to myself in the car, verbally articulating my thoughts without any audience in mind. Considerations like this one lead Bar-On to propose that expressive behavior, though intentional, does not require the intention *to express* one's state of mind (see, e.g., Bar-On and Chrisman, 2009, p. 136). Bar-On also points out that non-human animals, many incapable of harboring the sort of complex Gricean communicative intentions at work in the intention-based account, nevertheless are intuitively capable of expressing mental states (see Bar-On, 2013). The accounts of the expression relation under consideration here all explicitly focus only on instances where *sentences* are used to express mental states, and so one

[13] Schroeder himself seems to acknowledge this in a footnote (2008a, fn. 9), citing Ayer's point that an utterance of "I am bored" can simultaneously count as *expressing* my boredom and reporting on it.

might be tempted to simply set aside non-linguistic expressive behavior. I think that would be a mistake: although there are undoubtedly important differences between expression by non-linguistic vs. linguistically articulate means, we run the risk of losing out on important insights about expression in general if we focus just on expression in language.

Against the implicature-based account: This account is simply not suitable for combination with the mentalist assumption. This is just because implicature is a pragmatic notion, whereas the mentalist assumption has been construed as a semantic proposal. When an utterance U carries the implicature that P, the implicature content (that P) is *additional to* or somehow *distinct from* U's truth-conditional semantic content, not constitutive of it.[14]

There is an additional concern about the intuitive notion of expression and the implicature contents attributed to ethical claims. In general, when an utterance carries an implicature, the implicature content is propositional in nature. For instance, an utterance of "LeBron is big but fast" typically carries, as an implicature, the propositional content *that there is a contrast between being big and being fast*, in addition to its ordinary truth-conditional content (i.e. that LeBron is big and that LeBron is fast). To be useful to the expressivist, the implicature-based account will need to identify a connection to motivationally-charged non-cognitive attitudes, and a natural way to do so is by proposing that ethical claims carry, as implicature content, a proposition to the effect that the speaker is in the appropriate sort of non-cognitive mental state. As a toy example, consider an implicature-based account according to which a sincere and competent utterance of "torture is wrong" carries the implicature *that the speaker disapproves of torture*.

[14] Grice introduces the concept of implicature along these lines in his seminal paper "Logic and Conversation" (1975), by considering cases where the meaning of what someone strictly says is different from what they mean to communicate.

Expressivists, recall, were supposed to depart from subjectivists in maintaining that ethical claims *express* certain non-cognitive states rather than *stating that* the speaker has them. Implicature-based expressivism does not construe us as making *statements* about our mental states in making ethical claims, and so it rightly can claim distance from simple subjectivism. But the view nevertheless misses a second contrast between expressing a mental state and saying that one is in it: Namely, that the latter (reportive case) is propositional in character, being essentially connected to the proposition that the speaker is in the relevant mental state, whereas the former (expressive case) is not necessarily propositional. If one properly expresses one's M, it must be true that one is in M, to be sure: but what gets expressed, intuitively, is *one's M*, not the proposition that one is in M. Since the implicature-based account relies on the notion of implicature, which relates an utterance to a propositional content, the result is that the implicature-based account only qualifies as an expressivist view on the first way of contrasting expressivism with subjectivism.

This may not seem like such a problem. It might be thought that the whole *point* of ethical expressivism is just to identify a linguistic relationship other than truth-conditional equivalence between ethical sentences and propositions about the non-cognitive mental states of individuals. But this way of thinking obscures the possibility of an intuitive understanding of a kind of expression relation that is first and foremost a relation between a minded individual, her mental states, and an expressive vehicle. I discuss this notion of expression – key to the *neo-expressivist* theory developed by Bar-On – in Chapter 4. I further discuss the notion of implicature in Chapter 3, where I critically assess the prospects for theories that incorporate implicature as a tool for explaining the connection to motivation and action-guiding features of ethical thought and discourse.

Against the assertibility and accountability accounts (I treat these together, given their similarities): I briefly raise two issues. Let me first emphasize again that the capacity to express one's mental state is not a uniquely human trait, nor does it require sophisticated language abilities, as reflection on instances of animal expressive behavior illustrate. This can be easy to lose sight of in discussions of ethical expressivism, since in that context expressivism is often taken as a thesis about the semantics of a portion of (human) natural languages and concerns an area of inquiry (ethics) that itself is arguably a uniquely human concern. But it is essential to keep in mind the continuities between human and non-human, and between linguistic and non-linguistic expressive behavior, to avoid constructing an over-intellectualized account of what expressing one's mental states requires. We saw the importance of this above in the criticism of the intention-based account of the expression relation. A similar criticism applies here. But the most crucial point to consider is that, when taken in combination with the mentalist assumption, endorsing these accounts requires taking on controversial commitments about linguistic meaning. The propositional content expressed by an utterance will be determined by the mental state(s) one is supposed to be in, or liable to be in, when making that utterance, in virtue of the assertibility conditions on that type of utterance. In short, what is needed is an *assertibility semantics*. As Schroeder (2008a, pp. 32-34) points out, this is at the very least not obvious that assertibility semantics is correct, yet expressivists that accept the assertability or accountability accounts of expression would be committed to its truth.

The traditional expressivist may be happy to bite the bullet here and endorse assertability semantics. After all, assertability semantics does enjoy some support. For instance, it could figure into an explanation of the intuition (if it is indeed intuitive) that we have direct cognitive access to the contents of our claims. When we combine the assertibility/accountability account with the mentalist assumption, we have it that the content of a sentence would ultimately be given by the conditions under which it would be appropriate to assert that sentence. Given (i) a norm of belief on assertion, and (ii) that one has especially epistemically secure and direct knowledge of one's own beliefs, it would turn out that we have especially epistemically secure and direct knowledge of the assertability conditions and thus (given mentalism) the contents of sentences of our language. And the assertability/accountability account of the expression relation does seem to improve upon the other accounts considered in just the ways needed to avoid the difficulties listed. For these reasons, I shall understand traditional expressivism as accepting some version of assertability semantics and the assertability/accountability account of expression.

In the next section, I proceed to highlight some of the challenges for traditional expressivism. In Chapter 1, I found it useful to distinguish between the *semantic* continuities between ethics and other domains, and the *cognitive* continuities. I find a similar distinction useful here. I divide objections to traditional expressivism into two categories; those highlighting the semantic continuities, and those highlighting the cognitive continuities. While much of the criticism of expressivist theories, including the famous Frege-Geach problem, center around the expressivist's (in)ability to capture the semantic continuities, I think that perhaps the deeper problems for expressivism have to do with explaining the cognitive continuities, as illustrated in the problem of fundamental moral error (Egan, 2007). Even if expressivists can ultimately solve the semantic problems discussed in the next section, we still have reason to be skeptical that expressivists can overcome the epistemic challenges of accommodating the cognitive continuities discussed in section 2.5.

2.4 Accounting for the semantic continuities

As I have flagged in this and the previous chapter, the most immediate task for expressivism is to account for the apparent semantic continuities between ethics and other factual domains. We now turn to consider this problem, which in the literature most commonly goes under the heading ‘the Frege-Geach problem’ (owing to Geach, 1960, 1965), but which also encompasses certain other issues, such as the negation problem (Unwin, 1999, 2001) and various problems concerning moral inference and validity, to be discussed in section 2.5. I present the basic versions of these problems, and then assess some expressivist responses.

2.4.1 Truth-aptness

The simplest version of the difficulty for traditional expressivism, is just that at least some (early) iterations of traditional expressivism appear committed to denying the platitude that ethical sentences are truth-apt. Expressivism views the literal meanings of ethical claims as determined in some way by the non-cognitive states those sentences are said to express. But, at least on early versions of expressivism, those types of state are not assessable for truth or falsity. Where beliefs are paradigmatic bearers of truth, boos and hurrays, plans, states of approval and disapproval, and so on, are not.

Early emotivists saw the denial of truth-aptness for ethical claims as a feature of their view, rather than a bug. (Ayer, for instance, was happy to charge that ethical claims are strictly meaningless, according to his verificationist standard for meaning). But contemporary expressivists are concerned to preserve the surface appearances of ethical thought and discourse. So expressivists must provide some explanation of how ethical sentences could be truth-apt even though the mental states those sentences express and from which they derive their meanings are non-cognitive.

2.4.2 Embeddability

Ethical sentences appear semantically unremarkable in that, just like any other ordinary declarative sentence, they can be meaningfully embedded in a variety of contexts, including in conditionals, negation, propositional attitude reports, and so on. Moreover, ethical sentences, like any ordinary sentence, must make the same contribution to the meanings of any complex construction in an extensional context in which they can be embedded. One of the most significant constraints on any adequate account of semantic content is that of *compositionality*: it must turn out that the meanings of complex linguistic items are systematically related to the meanings of their parts. This applies to ethical sentences just as much as any other kind of sentence.

This leads us to the original Frege-Geach problem (as presented in Geach, 1960, 1965), which turns on what Geach calls the 'Frege point' that: "A thought may have just the same content whether you assent to its truth or not; a proposition may appear in a discourse now asserted, now unasserted, and yet be recognizably the same proposition" (1965, p. 449). As mentioned just above, any adequate theory of meaning must recognize that the content of a linguistic item remains constant across a variety of contexts – asserted and unasserted, free-standing and embedded, etc.; it must recognize that there is some content shared by the assertion "it is raining" and the question "is it raining?", even though these two utterances are standardly used to accomplish different speech acts.

Traditional expressivism, we have seen, associates the meaning of ethical sentences with the non-cognitive mental states those sentences are said to express. The basic problem for traditional expressivism here is that it is only when ethical sentences are *asserted* that they seem to express the meaning-giving non-cognitive states. So, it seems utterly mysterious how expressivism can account for the Frege point. To illustrate: the expressivist tells us that an assertion of “Lying is wrong” expresses a non-cognitive state M, and that M gives the meaning of that sentence. But the sentence “Lying is wrong” is *not* asserted when it appears in (for instance) a conditional, such as: “If lying is wrong, then getting your little brother to lie is wrong”. And so, intuitively, that conditional sentence does *not* express the non-cognitive state M, even in part. Since the mental state M expressed by the assertive utterance of “Lying is wrong” was supposed to give the meaning of that sentence, and M is not part of what is expressed by an utterance of the conditional sentence, it seems that “Lying is wrong” cannot have the same meaning when it appears embedded in the conditional sentence as it does when it is asserted on its own. And this flatly contradicts the Frege point.

Below, I present two strategies expressivists have adopted to account for the semantic continuities. In essence, the challenge for the traditional expressivist is to explain how to give a semantics for ethical sentences given the four expressivist theses discussed above, in conjunction with the mentalist and the Humean assumptions. Whatever semantics expressivists provide for ethical sentences, it must, at the very least, account for the seamless semantic integration of ethical claims with claims in other cognitivist domains.

2.4.3 The deflationist strategy

Starting in the early 1990’s,[15] some have thought that expressivists can appeal to deflationism about truth and truth-aptness in order to account for the semantic continuities between ethical discourse and other domains, without having to provide an entirely new semantic theory (such as an expressivist assertability semantics).[16] Call this the deflationist strategy

[15] This is actually not true. The innovation of applying deflationism about truth to rescue non-cognitivism from the Frege-Geach problem appears earlier in Smart (1984). Thanks to Bill Lycan for pointing me to this work in connection with the deflationist expressivist approach.

[16] Deflationism about truth and related semantic notions remains an important feature in contemporary discussions of expressivism, but it seems that deflationism’s enduring significance in metaethics has more to do with the alternative it provides to representationalist approaches to semantic notions expressivists are already keen to reject

(DS).[17] I now consider how deflationism about truth and related notions purportedly provides expressivists an easy way out of the Frege-Geach problem, and then why this easy out is ultimately unsatisfactory.

Deflationists about truth deny that there is any interesting answer to the question "what is the nature of truth?".[18] In so doing, deflationists deny that there is any substantive property of truth, some quality which all and only truths possess and in virtue of which they are true.[19] Rather, deflationists give an account of the predicate 'is true' that assigns it only an expressive and generalizing function. The guiding thought is that if deflationists can successfully account for our use of the truth predicate and concept without appeal to a substantive property of truth, we have no need to postulate such a property in the first place, obviating the need for a substantive theory of truth.[20] Now, clearly, the predicate 'is true' does have an expressive role: it enables us to express our commitment to or endorsement of some proposition. And it has a generalizing role; it enables us to endorse a whole slew of propositions at once, and to endorse propositions without knowing their content, for instance, when one asserts "What George said is true", not knowing what it was that George said. All that needs to be assumed for the truth predicate to play these roles are the various instances of the schema (Horwich, 1990, p. 6):

(E): It is true that p if and only if p.

as accurate to the moral realm, rather than because deflationism purportedly provides a 'fast track' for addressing the Frege-Geach problem.

[17] See Horwich (1993, 1994), Stoljar (1993), and Price (1994).

[18] For an influential defense of deflationism, see Horwich (1990).

[19] By 'substantive' property, I mean a property with an analyzable underlying nature. Instantiations of properties that are not substantive have nothing necessarily in common with one another other than that they instantiate that property.

[20] Note the similarity between the structure of the argument for deflationism, and that for expressivism in ethics. Expressivists argue that if we can successfully account for our use of ethical terms and concepts without appeal to substantive ethical reality, we have no need to postulate such a reality in the first place.

For then, instead of having to say "If what George said is 'snow is white', then snow is white; and if what George said is 'grass is green', then grass is green; . . .", one can simply say "What George said is true". In short: The truth predicate plays an important role for enabling blind and compendious assertion, and it does so by providing a tool for disquotation. What deflationists contend is that these facts about the truth predicate *suffice* to explain everything we want to explain about truth. *Inflationist* theories of truth do not need to disagree that the truth predicate has important expressive and logical functions, nor that there is any problem with the instances of the schema (E).[21] Rather, the point of contention is whether there is anything more to the concept of truth than the instances of the schema (E).

It will be useful to contrast the deflationist approach with a prominent competing inflationary approach: the correspondence theory of truth. According to correspondence theory, truth is not just an expressive device; it is a substantial property, one that relates truth-bearers to the world. The core correspondence intuition is just that, when a truth-bearer (such as a sentence, belief, proposition) is true, this is because it corresponds to some worldly *fact*. So, to say "It is true that grass is green", on this approach, is not just a way of endorsing the claim that grass is green; it additionally ascribes to that claim the property of corresponding to the fact that grass is green. This approach has the advantage of capturing the intuition that truth depends on reality (an intuition deflationists struggle to capture, since acknowledging such a dependence seems to assign an explanatory role to truth going beyond its use as an expressive device). However, when combined with a philosophical stance that only recognizes naturalistic entities as the inhabitants

[21] One thorny issue I set aside here is whether (or how) schema (E) applies when it comes to the semantic paradoxes. What should we say about sentences like "This sentence is false", for instance? Horwich articulates his theory so that it only recognizes the *uncontroversial* instances of schema (E), and so rules out paradoxical instances of (E). This seems rather unsatisfactory, since one might hope to have an *explanation* of what is defective about those instances. But a deflationist will likely want to respond that whatever is interesting and requires explanation about the semantic paradoxes will not ultimately have to do with the nature of *truth*.

of our world at the most fundamental level, correspondence theories face the challenge of locating the naturalistic truth-makers for certain claims. For instance, what would the natural world have to be like in order for it to be true that 2+2=4, or that Mitch Hedburg is funny? Inflationary views of truth (combined with naturalism) face ‘placement problems’: they have to somehow fit mathematical truths, truths about humor, ethical truths, and so on, into the natural world.[22] Deflationary views do not face this problem, for the simple reason that they do not view it as part of the job for a theory of *truth* to explain what the world has to be like for any particular claim to get to be true.

There are thus two main attractions of the deflationist approach for ethical expressivists. The first – our current focus – is that deflationism offers a potential fast-track to addressing the Frege-Geach problem (this is the Deflationist Strategy). The second is that deflationism offers a way to vindicate talk of moral truth, moral fact, etc., without realist metaphysical commitments – this is the *quasi-realist* strategy undertaken by Blackburn and Gibbard, which I return to in section 2.5.4. Following the deflationist idea that to call something true is just a way of endorsing it, there seems to be no substantial barrier to regarding ethical claims as truth-apt, even if they do not correspond (or not) to ethical facts. That is: we can hold, with expressivists, that ethical claims express (and get their meaning from) motivationally-charged non-cognitive mental states rather than (representational) beliefs, and also accept ethical claims as truth-apt.

There is an important additional wrinkle to this fast-track strategy: it is not enough simply for the expressivist to appeal to a deflationary theory of truth; she must additionally endorse a deflationary account of truth-*aptness*. Traditional expressivism maintains that “Stealing is wrong” and “Boo to stealing money!” have the same kind of meaning, since they each are

[22] See Price (2013) for a discussion of the placement problem in connection with expressivism.

supposed to express a motivationally-charged non-cognitive state. Thus, expressivism cannot explain why one of these sentences is truth-apt while the other is not simply in terms of their meanings. Accordingly, we are pushed to adopt a corresponding deflationism about truth-*aptness*, one that explains truth-aptness in non-semantic terms. In particular, the proponent of the deflationist strategy must hold that it is sufficient for a sentence's being truth-apt and having truth-conditions that it displays *syntactic discipline*, where for a sentence to display syntactic discipline just is for it to be "capable of significant embedding in negations, conditionals, propositional attitude operators and other subsentential constructions" and "subject to clear standards of appropriate and inappropriate useage" (Sinclair, 2006, p. 249). As Horwich (1994) puts the point: "Thus all it takes, from a minimalist perspective, to be truth-assessable, is having the appropriate syntactic and inferential properties" (pp. 19-20). Now, clearly, sentences like "Boo for stealing!" do not have the right syntactic structure to be truth-apt. But the sentences that would express an attitude of disapproval towards stealing, e.g. "Stealing is wrong", *do* have the right structure. After all, "Stealing is wrong" is parallel in syntactic structure with sentences like "Snow is white", "Grass is green", etc., which are paradigmatically truth-apt. In short, the idea behind the deflationist strategy is that the syntactic continuity we observe to hold between ethical sentences and other truth-apt sentences suffices to establish the truth-aptness of ethical sentences.

On the deflationist strategy DS, to judge that it is true that stealing is wrong is just to judge that stealing is wrong, where this judgment is understood to consist in having a certain non-cognitive attitude towards stealing. DS thus seems to account for truth-aptness. Moreover, since DS can hold that ethical sentences have (deflated) truth-conditions, and that ethical sentences express (deflated) propositions, it seems that DS can account for embeddability and respond to the Frege-Geach problem. For DS can say that the deflated propositions ethical sentences express are what those sentences contribute to the meanings of complex constructions in which they figure. This deflationist strategy generalizes, allowing the deflationist expressivist also to talk of ethical ‘fact’, ‘property’, ‘belief’, and so on, where these are each construed in a deflated way.

However, simply accepting deflationist conceptions of truth, truth-aptness, and related notions is not, on its own, sufficient to account for the semantic continuities. To highlight the difficulties here, consider the following problem for deflationist expressivism from Dreier (1996). According to DS, all that it takes for a sentence to be truth-apt is for it to exhibit certain syntactic properties. But, Dreier points out, these properties are too easy to come by. To show this, Dreier invites us to imagine that we introduce to English the predicate ‘hiyo’, which is stipulated to be used to accost others, as in “Hiyo, Bob!”. And Dreier further stipulates that, as a grammatical matter, “Bob is hiyo” is well-formed and can be used to accost Bob. In short, the rules introduced for ‘hiyo’ guarantee that, once the relevant linguistic conventions are adopted, hiyo sentences are syntactically well-disciplined, and so truth-apt according to DS (Dreier, 1996, pp. 42-44). But clearly, even if the relevant linguistic conventions were adopted, we would not understand what a sentence such as “If snow is white, then Bob is hiyo” means. Mere syntactic discipline plus deflationism about truth does not explain meaning.[23]

2.4.4 Sophisticated Expressivism

Gibbard, rather than going for deflationism about truth (at least in his earlier work)[24], instead undertakes the ambitious project of supplanting truth-conditional semantic analysis with

[23] See also Smith, Jackson, and Oppy (1994), who argue against minimalist conceptions of truth-aptness, thereby blocking the easy move from deflationism about truth and truth-aptness to capturing the semantic continuities between ethics and other cognitivist domains.

[24] Gibbard is open to deflationist truth, and is happy to make use of it to account for our tendency to describe ethical judgments as ‘true’ or ‘false’. But he avoids the mistake of using deflationism about truth and truth-aptness from the outset to give a ‘fast-track’ ‘solution’ to the Frege-Geach problem (see Gibbard, 2003, Chapter 4).

a mentalist alternative. Given that all semantic meaning is mentalistic, there is no difference in genus between the sort of meaning that descriptive sentences have (e.g. truth-conditional) and the sort of meaning ethical sentences have (expressivist). There is just a difference in species between the mental states expressed in uses of descriptive sentences and uses of ethical sentences. Since all meaning is explained in the same mentalist way, it is not so puzzling to see how the meanings of ethical sentences could contribute in a systematic way to the meanings of sentences of which they are a part.

However, if this strategy is to work, expressivists have to provide a model for the non-cognitive mental states expressed that attributes to them sufficient logical complexity to enable them to semantically integrate with the belief states expressed by utterances of descriptive sentences. So far, I have only construed the states ethical sentences are supposed to express as ‘motivationally-charged non-cognitive states’. But we must be more specific if these mental states are going to be semantically integrated with beliefs in such a way as to meet the compositionality constraint. For example, “Stealing is wrong” has to express a mental state with more logical complexity than simply mentally ‘booing’ stealing. For mere aversion to or disapproval of something (as mental states), lack the right structure to serve as constituents of the contents of truth-apt sentences. Relatedly, in order to avoid the problem Dreier illustrates with his ‘hiyo’ example, the semantic continuities cannot be accommodated simply by pointing out that ethical judgments have the appropriate syntactic form. It must be that the mental state the expressivist says constitutes ethical judgment has the right logical structure to be able to enter into logical relationships with other mental states in a way that ‘accostings’ do not. The key to showing that ethical judgments do enter into such relationships, Gibbard thinks, is showing that such judgments have the right structure to be able to *disagree* with other states (2003, pp. 65-71).

Gibbard's strategy, then, is to provide a complex account of the structure of normative judgment, such that these judgments and their constituent concepts fit into a compositional mentalist semantics that also encompasses descriptive judgment. The basics of this account are as follows.[25] Gibbard's norm-expressivist account holds that to judge something permissible (for instance) is to accept a system of norms that permit it (Gibbard, 1990, p. 26). And accepting this system of norms, in turn, amounts to accepting norms that rule out any combination of descriptive belief about the action with norms that forbid the action described by that belief. In later work, Gibbard re-works this analysis in terms of *planning* states. The idea is that thoughts about oughts are really planning states; to judge that one ought to phi in circumstances C is to plan, for the contingency of being in C, to phi (Gibbard, 2003, pp. 53-59). Much of the analysis hangs, then, on what it is to accept a system of norms, or what it is to be in a planning state.

Gibbard analyzes acceptance of a system of norms and of planning states using a variation on possible-worlds semantics. In particular, the idea is to explicate the contents of a mental state of *acceptance* in terms of the 'fact-plan worlds' that state does not rule out. These 'fact-plan worlds' are a combination of some possible way the world could be and a 'hyperdecided' plan about what to do in the case of being in the position of any given agent at any given time in that world (Gibbard, 2003, p. 57). So, for ordinary non-normative judgments (i.e. ordinary descriptive beliefs), Gibbard's analysis gives the content of the judgment (and therefore the content of the sentence that would express that judgment) just as a possible-worlds semantics would, albeit with additional machinery in place not relevant to the contents of

[25] I switch freely between Gibbard's characterization of the view in his earlier work (in *Wise Choices, Apt Feelings*, 1990) and in later work (*Thinking How To Live*, 2003). The former produces an analysis of normative judgment in terms of mental states of accepting systems of norms. The latter produces an analysis in terms of planning states. This should not cause problems, however, because my criticisms do not turn on the details of the analysis itself, and because Gibbard understands his later work "not as a change of position but as a shift of expository purposes" (2003, p. 181 fn. 3).

ordinary descriptive belief. When it comes to normative judgments, this additional machinery makes a difference to content: for normative judgments are analyzed not simply in terms of possible ‘factual’ worlds, but in terms of possible ‘fact-plan’ worlds. To judge “Stealing is wrong”, on this view, is (roughly) to accept a system of norms that rules out any fact-plan combination of a state describing some possible act A in situation S as one of stealing, and a state of planning in S to do A. The important point is that construing content in terms of mental states of permitting or rejecting fact-plan combinations is supposed to be “isomorphic with truth-functional ways of speaking”, and that fact-plan worlds “give us a way of displaying entailment relations among judgments that intertwine fact with plan” (Gibbard, 2003, pp. 57-58).

In sum: Gibbard proposes a mentalist alternative to truth-conditional analysis, where mental operations of ‘combining, rejecting, and generalizing’ on mental states are to take the place of the more standard logical notions of conjunction, negation, and quantification in semantic analysis. With a notion of ‘acceptance’ in hand that is neutral between normative and non-normative judgments, Gibbard’s norm-expressivism purports to be able to account for the logico-semantic features of ethical sentences that explain why, for instance “Stealing is wrong” is incompatible with “Stealing is not wrong”. What explains this incompatibility is supposedly the fact that the state of mind of accepting that stealing is wrong *disagrees with* the state of mind of accepting that stealing is not wrong.

This, at least, is the basic structure of Gibbard’s norm-expressivism. It contrasts with DS in that rather than appealing to deflated conceptions of truth and related notions to give a ‘fast-track’ solution to the Frege-Geach problem, it tackles the hard work of developing an expressivist compositional semantics head on. As Schroeder emphasizes in his (2008a) critique of the traditional expressivist’s semantic program, this is indeed hard work, because a satisfactory expressivist semantics not only needs to mirror the successes and progress of more conventional semantic programs; it must also meet a number of further constraints in virtue of the fact that it proposes a bifurcation between two fundamentally different kinds of semantic content – that which is representational, and that which is normative. Schroeder argues that Gibbard's and Blackburn's attempts to fill out this semantic program in detail leave much unanswered. To illustrate the serious challenge, let us consider *the negation problem*.

2.4.5 The Negation Problem (Unwin, 1999, 2001)

The negation problem, at its root, is a problem about the combination of mentalist semantics with non-cognitivist proposals about the mental states expressed by ethical claims. Traditional expressivism, we just saw, is saddled with the task of providing a comprehensive mentalist semantic program, as an alternative to truth-conditional approaches. One of the most basic constraints on an adequate semantic program is that it be able to account for *compositionality*, or how the meanings of complex linguistic constructions are built up in a systematic way out of the meanings of their parts. Part of providing such an explanation requires explaining how the meanings of atomic sentences combine with logical operators to contribute to the meanings of more complex sentences in which the atomic sentences figure. The simplest case will be to explain how negation operates on normative sentences. The negation problem is just the problem that leading versions of sophisticated expressivism – Blackburn's and Gibbard's – fail to adequately accomplish even this first step.

Let us first illustrate the problem with a simple emotivist view, according to which ethical claims express states of approval or disapproval. To illustrate this problem, it will be convenient to switch from analyzing atomic moral sentences (e.g. "Torture is wrong") to considering instead sentences that ascribe moral attitudes ("John thinks torture is wrong"). That is, it will help to ‘psychologically ascend’ (Yalcin, 2021) from moral judgments to ascriptions of moral judgments. (This is a typical expressivist move; the idea is to switch from talking about what the world would have to be like for a moral sentence to be true, to talk of what the mental state of accepting that sentence is like, where ‘acceptance’ does not always amount to having a representational belief. This is a useful move for the expressivist, who typically wants to avoid questions about what the world would have to be like for moral sentences to be true). So, consider:

(T) John thinks torture is wrong

This sentence describes John’s mental state, a state which our simple emotivist regards as one of disapproval of torture. So, the emotivist analysis of (T) is:

(T)* John disapproves of torture

Now, how should we analyze the following?

(N) John thinks torture is not wrong

What the expressivist would *like* to be able to say is that (N) describes John as being in a mental state that is inconsistent with the mental state (T) describes him as being in. But what state is that? Intuitively, it cannot be the state of disapproving of *not* torturing. But neither is it simply the state of not disapproving of torture. For (N) describes John as having a settled view on torture, having reached the conclusion that it is not wrong. This is different from John’s being completely *agnostic* as to whether torture is wrong or not, yet to read (N) as describing John as not disapproving of torture is consistent with his being agnostic. In sum, our options for analyzing (N) seem to be these:

(N)*^{1} John does not disapprove of torture

(N)*^{2} John disapproves of not torturing

And neither seems satisfactory as an analysis of (N).

One might hope that the problem just described only arises in the first place because of the simplicity of the emotivist view used as an example. If the project is to give a sophisticated semantic analysis in mental terms, it should perhaps come as no surprise that a simple emotivist view of the mental states expressed is not up to the task. Perhaps expressivists would do better to theorize with mental states that have more logical complexity than approvings and disapprovings, but which lack the representational features of belief as ordinarily understood. This is just the approach taken by sophisticated expressivism. On Blackburn's project, this is accomplished through a 'logic of attitudes'; in Gibbard's project, this is accomplished in terms of *planning* states and a normative variation on possible worlds-semantics. I continue to focus on Gibbard's project here.

Gibbard (2003) tackles the negation problem in the following way. Rather than construing the mental state of judging that torture is wrong as a kind of disapproval of torture, Gibbard explains the judgment that torture is wrong as a kind of planning state. In turn, Gibbard has a complex theory of planning states, a variation on possible worlds semantics, as described above. Roughly, the idea is that for John to judge torture wrong is for John to disagree with the set of 'hyperplanners' – fully decided planners with a contingency plan for every possible situation – who plan, for some situation, to torture. The general thought is that we can model the mental state expressed by an utterance in terms of the set of hyperplanners with whom one does not count as disagreeing just by being in that mental state. Now Gibbard's idea is that "to accept the negation of a plan is just to disagree with the plan" (2003, p. 74). This seems to get the right result: disagreeing with planning, for any situation, not to torture seems different from being agnostic about whether torturing is the thing to do, so we avoid the collapse into N*^{1}; and it is certainly not the same as disagreeing with planning *to* torture, avoiding the collapse into N*^{2}. In sum: by setting his analysis in terms of disagreement with planning states, Gibbard introduces more complexity than was found in the simple emotivist proposal, enabling his theory to model more complicated types of mental state.

While this feels like some progress, the expressivist strategy just described still has an important explanatory gap: Gibbard has helped himself to an unexplained notion of disagreement in mental state (Schroeder, 2008a, pp. 51-53). This is no oversight, but an essential aspect of Gibbard's strategy. One obvious way to explain mental state disagreement is in terms of inconsistent content; this is a clear way in which we can disagree in belief, for instance – if I believe something that is logically inconsistent with what you believe, we disagree. But Gibbard cannot appeal to this explanation because he needs to use disagreement in mental state to *explain* content, whereas the obvious explanation just considered *presupposes* content. So instead, Gibbard has to take disagreement in mental state as basic. As Gibbard explains (2003, p. 74),

> Proceeding this way might seem to be philosophical theft. The scheme amounts just to helping ourselves to the notion of disagreeing with a piece of content, be it a plan or a belief. A negation, we say, is what one accepts when one disagrees – and this explains negation. Now I wish, of course, that I could offer a deeper explanation of disagreement and negation. Expressivists like me, though, are not alone in such a plight. Orthodoxy starts with substantial, unexplained truth, eschewing any minimalist explanation of truth. I start with agreeing and disagreeing with pieces of content, some of which are plans. It's a thieving world, and I'm no worse than the others.

While it may be unsatisfying not to get an explanation of disagreement, this does not obviously sink the expressivist project. After all, everyone needs an account of disagreement, in the end. Unfortunately, the demand for an explanation of disagreement is particularly sharp for the expressivist.[26] As noted above, it is tempting to explain some instances of disagreement as arising from the fact that two individuals hold the same attitude (e.g. belief) towards

[26] The argument in this passage is loosely drawn from Schroeder (2008a, Chapter 3).

incompatible contents. As we've just seen, Gibbard seems to get this the wrong way around; he has to hold that it is the fact of disagreement which is basic, not the inconsistency in content. While everyone needs to account for disagreement, it is only the expressivist who is barred from the tempting explanation just considered.

Here's why this is a problem: In order to make the right predictions about when sentences are logically consistent or inconsistent with each other, Gibbard's theory must make the right predictions about when two mental states count as disagreeing with each other. But we just saw that mental state disagreement has to be taken as primitive in Gibbard's model. So, if Gibbard's model makes the right predictions about (in)consistency, it will be because the model makes the right stipulations about mental state agreement. But this does not seem to be a very promising theory, then, for it seems to stipulate what we wanted it to explain. As Schroeder puts it, regarding Gibbard's plan expressivism: "if we unpack it, what it is really saying is merely that 'murder is not wrong' must express a mental state that is inconsistent with all and only the hyperdecided mental states that 'murder is wrong' is not inconsistent with. And again, that looks more like a list of the criteria that we hope the attitude expressed by 'murder is not wrong' will satisfy, in lieu of a concrete story about which mental state this actually is, and why it turns out to be inconsistent with the right other mental state" (2008a, pp. 52-53). Adding complexity to the structure of the mental states expressed by utterances of ethical sentences does not help with the negation problem, then: for Gibbard still has to stipulate when these complex states count as disagreeing.

2.5 The Cognitivist Continuities

In this section, I discuss a variety of problems for traditional expressivism that center around the expressivist's non-cognitivist idea that ethical judgments are fundamentally different from ordinary descriptive beliefs. Briefly put, the main problem is that this non-cognitivist idea directly conflicts with the cognitivist appearances of the ethical domain, including that ethical claims appear to express moral beliefs – what I called *belief-expression* in Chapter 1. Additionally, ethical judgments appear subject to various *epistemic* assessments; we can describe people's ethical judgments as being justified to greater or lesser degrees, and we expect each other to be able to support the ethical assertions we make by providing reasons for thinking them true. This amounts to what I called *justification-aptness*. Moreover, it seems that, like beliefs about ordinary matters of fact, it is not 'up to us' whether or not a given ethical judgments is true. In this sense, ethics appears to be an *objective* matter. The challenge for traditional expressivism here is to account for each of these appearances while still maintaining a negative metaphysical thesis (anti-realism) and non-cognitivist thesis that ethical judgments do not represent some moral way the world is.

2.5.1 Belief-expression

Of the cognitivist continuities, belief-expression is the simplest for expressivists to account for. This is just because the 'data' supporting belief-expression as a feature of ethical thought and discourse does not strongly support conclusions that are clearly at odds with the expressivist view. In support of the belief-expression feature, in Chapter 1 I pointed out that it is quite natural to attribute moral beliefs to others; if Ahmad says "Torture is morally wrong", and appears sincere, we would ordinarily readily describe Ahmad as *believing* that torture is morally wrong. If there is any difficulty for the expressivist here, it is only because and to the extent that the ordinary folk notion of belief invoked is *representational*.

One reason to think it is representational is that the folk concept of belief figures into folk psychological explanations of action, according to which beliefs *about how the world is* combine with desires that the world be a certain way to generate action. (For instance, my act of putting the kettle on is explained by my desire to have tea, together with my belief that the best means for me to have tea involves putting on the kettle).

However, I don't think this provides very strong support for treating the folk concept of belief as representational. After all, *mathematical* beliefs can equally enter folk psychological explanations. (My belief that 5+5=10 seems to enter into the explanation of why I bought 2 bunches of 5 bananas, given that I know I need 10 bananas for a rather large loaf of banana bread.) But it is not obvious that mathematical beliefs represent features of the natural world. It seems to me, then, that the notion of belief required for folk-psychological explanations does not need to be representational.[27] The simplest option for expressivists to accommodate belief-expression is simply to concede that ethical claims do express beliefs, but then articulate a non-representational notion of belief consistent with the folk conception.[28]

2.5.2 Justification-Aptness

Similar considerations apply in accounting for the justification-aptness of ethical thought and discourse. We do ask for and sometimes provide reasons for the ethical claims we make. Simply admitting this would be a problem for expressivists if epistemic reasons to accept ethical claims had to ultimately refer to moral facts or properties. But in general it is not obvious that epistemic reasons must be representational. Again, consider mathematical beliefs: it is natural to

[27] At any rate, even if the folk concept of belief is essentially representational, another way for the expressivist to proceed is to concede that there is a descriptive, representational aspect to ethical judgments, but that what they describe is not moral reality, but just regular reality, thereby avoiding what expressivists regard as the 'spooky' metaphysical commitments of robust realism. So for instance, an expressivist might propose that Ahmad's utterance of "Torture is wrong" does express a belief – the belief that torture has some *naturalistic* property, such as the property of causing undesired pain – and *additionally* expresses something like disapproval of doing whatever has that property. This move is favored by *hybrid* expressivism, which I discuss in Chapter 3.

[28] Still, whatever account of belief is offered, it should turn out that belief states are something other or more than desire or desire-like states, if it is to be useful to provide folk-psychological explanations of action. I return to this point in Chapter 4, in support of my own hybrid view.

think that our mathematical beliefs are sometimes epistemically justified, but barring Platonism about mathematics, there don't seem to be any mathematical entities around for our epistemic capacities to track. So they must be justified in some other way – perhaps by coherence relations within a body of mathematical beliefs. When it comes specifically to moral belief, we must also consider that a leading approach to moral epistemology –reflective equilibrium[29] – is an essentially coherentist approach and as such does not require a realist moral metaphysics in order to explain moral knowledge.[30]

There is one aspect of the feature of justification-aptness that expressivists cannot handle so easily, however. Validly drawing an inference from premises one accepts to the conclusion they entail is one paradigmatic way to be justified in accepting that conclusion. A paradigmatic way to subject a set of judgments to epistemic assessment is by examining the logical relationships between those judgments for consistency. We subject ethical judgments to these sorts of epistemic assessment, and traditional expressivism owes an explanation of how we are able to do this. However, the Frege-Geach problem arises again here, under a new guise.

Earlier, I introduced the Frege-Geach problem in semantic terms, as a problem about explaining the linguistic significance of ethical sentences in non-assertoric contexts. There, the Frege-Geach problem was understood as a barrier to expressivists explaining the semantic continuities. However, some variations on the Frege-Geach problem I think are best understood as problems for the non-cognitivist commitment of traditional expressivism; these variations challenge expressivist attempts to explain the cognitive continuities. 'The' Frege-Geach problem

[29] The method of reflective equilibrium is developed in Rawls (1971).

[30] Thanks to Bill Lycan for pointing out that I need to consider non-tracking epistemologies as possible expressivist-friendly models for moral knowledge. See Lycan (1988) for defense of a coherentist epistemology.

is presented as a concern about how expressivists can account for the validity of moral arguments such as the following:

1. If lying is wrong, then getting your little brother to lie is wrong.
2. Lying is wrong.
3. Therefore, getting your little brother to lie is wrong.

This is an unproblematic instance of a *modus ponens* argument. It is clearly logically valid. Traditional expressivist theories should be able to account for its validity; this is one aspect of capturing the semantic continuities between ethical and non-ethical thought and discourse. The most immediate concern is that a standard way of understanding validity is in terms of truth-preservation, yet, at least according to early versions of expressivism, ethical claims are not truth-apt, and so cannot enter into relations of truth-preservation. But again, most contemporary expressivist admit that ethical claims are truth-apt in some sense, so this is not the most pressing version of the Frege-Geach problem.

Apart from accounting for logical validity, there are several other features of good moral arguments that expressivists must explain. For instance, as both Schroeder (2009) and Ridge (2006) note, good moral arguments have the following two features: they are *inconsistency-generating*, in that any possible reasoner who accepted the premises yet rejected the conclusion would be guaranteed to have logically inconsistent intellectual commitments. And good moral arguments are *inference-licensing*, in that it can at least sometimes be rational for reasoners to come to accept the conclusion of a good moral argument by first accepting the premises and reasoning to the conclusion from those premises.

A tempting approach would be to explain the inconsistency-generating and inference-licensing features is to appeal to the logical relations between the contents of the premises and conclusion; but this is not open to Gibbard, for the reasons discussed in treating the negation problem above. Instead, Gibbard explains these features in terms of *disagreement*. The idea is that to accept the premises while rejecting the conclusion of a valid argument is necessarily to have inconsistent attitudes, because the mental state one is in in virtue of accepting the premises disagrees with the mental state one is in in virtue of rejecting the conclusion. A similar analysis can be proposed for the inference-licensing feature of good moral arguments: coming to accept the premises of a valid argument makes accepting the conclusion the thing to do. However, as discussed in the previous section, it seems to be a problem that Gibbard takes disagreement in attitude as basic; we would like an explanation for *why* certain attitudes admit of disagreement. This remains a lacuna in the sophisticated expressivist program.

2.5.3 Objectivity

In motivating the expressivist project, we considered the advantages of expressivism over another superficially similar view: simple speaker subjectivism. Simple speaker subjectivism clearly denies that ethics is objective: what makes an ethical claim true, according to that theory, is the attitude of the person making it. One strong reason for rejecting simple speaker subjectivism is exactly that it fails to account for the apparent objectivity of ethics. If simple speaker subjectivism is true, then it would seem that ‘anything goes’ when it comes to ethics. But it is not the case that anything goes in ethics: some actions are just horribly morally wrong.

Is expressivism similar to subjectivism in respects that would prevent it from acknowledging the objectivity of ethics? We get an easy “yes” if we think that objectivity entails realism, and then note that neither subjectivism nor expressivism are realist views. But in Chapter 1, I cautioned against thinking of objectivity this way, preferring instead to characterize objectivity as roughly the idea that what is morally right or wrong is not determined by what any individual thinks is right or wrong. This clearly contradicts simple subjectivism, but it does not clearly conflict with expressivism, since expressivists deny that the truth-conditions of ethical claims are relativized to the attitudes of the claim-maker. One other obvious concern is that because (early) expressivism held that ethical claims were not truth-apt, it could hardly explain the existence of such a thing as objective ethical truth. But this is just an instance of the truth-aptness issue and seems to pose no additional worries when it comes to explaining objectivity. The Deflationary Strategy, for instance, provides one way around this concern.

Still, one might have a lingering concern that expressivism must somehow in the end collapse into subjectivism. One gets the sense that if moral judgments fall on the side of the ‘passions’, if they are ‘just feelings’, then morality could hardly be an objective affair. After all, what offends my moral sensibilities may not offend yours. Even if expressivism is not subject to the two simple concerns flagged above, we should expect the expressivist to provide us with a positive account of objectivity. I have taken care not to conflate objectivity with realism, but it has to be admitted that realists can account for objectivity fairly straightforwardly: claims in realist domains are ultimately made true (or false) by instantiations of mind-independent properties, not what any individual happens to think. Expressivism, qua antirealist position, seemingly cannot appeal to this straightforward explanation of the objectivity of ethics. What can the expressivist say?

There is an important general expressivist strategy for accounting for objectivity, along with other ‘realist-seeming’ features of ethical thought and discourse. This is the ‘quasi-realist’ strategy developed independently by Blackburn and Gibbard. Quasi-realism is a research program wherein the realist-seeming or realist surface appearances of some domain are explained without commitment to error theory/fictionalism, on the one hand, nor to realism, on the other. Quasi-realists, in short, aim to 'earn the right' to speak like realists, while in some crucial respect avoiding commitment to full-on realism. We are already familiar with the main ingredients in a quasi-realist analysis, appearing under the labels 'the Deflationary Strategy' (DS) and Sophisticated Expressivism (SE). Quasi-realism can be understood to combine these approaches; Blackburn, for instance, presents his quasi-realist strategy as integrating elements from what he calls 'fast-track' (in my terms, DS) and 'slow-track' (SE) routes to quasi-realism. As he puts it, "the fast track can benefit from some of the security of the slow, and the slow track can benefit from some of the shortcuts of the fast" (1993b[1988], p. 186). In what follows, we shall see how these can be applied in an expressivist-friendly account of objectivity.

Quasi-Realism, Objectivity, and Error

One feature of the truth-predicate often emphasized by deflationists is that it can be used as a tool for endorsing some bit of propositional content, especially in situations where one does not know or cannot state that content explicitly. For instance, asserting "It is true that grass is green" seems to be just another way of asserting "Grass is green". Similar considerations apply for the locutions "It is a fact that X", "It is the case that X"; these seem like just more tools with which to assert "X". This feature of the truth-predicate is entirely neutral with respect to subject matter. So long as a sentence meets the conditions for truth-aptness, it is apt for embedding in these various constructions (*viz*., "It is true that . . .", "It is the case that . . ."). Expressivists can leverage this point, together with the observation that ethical sentences are truth-apt, to 'earn the right' to speak of moral *truth* and moral *fact*. According to the expressivist, to say that it is *true* that (or is a *fact* that) torture is wrong just amounts to another way of saying (perhaps more emphatically) that torture is wrong. We can talk of moral truth, moral fact, moral reality, and so on, the expressivist can now say: but such talk is to be understood as just another tool for *making* moral claims. Quasi-realists capitalize on this strategy to make sense of realist-seeming talk in expressivist-friendly terms.

Can this deflationary strategy offer any purchase in accounting for objectivity? Not on its own, it seems; a similar strategy could be run for *taste* discourse, but this should not convince us that matters of taste are objective. What the deflationary strategy shows, if it succeeds, is that there are at least no *formal* barriers to capturing realist-sounding talk in an expressivist framework. The deflationary strategy enables expressivists to convert a demand to account for objective moral truth – which may seem a metaphysical demand that expressivists are in no position to meet – into a demand to account for the objective character of ethical *judgment*. That is, when we say that moral truth is objective, this is not to be understood as a claim about the moral features of the world, but as a claim about what it is to engage in ethical judgment.

As Gibbard understands it, to think that a norm is objective is to think that it would still be valid even if one had happened to reject it instead of accepting it (1990, p. 155). The idea, then, is that the objectivity of morality consists in the following feature of ethical judgment: That necessarily, when one makes an ethical judgment, one thereby accepts a system of norms as valid without qualification, such that they apply even to those who do not accept that system of norms (including oneself if one had happened not to accept that system). It is in this sense that expressivists can say that the demands of morality are not ‘up to us’: to make a moral judgment is, in part, to view oneself as subject to certain norms, and to think that one would still be subject to those norms even if one had thought one was not. (Put in terms of Gibbard’s later plan-expressivist account: to judge that one morally ought to phi is, among other things, to plan to phi even for the circumstance of being in one’s exact situation except that one does not judge that one morally ought to phi).

This seems to be a promising strategy for capturing objectivity. Here are three features recommending it. First: we can see that a parallel strategy cannot spuriously account for objectivity in taste discourse. Ethical judgments are, we might say, counterfactually insensitive, in that when one makes an ethical judgment one has to understand it to be valid for one even if one hadn't made that judgment. Taste judgments are counterfactually sensitive, as is illustrated by the strangeness of assertions like "Spinach is tasty. In fact, spinach would be tasty even if I had happened to strongly dislike it." Second: this strategy leaves room for the expressivist to distance herself from moral realism. What makes ethics objective, on this approach, is the nature of ethical judgment, not the metaphysics of moral properties.

The third recommending feature of Gibbard's approach to objectivity is that it seems also to capture an important conceptual connection between acknowledging a domain as objective and acknowledging the possibility of error in our judgments in that domain. This is an aspect of the sense that objective truths aren't 'up to us'. When a subject matter is objective, it is possible for one to get things *wrong* in that domain even if one thinks one has the correct view. Properly acknowledging the objectivity of a domain seems to rationally mandate also acknowledging the in-principle fallibility of judgments in that domain. Gibbard's proposal makes sense of this as follows. Suppose I judge torture to be morally wrong. On Gibbard's plan-expressivist account, I thereby plan, for any contingency, not to torture.[31] This includes planning, for the contingency of being in someone else's exact circumstances, not to torture; and it includes planning not to torture even for the counterfactual contingency of being in one's own exact actual circumstances except that one did not accept that torture is wrong. In the actual world, where I do judge torture to be morally wrong, I must regard my counterfactual torture-endorsing self to be in error,

[31] This is a rough gloss of how Gibbard's account would apply. Strictly speaking, my judgment that torture is wrong counts as disagreeing with the set of hyperdecided planners who plan, for at least some contingency, to torture.

because the norms against torture I acknowledge in the real world I must view as binding on my counterfactual self. When we say that someone is in error in their moral judgment, we are saying that they are in roughly the position of my counterfactual torture-endorsing self. We are saying that they are subject to a norm (belonging to the system of norms *we* accept), even though they do not think that they are.

2.5.4 The problem of fundamental moral error[32]

Although Gibbard's account does well in the three respects just discussed, there is an aspect of the objectivity of ethics that it seems to miss. As noted in the previous section, there is a conceptual connection between intellectual humility and objectivity; where a domain is objective, it is possible and sometimes appropriate to manifest a degree of humility about the correctness of one's judgments in that domain. Since ethics appears to be an objective matter, it should be possible and sometimes appropriate to manifest humility about the correctness of our ethical judgments. And intuitively, we do sometimes exhibit such humility. For instance, I might judge that it is wrong to allow children under the age of 5 to use smartphones, while acknowledging that I may be wrong about that. The challenge for expressivists is to explain how we can take up an attitude of humility regarding our own moral convictions.

In realist domains, the facts are what they are independently of our epistemic relation to those facts. Thus, realists about a domain can acknowledge that truths in that domain may outstrip our abilities to know them. This is a simple route to acknowledging the possibility of error in our own moral judgments. Traditional expressivism seeks to acknowledge the objectivity of ethics, including the possibility of exhibiting humility about one's moral convictions. But expressivism seemingly cannot appeal to the simple explanation of objectivity that is available to

[32] The ideas to follow in this section have benefitted from comments from Mike Ridge.

realists. There is a simple dilemma: if expressivists fully mimic the realist, *including* in acknowledging that the moral facts are what they are completely independently of our own attitudes and epistemic position with respect to them, the expressivist then confronts a sharpened version of the problem of creeping minimalism (Dreier, 2004) – roughly, the problem of then stating exactly how expressivism and realism are supposed to differ once expressivists 'earn the right' to talk like realists. But if expressivists do not mimic the realist with respect to objectivity, then the expressivist risks losing out on the attractive realist explanation of how we can acknowledge the possibility of first-personal moral error.

Beginning with the first horn, let us consider the problem of creeping minimalism: that insofar as the quasi-realist project succeeds in 'earning the right' to talk just like a moral realist would, it is no longer clear what separates quasi-realism from ordinary moral realism. Once the quasi-realist project gets started, any proposed point of contrast between quasi-realism and realism becomes a point that the quasi-realist is supposed to accommodate and 'earn the right to', thereby erasing the contrast that was proposed in the first place.[33] Using a combination of the deflationary and the sophisticated strategies discussed above, expressivists end up getting to speak of moral 'facts', 'properties', 'truths', 'beliefs', and so on. Quasi-realism can even go so far as to say that ethical judgments 'represent' moral 'reality'. One can be forgiven for wondering exactly where, then, the quasi-realist disagrees with the realist. Now, the lesson Dreier takes from this problem is that (2004, p. 42):

> Crucial to maintaining the distinction in meta-ethics, in the twenty-first century, between realism and irrealism is the possibility that concepts (and meanings) can differ in ways other than by their content. Or, if the difference between normative (or evaluative, or "planning") concepts and descriptive (naturalistic) ones can also be stated as a difference in content, then at least it must be a comprehensible, substantive question whether the difference in concept is *explained by* (or if you prefer *amounts to more than*) a difference

[33] I owe this way of posing the problem to conversation with Bill Lycan.

> in content, on the one hand, or rather it is explained by (amounts to) something else entirely, which in turn explains the difference in content.

As Blackburn puts it, “it is not what you finish by saying, but how you manage to say it that matters” (1993a[1998], p. 168). The problem of creeping minimalism, then, is not thought to be insurmountable; the trick is to provide an expressivist-friendly *explanation* for the realist-sounding theses that the quasi-realist accepts. But it seems to me that there is a barrier to expressivists accommodating the realist appearances with respect to objectivity. If expressivists do not just want to ‘talk like’ realists, but also want to appeal to moral reality to do real theoretical work in explaining how we could regard ourselves as possibly in fundamental error, the problem of creeping minimalism would return with a vengeance. For then expressivists would be making use of the realist’s very own metaphysical commitments in their explanation of objectivity. What started out as mimicry has become outright theft. There is a difference between explaining why one accepts something that sounds like realism without committing oneself to a realist metaphysics about some domain, and using the theses of realism to do real explanatory work.

Let us now consider how expressivists have attempted to respond to the second horn of the dilemma by providing a genuine alternative to realist explanations of fundamental moral error. I shall argue that the strongest expressivist responses are inadequate. I regard this as one of the most pressing problems for traditional expressivism, and it is one of my primary motivations for rejecting the traditional expressivist project. I devote the remainder of this chapter to explaining what the problem is, and why it is so difficult for the expressivist to address.

Expressivists have had something to say about acknowledging the possibility of error in first-personal cases. We saw above how Gibbard can account for this, in terms of planning for the situation of being in someone else’s exact circumstances, or one’s own circumstances except that one accepted something other than what one does in fact accept. But this does not go far enough. Planning what to do for the circumstance of being in one's own exact situation except that one accepts something other than what one in fact does accept, seems rather like planning what to do for the circumstance of being someone *else* remarkably like oneself except in certain commitments. It is not clear that we have really made progress in explaining how we can acknowledge *first-personal* moral error. Or at least, there remains something further to be explained. How is it possible, in the expressivist account, for S to think that her very own, actual, current moral judgment M might be mistaken?

I turn now to consider Blackburn's treatment of these issues, since Blackburn has had more to say on the topic, and explicitly engages in the task of accounting for humility in moral conviction. According to Blackburn, to regard oneself as possibly in moral error is to judge that changes in one's epistemic position that one would regard as improvements might lead one to abandon one's moral judgment (1998, p. 318). This seems to be the sort of position a quasi-realist has to take: for to admit the possibility of error even for a maximally subjectively well-justified moral belief would seem to require recognizing an independent realm of moral fact that the belief in question could fail to meet no matter its epistemic pedigree.[34]

Egan (2007) raises an important challenge to Blackburn's account of moral humility, known as the problem of fundamental moral error. In short, the challenge is that Blackburn's account cannot satisfactorily explain all that we want explained when it comes to moral error. While Blackburn's account can explain the acknowledgment of the possibility of third-personal

[34] In Blackburn's reply to Egan, he clarifies that he views moral truth along the lines of Wright's notion of 'superassertibility' – an anti-realist, epistemic notion of truth according to which something is true if the epistemic warrant for it survives all epistemic scrutiny and improvements in information (Blackburn, 2009, p. 206). This invites the question of whether expressivists might be better off to abandon the Deflationary Strategy altogether, and align themselves instead with *pluralism* about truth, which would enable them to acknowledge moral truths without the metaphysical commitments they are concerned to avoid. See Lynch (2013) for discussion.

moral errors, and while it can explain acknowledgment of the possibility of *some* first-personal errors, there is an important kind of error that it cannot handle: *fundamental* moral error.

Let us distinguish between ordinary error and fundamental error in general terms. An ordinary error, we can say, occurs when an individual S accepts that P, but it is incorrect to think that P. It could be that S has insufficient justification to accept P in an epistemically responsible manner; it may be that P is false; it may be that there is evidence for not-P that S should be aware of, etc. Generally, our ordinary errors can be corrected (at least in principle, anyway) if we can manage to get ourselves into a stronger epistemic position on the matter in question. If my friend tells me it is raining outside, this can put me in a good epistemic position to know that it is raining outside. But my friend can get things wrong, as can I, in which case I would be in error to believe what they say. Now, if I were to look out the window instead, it seems I would be in a much stronger epistemic position with respect to the weather (though my vision is not infallible either). But not all errors can be corrected just by getting oneself into a stronger epistemic position. The errors that cannot be so corrected are *fundamental*. If the radical skeptical hypothesis that I am currently a brain in a vat being stimulated to think I am currently writing my dissertation were true, I would be in fundamental error regarding most of my beliefs about the external world, because no matter what I might do to improve my epistemic position, I could never get myself in a position to know that I am a brain in a vat.

Intuitively, we seem as capable of fundamental error on *moral* matters as we are in any other realist domain. Thus, the capacity to recognize one’s views as subject to fundamental error seems to be one of the ways in which the ethical domain is epistemically continuous with other realist domains. For example: While I am very certain that it is morally wrong to cause undeserved suffering, it at least seems coherent for me to wonder whether I could be wrong about that. At the very least, given that the truth of moral nihilism is logically compatible with how things seem to me, it seems that for all I know I could be massively deceived in my moral judgments.[35] Slightly less fantastically, it is conceivable that sometimes it is not wrong to cause undeserved suffering; perhaps it is valuable to sometimes cause such suffering to oneself to strengthen one's character.[36] Yet even if one of these hypotheses were correct, I would still likely be just as confident in my belief that such suffering is wrong as I currently am. (As with the radical skeptical problem, it does not matter if these hypotheses seem implausible or far-fetched – they only need to be compatible with how things seem to me).

Egan explains the difference between ordinary and fundamental error in terms of what he calls 'stable' belief. A belief is stable just in case "no change that the believer would endorse as an improvement would lead them to abandon it" (Egan, 2007, p. 212). An ordinary error occurs when one holds a mistaken unstable belief. A fundamental error occurs when an individual holds a mistaken *stable* belief. When it comes to ordinary belief – such as my belief that human activity has contributed to climate change – I can regard myself as possibly mistaken, *even if* I am entirely justified in my belief by my own lights, and even if I would continue to be so justified through any changes in my epistemic position on the issue that I would regard as improvements. I can do everything in my power to arrive at what I think is the position best supported by the evidence, and yet still end up with a false belief, through no epistemic fault of my own.

The crux of the problem Egan (2007) articulates is that Blackburn's account can at best explain acknowledgement of *non*-fundamental first-person moral error. In Egan's terminology, a

[35] See Walter Sinnott-Armstrong (2006) for an argument that moral nihilism is logically compatible with how things appear to us and therefore cannot be ruled out. I respond to Sinnott-Armstrong's argument in Chapter 6.
[36] Ridge (2015, p. 19) considers some examples along these lines.

belief that would survive any improving changes in epistemic position (that is, improving by one's own lights) is a *stable* belief. Accounting for *fundamental* error amounts to accounting for the possibility of mistaken stable belief. But Blackburn's quasi-realism accounts for regarding oneself as possibly in moral error in terms of *unstable* belief – for to be in moral error just is for there to be some epistemic improvement that would lead one to abandon one's moral judgment. Expressivists, it seems, have to be 'unpardonably smug', as Blackburn puts it (1998, p. 318), when it comes to their most fundamental moral convictions, since they cannot account for first-person judgments of fundamental moral fallibility.

Blackburn's initial response to Egan's worry is to emphasize that 'improving changes' to one's epistemic position must be understood as changes that *one would regard* as improving, rather than as changes that are *objectively* improving, independently of one's recognizing them as such. With this clarification in hand, a *stable* moral belief is one that one would continue to hold through any changes to one's epistemic position that one would regard as an improvement. Blackburn's response to Egan's challenge then is to point out that one could still regard a stable belief as possibly mistaken, by considering the possibility that the changes that one would *regard* as improving might not *in fact* improve one's epistemic position. A moral belief that would not be revised through any changes that are *actually* improving must be true[37]; but we do not always know whether the changes we regard as improving our epistemic position in fact improve our epistemic position (Blackburn 2009, pp. 205-206).

Kohler (2015) argues that this response is inadequate, since at best it can account for how we might regard *others* as in fundamental moral error, but not ourselves. To judge of some other person S that her moral judgment M is possibly fundamentally mistaken is (i) to judge that there

[37] At least, on the anti-realist epistemic notion of truth Blackburn seems to accept for the moral domain (Blackburn, 2009, p. 206).

are no changes to her epistemic position that she would regard as improving which would lead her to reject M – i.e., S's belief is stable – and (ii) to judge that there might be epistemic changes that you yourself would regard as epistemic improvements which would warrant rejecting M. (This is just for you to think that there is some epistemic improvement S is not able to recognize that might warrant rejecting M). But this analysis cannot be applied to one's own case. To judge of yourself that your moral judgment M is possibly fundamentally mistaken is (i) to judge that there are no changes to your epistemic position that you would regard as improving and which would lead you to abandon M, and (ii) to judge that there might be some improving changes that would lead you to reject M. But in the first-person case, the 'improving changes' in (ii) advert to changes in your epistemic position that *you* would regard as improving. So, accepting both (i) and (ii) is contradictory when applied to one's own moral judgments. Thus the problem re-emerges.

Ridge (2015) is in agreement with Kohler's reinstatement of the problem for Blackburn's quasi-realism. Ridge, *contra* Blackburn, recognizes that the quasi-realist position cannot be defended simply by defusing various articulations of the fundamental error problem. Since quasi-realists need to 'earn the right' to talk like realists, they owe us a positive account of fundamental moral error. Ridge aims to provide just such an account. In the course of doing so, he adds several innovations to Blackburn's account deriving from Ridge's complex hybrid expressivist view (drawing from Ridge, 2006, 2007, 2014). In what follows, I shall simplify various aspects of Ridge's complex hybrid expressivist account, in order to focus on the essence of Ridge's solution to the problem of fundamental moral error. I consider Ridge's own hybrid expressivist account in more detail in Chapter 3.

Whereas traditional expressivism (as found in Blackburn) holds that normative judgments are non-cognitive, desire-like states, hybrid expressivism maintains that in addition, normative judgments are partly constituted by a representational belief, where, crucially, there is no particular belief that constitutes a given moral or epistemic judgment.[38] Different agents may make the same judgment and yet have corresponding representational beliefs with different contents. This is because the content of the belief that is partly constitutive of a token ethical judgment, in Ridge's view, is determined as a function from the agent's *normative perspective*: a stable intention *not* to endorse certain standards, and to only act/deliberate in ways permitted by standards not thereby ruled out (see Ridge 2014, Chapter 1; 2015, pp. 8-9).

The hybrid innovation introduced by Ridge does not seem to be able to offer any help concerning the problem of fundamental moral error. After all, a fundamental error in the representational belief component of moral judgment is not itself a *moral* error, but an ordinary cognitive one. Anything specific to fundamental *moral* error will have to concern the certainty of one's normative perspective itself. Insofar as the solution that Ridge offers provides no insight on how one could admit fallibility concerning one's normative perspective itself (rather than an associated non-moral belief), it is unclear how the hybrid expressivist view can offer any real progress on the problem of fundamental moral error. This, it seems to me, is a serious problem for Ridge's response to the problem of fundamental moral error.

The problem generalizes to apply to any expressivist attempts to account for fundamental error that follows a certain structure. Expressivists might hope to separate out an expressive from a descriptive aspect of moral judgment and rely upon the descriptive aspect to explain

[38] This qualification that no particular belief constitutes a normative judgment makes normative judgments 'massively multiply realizable'. This qualification is important for distinguishing the view from hybrid versions of *cognitivism*, which Ridge criticizes in other work (see Ridge, 2006; 2014, Chapter 3). See also discussion in Chapter 3.

fundamental error, even without ‘going hybrid’ like the views discussed in the next chapter. To illustrate: Expressivists like Gibbard have viewed it as an advantage of their theory that it nicely accounts for the supervenience of the moral on the non-moral; he argues that two acts can only differ in being ‘okay’ (permissible) to perform if they also differ in their prosaically factual properties (2003, p. 90). For instance, if I regard skiing as an okay activity to participate in, but Russian roulette as not an okay activity to participate in, I am committed to thinking there must be some factual property possessed by one of these activities but not the other that explains this difference in ‘okayness’ from my normative perspective. For an ideally hyperdecided planner (with a contingency plan for every possible situation), we can then identify a *property* of ‘okayness’: the prosaically factual property possessed by all and only those acts the planner regards as okay to do.[39] So for instance, suppose Jeremy endorses a simple utilitarian theory; Jeremy plans to act so as to maximize pleasure and minimize pain for the greatest number. We can then identify a property that all and only the actions Jeremy endorses have in common: that they maximize pleasure and minimize pain.

Given this take on supervenience, Gibbard can make room for acknowledging the possibility of a kind of fundamental error in moral judgment. Let’s say Jeremy judges that we ought to have a capital punishment program in the US because Jeremy believes it would deter future violent crime (this contributes to it promoting the best consequences, in Jeremy’s view). Jeremy could be in fundamental error: it could be *false* that capital punishment would best deter future violent crime, even if Jeremy would continue to think so through any changes he would

[39] As Gibbard explains: Hera, a hyperdecided planner, “accepts hyperplan *p*. She thus regards an act *a* as okay to do in a situation *s* if and only if her plan *p* permits *a* in *s*. But a plan can distinguish between situation only in terms of the prosaically factual properties of those acts. If two acts in two possible situations differ in no prosaically factual way, a plan can’t distinguish them, permitting one and ruling out the other” (2003, pp. 91-92). What is permitted to do according to *p*, then, supervenes on the factual properties of situations.

regard as epistemically improving. (Assume that Jeremy's access to information is limited, such that, for all Jeremy can know, capital punishment best deters crime). My point is just that this is a kind of fundamental error, but it is fundamental error in prosaically factual belief about the efficacy of capital punishment as a deterrent, when what we wanted was an account of fundamental error specifically in *moral* commitment – e.g. an error about whether the fact that capital punishment would deter crime counts in its favor morally speaking.

Below, I sketch a second problem for Ridge's response, but I place less argumentative weight on this secondary problem, for two reasons, one substantive, one organizational. The substantive point is that my argument below admittedly rests on a highly controversial account of certainty. In my opinion, that account is correct. But I do not wish to hang the core argument against traditional expressivism on it. The organizational point is that I discuss this account of certainty at much greater length in Chapter 6, in the course of defending the possibility of moral knowledge. That will be the appropriate occasion to defend my views on certainty. Here, my aim is to indicate the implications of these accounts of certainty for Ridge's proposed solution to the problem of fundamental moral error.

Given his hybrid expressivist account, Ridge takes it that he must reconstruct the problem of fundamental moral error in terms of the normative perspective/representational belief pair that make up ethical judgments. As Ridge construes it (but see misgivings above), the challenge for the expressivist is to explain how a subject S could at once make a judgment of first-person fallibility about her moral belief M, while remaining certain that her judgment that M is stable (in Egan's sense).[40] That is, the puzzle is one of explaining how I can at once be certain that my

[40] One might wonder why Ridge construes the problem in terms of an agent's *certainty* that her judgment is stable, instead of the *stability* of the judgment. It seems that it is not sufficient to generate the problem for her belief to just *be* stable, whether or not she realizes it. For if one were unaware, or unsure of, the stability of one's belief, there would be no internal tension between holding that belief and wondering whether one might be mistaken in holding

moral judgment that M will survive changes to my epistemic position I would regard as improvements, while also recognizing the possibility that M is mistaken. There are thus two components to Ridge's solution; an account of acknowledgments of fallibility in ethical judgment, and an account of what is involved in being certain that a given belief one has is stable.

My certainty that my belief that M is stable is explained as follows: according to Ridge, being certain that P should be understood as assigning a much higher level of credence to P than to not-P. That is, certainty should *not* be construed as assigning a credence value of 1 (as one might have expected). Ridge's justification for this qualification is that, first, given the prevalence of radical skeptical hypotheses, not just in philosophy but in popular culture (such as in the film *The Matrix*), probably most people do not assign a credence of 1 to those judgments they consider certain. And second, as a matter of fact, many ordinary judgments about what is certain tend not to be as certain as the most secure pieces of knowledge, such as that 2+2=4, and so it does better justice to our ordinary use of 'certain' to construe it as assigning something less than a credence of 1 (Ridge, 2015, p. 13). With this understanding of certainty, it turns out that being certain that a belief of mine is stable does not amount to my thinking that there is *no* possible evidence that might make it rationally permitted for me to reject that belief. Instead, it is to think that it is very *unlikely* that I will come across evidence that would make it rationally permitted for me to abandon that belief.

Now we turn to consider the other component of Ridge's solution; the account of first-person fallibility. Since a judgment of first-person fallibility is *itself* a normative judgment, according to Ridge, it is also composed of a normative perspective/representational belief pair.

it. This emphasizes that the problem of fundamental moral error is essentially about a certain *reflective* stance we can take towards our moral judgment.

So, my judging that I might be wrong on some moral matter M involves: a representational belief that some epistemic standard not ruled out by my normative perspective would permit *not* assigning a much higher credence to M than to not-M, if I assign them credences at all (2015, p. 14). That is, the fallibility judgment ranges over the epistemic standards not ruled out by my normative perspective, and it is a judgment that on at least one of those standards, it would be epistemically permissible not to hold M as much more likely true than not-M.

We are now in a position to see the structure of Ridge's solution. Recall that there are two parts – (i) the certainty that one's moral judgment M is stable, and (ii) the judgment of first-person fallibility concerning M. The main point is that one's *certainty* that one's moral judgment M is stable is explained in *counterfactual* terms – what I *would* be rationally permitted to believe if I had different evidence that I would regard as improving on my actual epistemic situation. Ridge accepts a possible-worlds semantics for such counterfactuals, so that to assess the judgment of certainty we only assess possible worlds nearby to the actual world, and disregard distant possibilities (Ridge, 2015, p. 16). By contrast, the descriptive belief involved in the fallibility judgment ranges over *all* epistemic standards not ruled out by one's normative perspective. Thus, the fallibility judgment reaches out further into modal space than the certainty judgment. This gap in modal space between the two judgments is supposedly what makes recognizing the possibility of first-person fundamental moral error possible. So long as there is some very unlikely circumstance in which I would have evidence that would make it rationally permissible for me to reject M, I can regard my belief that M as possibly in fundamental error. Since the possibility that I would come to have evidence speaking against M is unlikely, I can recognize that possibility while also being certain of the stability of my belief that M, since certainty judgments ignore such unlikely possibilities.

Ridge's solution to the problem of fundamental moral error does have intuitive plausibility. The main idea of this solution is that to be certain of something is compatible with recognizing the possibility that one might be mistaken, since the possibilities for error are more wide-ranging than the grounds for certainty. This seems like a promising approach not just to accounting for fundamental moral error, but for developing an account of intellectual humility more generally. However, I think that Ridge's solution is nevertheless based on an assumption about the nature of certainty that can and should be challenged.

In particular, I think we should question Ridge's construal of certainty in terms of having a *much higher* degree of credence in some proposition than in its negation. Though Ridge's construal enjoys some plausibility, I think the alternative conception of certainty Ridge rejects – as having a credence of 1 towards some proposition – is the correct one, at least in the context of thinking about fundamental error. This alternative understanding of certainty appears similar to what Ridge labels *hyper-stability* (2015, p. 17). A belief is hyper-stable if it would survive any changes that the agent would regard as epistemic improvements *at the time of the change*. So understood, if a belief is hyper-stable, it is impossible for one to ever regard it as rational to abandon the belief.[41] Now, it does seem there are ordinary cases where people rationally change their minds on matters about which they would have correctly described themselves as certain in the past, so hyper-stability seems too strong. Perhaps a better model for certainty, then, would be something like 'super-stability', where a belief is super-stable just in case it would survive any changes that the agent would regard as epistemic improvements given the agent's *actual current* epistemic position, not relative to her epistemic position after the change.

[41] Here is why: At any point of change in epistemic position, no change would be regarded as improving with respect to a hyper-stable belief, by definition. So it cannot be rational to lower one's confidence in a hyper-stable belief in response to a change in epistemic position.

Ridge understandably cautions against understanding certainty in terms of hyper-stability (and presumably super-stability as well) noting, for instance, that hyper-stability appears to be an epistemic *vice,* and at any rate not a necessary commitment of expressivism. One of the main reasons for preferring the ‘much-higher-credence’ account over a super-stability account is that certain radical skeptical hypotheses seem to show that it would not be rational to form super-stable beliefs. All that is needed to show the supposed irrationality of S’s super-stable belief that P is to construct a skeptical hypothesis H incompatible with P, yet which S cannot rule out as a possibility. S should then be able to infer, from the recognition that she cannot rule H out, and her knowledge that H is incompatible with P, that she cannot be sure that P is true – which is incompatible with her having the super-stable belief that P. By contrast, it is consistent with S’s assigning P a much higher degree of credence than not-P that there is some skeptical hypothesis H incompatible with P that S cannot rule out – S presumably would also assign P a much higher credence than H.

In Chapter 6, I shall argue that having super-stable beliefs is in fact *not* an epistemic vice, *not* irrational, does *not* lead to skeptical problems, and moreover, is a necessary condition for rational evaluation to occur at all. Given that most of us have some super-stable moral beliefs, and are not thereby irrational, the problem of fundamental moral error remains, for there is then no ‘modal gap’ between the judgment of fallibility and the judgment of certainty crucial to Ridge’s proposed solution. But the problem takes on a new significance, in that it is no longer specific to the moral domain, or to quasi-realism. It becomes a general problem of reconciling our most fundamental, certain commitments with a recognition of the possibility of error. My suggestions for resolving this general problem, too, will have to wait for Chapter 6.

2.6 Conclusion

Let us take stock. Traditional expressivism, a view championed by Blackburn and Gibbard, is comprised of four logically distinct theses: the positive expressivist thesis, the constitutivist thesis (motivational internalism), the negative ontological thesis, and the anti-representationalist non-cognitivist thesis. The main insight of this view I wish to preserve is the explanation it provides of the apparent close connection between ethical claims and motivation to act. In future chapters, this connection will be examined in more detail; as we shall see, standard understandings of the motivational internalist idea cannot stand as they are, but require important qualification. Still, what the expressivist is right about, in my view, is that the connection between ethical claims and motivation is best explained in terms of the expressive character of ethical claims.

Traditional expressivism, we have seen, embeds the positive expressivist thesis just mentioned within an irrealist metaethical theory that takes on substantive commitments in philosophy of language, philosophy of mind, and epistemology as they apply to the ethical domain. It holds that ethical judgments are motivationally-charged non-cognitive states of a certain sort, that ethical sentences express these states, and that this is what the meaning of basic ethical sentences consist in. It holds that ethical judgments do not represent some moral way that the world might be, and that ethical claims do not describe the world as being such a way.

The central challenge for traditional expressivism I have considered in this chapter is that of adequately accounting for the continuities that appear to exist between the ethical domain and other domains. These included the semantic continuities: truth-aptness and embeddability, and the cognitive continuities: belief-expression, justification-aptness, and objectivity. As we have seen, in order to succeed in this project of accommodating the appearances of ethical thought and discourse, expressivists have endorsed controversial mentalist semantic theories along with controversial deflationary accounts of truth, belief, fact, and so on. Finally, even granting that the expressivist mentalist semantics and minimalist quasi-realist program succeed as intended, we saw that there still remains an explanatory challenge for traditional expressivism: that of accounting for the possibility of regarding oneself as subject to fundamental moral error. I argued that extant attempts to address this problem do not succeed. However, one of my objections to Ridge' s response to the problem of fundamental moral error turns on a highly controversial view of the nature of certainty. I defend this view in Chapter 6, where I argue for the possibility of rationally grounded moral knowledge.

Since traditional expressivism is comprised of four logically independent theses, one might think that the expressivist explanation of the connection to motivation could be preserved while rejecting the theses responsible for the challenges to traditional expressivism. None of the problems discussed in this chapter arise for cognitivist realist theories of ethical thought and discourse. So, one might think, combining the positive expressivist thesis, and possibly the constitutivist thesis (motivational internalism), with an otherwise cognitivist realist theory, would yield an account that preserves the expressivist insight while avoiding all of expressivism's problems. Such a theory, it seems, would have the best of both worlds as between expressivism and cognitivism. This is the general strategy taken by recent work in hybrid metaethics. The next chapter is devoted to examining hybrid theories.

Chapter 3

Hybrid Metaethical Theories

3.1 Introduction

In the previous chapter, I reviewed some of the most pressing challenges for traditional expressivism, including the Frege-Geach problem. I argued that traditional expressivism takes on some difficult explanatory challenges. First, traditional expressivism must provide a fully worked out mentalist semantics to supplant more conventional truth-conditional and possible-worlds semantics, where the expressivist's semantics must also include a bifurcation in its analysis of ethical language and ordinary descriptive language. And second, I have argued that traditional expressivism has not yet adequately accounted for all the cognitive continuities we observe between ethics and other domains, including the possibility of fundamental error. However, I noted that it may be possible to retain what I see as the primary attraction of traditional expressivism – the explanation it provides of the connection between ethical claims and motivation – while rejecting the expressivist commitments that lead to the explanatory challenges just mentioned. In this chapter, I examine some of the extant attempts in the recent literature to do just this.

Hybrid metaethical theories hope to retain an expressivist-style explanation of the distinctive features of the ethical domain, while avoiding the problems for pure expressivism by incorporating a cognitivist element in their analysis of ethical thought and discourse. The purpose of this chapter is to review the most prominent hybrid theories and assess their plausibility. I shall first, in section 3.2, summarize various theories that broadly aim to accommodate intuitions driving both pure cognitivism and pure expressivism. In section 3.3, I consider Schroeder's influential 'Big Hypothesis' argument, intended to set a challenge for hybrid theories (2009, 2014). I point out that Schroeder's way of setting up the issues neglects a crucial distinction (to be discussed fully in Chapter 4), and as a result obscures the possibility of certain promising avenues for hybrid theories. Schroeder's way of setting up the dialectic is nevertheless understandable, since the crucial assumptions he makes are fairly widespread in the literature itself. I conclude by briefly proposing my own taxonomy for distinguishing among hybrid views, one that makes space for the possibility of the view to be proposed in the next Chapter.

3.2 Taxonomy of views

Hybrid metaethical theories are generally understood to be theories holding either (i) that ethical *claims* express both a representational belief and a motivationally-charged non-cognitive attitude, or (ii) that ethical *judgments* are constituted by both representational beliefs and motivationally-charged non-cognitive attitudes, or both. Additionally, hybrid theories are often distinguished according to whether they give priority in analysis to either the cognitive element or the motivational element of ethical discourse or thought; views of the former type are versions of 'ecumenical cognitivism', while views of the latter type are versions of 'ecumenical expressivism' (terminology due to Ridge, 2014). And finally, hybrid views can be further distinguished according to the psychological or linguistic mechanisms by which they propose that the 'secondary' aspect of ethical claims or judgments are generated. I shall return, at the end of this chapter, to provide a more precise set of taxonomical questions after summarizing some central hybrid theories.

3.2.1 Barker (2000)

Stephen Barker's "Is Value-Content a Component of Conventional Implicature" is generally recognized as an early hybrid view.[1] Barker's primary motivation is to reject an argument due to Jackson and Pettit (1998) that expressivism collapses into subjectivism. It is worth summarizing Barker's construal of the argument from Jackson and Pettit, at least because it represents a further criticism of pure expressivism, as discussed in Chapter 2. Additionally, Barker's response to the argument, which informs much of the ensuing literature in hybrid theories, already contains some of the crucial assumptions to be challenged by the neo-expressivist framework. I summarize Barker's representation of Jackson and Pettit's argument as follows (Barker 2000, pp. 268-269):

1. Expressivism claims that the 'value content' of an uttered evaluative sentence 'V' resides in a speaker U's conveying that she is in mental state ψ.
2. The conveyance in (1) is associated with clear-cut linguistic rules. (Expressivist premise)
3. From 2: Competent use of 'V' involves U's capacity to recognize that she is in ψ. So, U must believe that she is in ψ in uttering 'V' sincerely.
4. If linguistic rules associate U's believing P with U's uttering 'V', P is a component of 'V's truth-conditions.
5. From 3 and 4, U's belief that she is in ψ is part of the truth-conditions of 'V'.
6. But 5 is just a statement of subjectivism.

Before proceeding with Barker's response, let us flag several features of this argument that will prove significant later (setting aside initial worries about the argument itself). It is important to raise these assumptions here, as they are also present in many of the other hybrid theories to be examined shortly.

[1] But several authors identify hybrid-like elements in earlier views that are often classified under the heading of 'non-cognitivism' – including in Stevenson and Hare (see Chrisman, 2013; Eriksson, 2009 for discussion).

Barker begins by noting that premise (4), as a general linguistic claim, is false. In particular, the phenomenon of conventional implicature (as discussed in Grice, 1989) represents a straightforward counterexample, since conventional implicature is (i) an aspect of conventional meaning that (ii) conveys propositional contents, yet (iii) does not contribute to the truth-conditions of the utterance carrying the implicature. As a standard example of conventional implicature, consider ‘but’: “LeBron is big but fast” has the same truth-condition as “LeBron is big and fast” – yet the sentences differ in meaning, since the first, but not the second, indicates that the speaker believes there is a *contrast* between someone’s being big and their being fast (compare to discussion in 2.3 of implicature-based accounts of the expression relation).[2]

Now, it may be admitted that premise (4) does not hold, as a general thesis. Still, for the particular argument provided by Jackson and Pettit to fail, it should also be shown that premise (4) does not hold even when restricted to ethical sentences. This is exactly what Barker intends to show, by arguing that ‘value content’ is a component of the conventional implicature of ethical terms. The resulting view, which Barker labels ‘the implicature theory’ (IT), is as follows (2000, p. 271):

> **Implicature Theory (IT)**: If U asserts an evaluative sentence such as ‘T is good’, then U denotes a property F by ‘good’ and:
>
> (i). U expresses-as-explicature the content that T is F. (This is the truth-conditional content of the sentence ‘T is good’).
>
> (ii). U expresses-as-implicature the content that U is committed to approval of F-things.
>
> (iii). U conveys that she believes the contents in (i) and (ii).
>
> (iv). U conveys that she approves of T.

[2] It should be noted that the existence of conventional implicature itself is not beyond doubt (see Bach, 1999), so the reliance on conventional implicature may be a dialectical weak point in Barker’s theory.

Barker takes it that, in the first instance, it is *speakers* that denote properties by their use of words. This is broadly in keeping with Grice's notion of speaker-meaning, and may ultimately figure in a metasemantic explanation of how speaker-meaning 'ossifies' over repeated use to establish linguistic conventions fixing the conventional meanings of linguistic items (Grice, 1989). However, we should be careful not to assume this Gricean metasemantic picture from the outset, and so we would do well to more neutrally hold that if anything has the 'job' of denoting properties, it is predicates and concepts, not speakers, regardless of how they come to have this job. (Similar comments apply for the idea that it is the *speaker*, U, who expresses the content that T is F, which is the truth-condition for the sentence she utters, on Barker's account). (More on this in Section 3.3)

In categorizing IT, it is important to note that the view is a version of *ecumenical cognitivism*, in that the 'primary' truth-conditional content of normative utterances is *descriptive*; an assertion of 'T is good' serves to describe T as being F. Additionally, since the view maintains that *which* property evaluative predicates pick out (as their 'explicated' content) is a matter of the *speaker's* moral view, it seems to provide a contextualist semantics for evaluative predicates, and so in this way is similar to speaker subjectivism (though Barker denies that the view is subjectivist, given its other features).[3] I shall not hang my criticisms of Barker's view on this point, but it is worth taking seriously the possibility that standard problems for subjectivism might also be problems for IT (unless features (ii)-(iv) of IT can somehow be employed to stave off such worries, as Barker thinks they can).

[3] Finlay (2005) instead labels it an 'indexical' theory, since it holds that 'good' is indexical in a manner similar to other indexical terms, such as "I"; Barker takes 'good' to have a *character*, in Kaplan's sense (1989). In this way, Barker's view is similar to Dreier's speaker-relativism (1990).

Returning to exegesis: Barker holds (i) and (ii) to be aspects of the *locutionary* content of an assertion of ‘T is good’; (i) concerns the assertion’s truth-conditions, while (ii) is an aspect of its meaning that is not truth-conditional. By contrast, (iii) and (iv) are intended as aspects of the *illocutionary act* performed. For this reason, Barker predicts that (i) and (ii) will contribute to the meaning (truth-conditional and otherwise) of ‘T is good’ in *embedded* contexts (such as in conditionals), while (iii) and (iv) will not.[4] As aspects of illocutionary force, (iii) and (iv) are absent in ‘force-stripping’ contexts such as embedding in conditionals.

Some significant features of IT advertised by Barker include: (a) how it handles embedding contexts and (b) how it handles normative disagreement.

Regarding (a): Barker points out that in general, implicatures ‘project through’ certain embedding contexts, such as logically complex constructions like conditionals. As a result, a commitment of IT is that an assertion of a complex sentence containing evaluative terms, such as “If T is good, then R is good” carries the implicature that the speaker is committed to approval of F-things. This is surprising, especially since a main point of the Frege-Geach idea is that embedding evaluative sentences in logically complex constructions is supposed to *remove* the expression of non-cognitive attitudes. It seems that the only way Barker’s proposal here could be plausible is if the attitude that is supposed to be expressed by utterances of moral sentences is a quite *general* attitude towards things insofar as they have some feature, rather than an attitude directed at particular acts, agents, or policies. So, even in the assertion of a simple moral sentence such as “Stealing is wrong”, the attitude that is implicated is not simply disapproval of

[4] While conventional implicature does not contribute to truth-conditions in unembedded contexts, Barker holds implicature content does contribute to truth-conditions in embedded contexts. So, it is not the case that “John believes that LeBron is big but fast” is true if and only if “John believes that LeBron is big and fast”, presumably because the first sentence is only true if John thinks there is a contrast between being big and being fast while the second sentence might be true even if John does not think there is such a contrast.

stealing, but disapproval of *F-things*, or perhaps *things insofar as they are F*. The attitude then gets to be directed at stealing in virtue of the fact that in asserting “Stealing is wrong”, a speaker also conveys that stealing is F.

A further controversial commitment of this approach, that we also see in some other hybrid theories, is that it appears to require a non-standard account of validity. Barker notes that implicated content can make a difference to validity of arguments, and that therefore validity cannot be understood purely in terms of truth-preservation, which ignores implicated content.[5] A consequence of this – which Barker endorses – is that one might accept the premises in a valid moral argument *as true*, yet rationally *reject* the conclusion, because one does not also have the relevant non-cognitive attitude that would be implicated by acceptance of the premises or conclusion.

However, this move, when understood as a proposal about how validity should be defined, carries significant and controversial commitments not fully appreciated by Barker. For instance, in defining validity in terms of *correctness*- rather than truth- preservation, arguments that would intuitively count as valid seemingly turn out invalid. Consider, for instance:

1. If LeBron is big but fast, then so is Michael.
2. LeBron is big.
3. LeBron is fast.
4. So LeBron is big and fast.
5. Therefore, Michael is big but fast.

[5] Barker’s example: “(i) If the hostages are released, *even* John will be given amnesty. (ii) If *even* John is given amnesty, there will be a peace deal. So, (iii) if the hostages are released, then there will be a peace deal. [This argument] is most plausibly seen as a straightforward case of transitivity. The implicature content of ‘even’ is essential to the argument” (p. 273).

Intuitively, this seems logically valid. If it is somewhat strange, this at least does not appear to be due to any *logical* fault with the argument. Yet according to Barker's 'correctness-validity', it is not a valid argument, because premise 4 lacks the implicature content in the antecedent of premise 1, blocking the modus ponens inference from 1 and 4 to 5. Concerns like this should make us wary of hybrid views that require developing non-standard accounts of validity (such as Ridge's ecumenical expressivism, to be discussed below).

Setting aside this concern about validity and embedding, IT does still retain a more conventional approach to addressing the Frege-Geach problem. Even though the implicated value-content is a general sort of attitude towards things insofar as they have a certain property, IT also maintains that assertions of simple ethical sentences convey an attitude directed at a particular act, person or policy, as per feature (iv) of the view. This feature is supposed to be an aspect of the illocutionary act of moral assertion, and so this *is* cancelled in 'force-stripping' contexts, as any adequate response to the Frege-Geach problem should predict. Thus, an assertion of "If T is good, then R is good" carries the implicature that the speaker is committed to approval of F-things, but it does *not* convey any *particular* attitude towards T or towards R. This reveals an important choice point for hybrid theories: do they hold that the relevant non-cognitive attitude is 1) directed at a general quality that acts, persons, or policies might have or lack, 2) directed at particular acts, persons, or policies but only insofar as they have some quality, or 3) directed simply at some act, person, or policy, without reference to their qualities. Barker appears to hold that simple moral assertions exhibit both (1) and (2), but that even utterances of logically complex sentences with moral content exhibit (3).

Regarding (b): A crucial insight of Stevenson's emotivism (1937) is that an interesting feature of *moral* disagreements is that they are not (or at least not *only*) disagreements in (representational) belief, but disagreements in *attitude*. Barker's IT aims to capture the intricacies of moral disagreement by accounting for both disagreement in attitude and factual disagreements. Consider the following case from Barker (2000, p. 277): Imagine that a racist, Norm, gestures at SS officer Schmidt in a film, and makes the following claim:

Norm: "Schmidt was good."

We might respond to this in two ways:

Us: "What Norm said was true," or "What Norm said was false".

Now, Barker views the descriptive content of normative claims as determined by a function from the speaker's attitude to some property F (2000, p. 277). So, if we focus on the *attitude* implicated in Norm's utterance – approval of racism – the descriptive content of Norm's utterance would be that Schmidt was racist, and we should judge Norm's utterance as true. Nevertheless, Barker hastens to add, there would be reason not to *voice* the judgment that what Norm said is true, since doing so would generate the implicature that we ourselves *share* Norm's attitude. If instead we focus on what Barker takes to be a further presupposition of Norm's utterance – namely, that the audience or interpreter *shares* Norm's attitude – we should be inclined to judge that what Norm said was *false*. Thus, Barker purports to capture both the intuition that we would regard Norm's utterance as false, as well as the (subjectivist) thought that Norm has uttered something *true*, but that we would not accept ourselves.

However, I think we should question whether this latter 'intuition' is really a surface feature of moral disagreement. Even when it is clarified that in this context, by 'good', Norm means 'racist', *we* should resist characterizing what Norm said as true, insisting that it would still be linguistically (and morally!) inappropriate to describe someone as morally good in virtue of being racist.[6] This should suggest that, *contra* the subjectivist idea, whatever descriptive content ethical terms carry does not vary from speaker to speaker, or at the very least that there are restrictions on the legitimate variations.

In sum, Barker's IT introduces the innovation of employing conventional implicature to explain how ethical claims might express motivational states in addition to beliefs. However, even apart from the problematic features of this conventional implicature idea (to be explored below in discussing Copp's proposal), IT has a number of other controversial commitments that should give us pause. First, the view risks shouldering the worries of traditional subjectivism in holding that the descriptive content of ethical claims are determined as a function from the speaker's attitudes. Second, the view makes some questionable predictions about what sorts of claims express non-cognitive attitudes; in particular, we might question the idea that utterances of *conditional* sentences containing ethical terms carry the implicature that the speaker has a certain non-cognitive attitude. Third, the view makes implausible predictions about moral disagreement, although it might be argued that it does better than simple subjectivism. And fourth, the view requires a revision to standard conceptions of validity that risks ruling as invalid arguments are clearly valid.

3.2.2 Copp (2000, 2009, 2014)

In several articles, David Copp has proposed a hybrid view he labels 'realist-expressivism'. This view is like Barker's in that it is a version of ecumenical cognitivism that makes central use of the notion of conventional implicature. However, Copp's view does not take on Barker's pseudo-subjectivism concerning the *descriptive* content of ethical claims (but as

[6] Thus much, of course, quasi-realist expressivists will be happy to agree with, as they would maintain that the question of whether 'morally good' could denote the property of being racist is itself a substantive, first-order normative question.

we shall see, Copp's view may still best be thought of as a sort of *neo*-subjectivism, for other reasons). Instead, Copp intends his view to be objective and realist, in that it takes the 'primary semantic function' of moral predicates to be that of picking out moral properties that have the same metaphysical status as non-moral properties. The descriptive, truth-conditional meaning of an ethical sentence of the form "T is good" can thus be schematized as "T is F", where this content is constant across speakers, so that in making ethical claims, my utterances of 'good' refer to the same property as your utterances (even if we disagree in what we think that property might be).[7] Further, Copp sees ethical claims as like standard assertions in that when they are sincere, they express representational beliefs with the same truth-conditions as the sentence asserted.

Thus far, then, Copp's position is conventionally morally realist and cognitivist. Copp's central addition – what renders his view realist-*expressivism* – is that he also takes ethical assertions to standardly express a 'conative-motivational' state of the speaker. The motivation for adding this feature is to provide an explanation of the *practicality* of ethical discourse. Dialectically, the success of this proposal bolsters the case for realism, at least insofar as the apparent close connection between making ethical claims and being in an appropriate motivational state are sometimes taken to support pure expressivism as opposed to moral realism. Realist-expressivism disarms the internalist argument for expressivism by showing how a cognitivist theory can also give an expressivist-style explanation of moral motivation. Despite all of this, Copp officially endorses motivational *externalism,* since he holds that one can make a genuine and correct moral judgment while lacking all motivation to act in accordance with that

[7] Although I characterize Copp's view as *objectivist* realism, Copp (1995) has developed a complex proposal about moral properties, the 'society-centered theory', which seems to allow for some variation in moral properties across cultures. Thus, Copp's view might be better thought of as a tempered version of cultural relativism. Still, the overall proposal is at least more objectivist than Barker's.

judgment (later, we shall question this way of distinguishing between externalism and internalism). Still, Copp maintains that to make an ethical assertion expressing a moral judgment that is not accompanied by the relevant motivational state would be *linguistically* inappropriate, even if not irrational. Copp labels the resulting position *discourse internalism*, because it posits the following necessary connection between ethical *discourse* and motivation:

> *Discourse internalism*: necessarily, if one makes a fully linguistically appropriate ethical claim, one must be in a certain motivational state.

This idea explains, for instance, why we would judge it strange if someone were to claim that it is good to donate to charity, yet immediately goes on to add that they are not at all motivated to do so. Thus, Copp hopes to explain the sorts of cases that have been taken to provide support for motivational internalism, in a way compatible with motivational externalism. In sum, Copp's view is an objectivist version of ecumenical cognitivism that advances a hybrid view of ethical *discourse*, but not of ethical *thought*.

What is most relevant for our purposes is to examine Copp's proposal about the linguistic mechanism linking ethical claims to motivation. Copp, like Barker, proposes that this connection is forged in the conventional linguistic meanings of ethical terms, but does not contribute to the truth-conditions of (unembedded) sentences in which those terms appear. So, Copp's proposal is also that ethical terms carry, as a *conventional implicature*, the content that the speaker is in a certain motivational state. (Note that Copp's view thereby also takes on the assumption, noted in regard to Barker's view, that *expressing* a mental state amounts to conveying that one is in it).

In developing this proposal, Copp makes an analogy with pejorative terms. The idea is that "Donating to charity is good" conveys that the speaker is at least somewhat motivated to donate in the same way that "Mark is a cheesehead" conveys that the speaker has a derogatory attitude towards people from Wisconsin. Pejorative terms such as ‘cheesehead’, in turn, are supposed to convey such attitudes through conventional implicature. Copp labels this aspect of a term’s meaning ‘Frege-coloring’, after Frege’s discussion of how a term can be ‘colored’ by an affective dimension going beyond its denotation. (For instance, “your cur howled all night” has the same truth-condition as “your mongrel dog howled all night”, but additionally conveys that the speaker has a derogatory attitude towards mongrel dogs) (Copp, 2001, p. 15).

Copp proposes four tests for determining whether a term exhibits ‘Frege-coloring’. A term t, in “X is t” has ‘Frege-Coloring’ if it passes:

1. The truth test: The belief expressed by someone uttering “X is t” might be true even if the implicated content is not.
2. Detachability: One can make an assertion with the same propositional content as “X is t” using a different sentence that does not generate the implicature.
3. Cancellability: A speaker can *intelligibly* deny the implicature without retracting her assertion of “X is t”.
4. Misuse: Asserting “X is t” when the implicated content is false is a misuse of ‘t’ and so is linguistically inappropriate.

Note that Copp’s cancellability test is non-standard. It is generally thought that conventional implicatures *cannot* be cancelled, where cancellability is understood not in terms of ‘intelligibility’, but rather linguistic appropriateness.[8] Copp’s ‘misuse test’ is actually closer to the standard understanding of cancellability. This is theoretically significant, given the point that there is a connection between the tests of detachability and cancellability: it seems that part of the explanation for why conventional implicatures in general are non-cancellable is that they are

[8] See Finlay (2004, p. 15) for discussion of this point and how it relates to the plausibility of Copp’s account.

detachable. I cannot felicitously *deny* that I have a derogatory attitude towards people from Wisconsin after I assert “Mark is a cheesehead”, because if I lacked such an attitude, it would have been more linguistically appropriate for me to have asserted something with the same truth-conditional content but that did not carry the implicature, as in: “Mark is from Wisconsin” – and I could only do that if the implicature in question is detachable.

Setting this issue aside for now, consider whether ethical terms such as ‘good’ pass Copp’s tests. Copp, of course maintains that they do:

1. The truth test: The belief expressed in an assertion of “T is good” (on Copp’s account, roughly, that T conforms to the relevantly justified moral standards) may be true while the implicated content – namely, that the speaker subscribes to those standards – is false.
2. Detachability: Copp maintains that the expression of a motivational state in asserting “T is good” *is* detachable. One might say “T is ‘good’”, or “T has the property of conforming to the relevantly justified moral standards”, without thereby conveying that one subscribes to those standards.
3. Cancellability: Copp maintains the implicated content is cancellable; he maintains that after asserting “T is good”, one could *intelligibly* continue: “Not that I care about doing what is good”.
4. Mis-use: To disavow the implicated content would be linguistically inappropriate. Even if one could *intelligibly* disavow the implicated content (as discussed just above), to do so would nevertheless be infelicitous, a misuse of moral terms.

Given these results, ethical terms would indeed have Frege-coloring. However, it seems unclear that ethical terms really do pass these tests, and it is unclear that these tests are the correct tests for conventional implicature. For instance, those drawn to motivational internalism of a strong sort hold that as a matter of conceptual necessity, one cannot rationally and sincerely judge that phi-ing is good without being at least somewhat motivated to phi. Accordingly, internalists will likely deny Copp's predictions when it comes to cancellability, even given Copp's non-standard construal of that test. Presumably, an internalist would predict competent speakers to judge an attempted cancellation not only as linguistically inappropriate, but as not even conceptually coherent.

Likewise, Copp's idea that the implicated content is detachable is questionable. Consider "T is 'good'". Using inverted-commas around 'good', if it is to detach the implicature that the speaker has a certain motivational state, should not affect the truth-conditional content asserted (namely, that T conforms to relevantly justified moral standards). But we would ordinarily take someone who asserted "T is 'good'" to precisely *not* be asserting that T in *fact* conforms to the justified moral standards. Instead, we would take such a person to be indicating that although many people might *think* T is good, or that T is *considered* good in their society, T is not in fact good.[9] At the very least, it seems that intuitions are just not clear here, both regarding cancellability and detachability. Given that we don't have clear (or unified) linguistic intuitions, it seems premature to accept these results on the tests for 'Frege-coloring'. The lack of clear and unified intuitions is not just a problem for Copp's view, but for any implicature view, since we categorize implicatures according to features such as cancellability and detachability.

In sum: realist-expressivism, like Barker's implicature theory, employs conventional implicature to explain the practicality of ethical discourse. The view is realist, objectivist, and externalist, but seeks to explain what is plausible in the internalist idea by proposing that ethical claims conventionally implicate that the speaker is in a certain motivational state. However,

[9] This criticism of Copp's proposal appears in Finlay (2005, pp. 13-14).

Copp's analysis of conventional implicature itself is non-standard, and moreover, it is just not clear that ethical terms do in fact pass Copp's own proposed tests for whether a term has 'Frege-coloring'. Employing conventional implicature to explain moral motivation is an innovative idea, but it is not clear that it succeeds. We turn now to examine a family of ecumenical cognitivist views that employ an alternative innovation: *conversational* implicature.

3.2.3 Finlay (2004, 2005)

Finlay proposes his end-relational theory as an improvement over others of what Finlay calls 'indexical' theories of ethical language (including, for instance, simple speaker subjectivism, or Dreier's (1990) more sophisticated speaker relativism). At the same time, Finlay hopes to avoid the challenges to convention-based theories such as Copp's realist-expressivism and Barker's implicature theory. To this end, Finlay uses the notion of *conversational* implicature, rather than conventional implicature, to explain the practicality of moral discourse, from within his cognitivist 'end-relational' semantics for normative language.

Conventional implicatures, as discussed in the previous two sections, are generated by linguistic conventions governing the use of particular terms of phrases. By contrast, *conversational* implicatures are generated by features of conversations, rather than the meanings of the terms used alone. A *conversational implicature* is content that a speaker conveys by an utterance that (i) is different from the literal meaning of the utterance, where (ii) the conveyance of that content is generated by the assumption that the speaker is following the maxims of cooperative conversation (introduced in Grice, 1975).

The two central features of Finlay's view, then, are the end-relational semantics he proposes to account for the truth-conditional content of normative sentences, and the conversational implicature explanation he provides for moral motivation. The basic idea of

Finlay’s end-relational semantics is that normative predicates (such as ‘good’) refer to *relational* properties. Strictly and literally speaking, something is never ‘good’ *full stop* – it is always good *for* some end. A knife, for instance, is never *just* good, but ‘good for cutting meat’, or ‘good for stabbing enemies’, or ‘good as a decorative piece’, in virtue of some qualities of the knife. The relational properties Finlay has in mind, then, are between a set of standards, on the one hand, and objects, acts, states of affairs, persons, etc., on the other, which can meet or fail to meet the set of standards.

Now, when we use normative terms in conversation, we do not always explicitly articulate the relevant standards. According to Finlay, which standards are relevant is not always made explicit when those standards can be assumed in the context of the conversation. In particular, it is assumed that when a speaker makes a *moral* judgment, the relevant standards are the ones that the speaker accepts. Finlay uses this proposal to explain how it is that speakers express motivational states in making ethical claims. The idea is that in making an ethical claim, such as “donating to charity is good”, when the standard that donating to charity is supposed to meet in virtue of that action’s qualities is not specified, it can be assumed that the relevant standard is a moral standard that the speaker herself accepts, and that the speaker is therefore at least somewhat motivated to act on.

We can now see how Finlay’s view improves upon other ‘indexical’ theories. Take, for instance, a crude version of speaker subjectivism, and now consider another variation on the problem of lost disagreement. Suppose that S asserts “T is good”, and S approves of T because T has feature F. Now imagine another judger, W, who agrees with S at least that T has F. However, W does not share S’s pro-F attitude. And moreover, W does not believe that T has any further properties towards which W has a pro-attitude. According to simple subjectivism, it seems that

W should think that S speaks *truly* when S says that T is good. After all, W knows that S *does* approve of T, and that is the truth-condition for S's utterance according to simple subjectivism. But this is the wrong result, for simple subjectivism therefore predicts that there is no disagreement between S and W when, intuitively, we would think S and W *disagree* about whether T is good.

Finlay's diagnosis of this case is that indexical theories confuse the property of *being* good with the property of *making* good (2005, p. 8). According to the end-relational theory, we can say that *relative* to some end E, T either *is* good or it is not (no matter who is making the judgment). So, end-relational theory is no simple subjectivism. The disagreement between S and W is really a disagreement about whether F is a feature that *makes* T good. Thus, substantive disagreement has been regained.

We can also see how Finlay's end-relational theory is meant to improve over conventional implicature theories (such as Copp's and Barker's). As Finlay notes (2005, p. 13), conversational implicatures, unlike conventional implicatures, are *non-detachable* and *cancellable*. That is, when some content P is conversationally implicated by U's utterance, U can, with full linguistic appropriateness, go on to deny that P is true (so P is cancellable). So for instance, ordinarily if I say "Sam drank some of the beer", my utterance carries the conversational implicature that Sam did not drink all of it. But I could go on to cancel the implicature by adding; "and in fact she drank all of it". Additionally, if P is conversationally implicated by S's utterance U, there is no other utterance S could make with the same content as U that would not also carry the implicature. There is no other way for me to make an assertion with the very same propositional content as is expressed by "Sam drank some of the beer" without also generating the implicature. For instance: "Sam drank *a bit* of the beer" has the same implicature. And “Sam drank *all* of the beer”, though it does not have the implicature, expresses a different propositional content.

Finlay’s contention is that the expression of a motivational state in making an ethical claim is cancellable and non-detachable. The intuitive possibility of amoralists who make ethical claims is supposed to show, according to Finlay, that the implicature is cancellable. One can claim “donating to charity is good”, but then felicitously cancel the implicature by continuing, “but I do not accept any moral standards” (2005, p. 15).[10] Moreover, Finlay adds, the implicature is non-detachable (this comes out in Finlay’s criticism of Copp on detachability – see previous section). Accordingly, Finlay takes it that conversational implicature better explains the ‘data’ about the expression of motivational states in making ethical claims than conventional implicature.

Nevertheless, Finlay’s view is still subject to many of the issues already raised. For one, note that conversational implicature, like conventional implicature, amounts to a way for an utterance to *convey that* a speaker is in some mental state, which stands in contrast to the original expressivist idea that expressing a mental state is different from saying that one is in it (see 2.3, against the implicature account of the expression relation). Now, this is not immediately an issue for Finlay, since he makes no claim to be articulating an expressivist view in any way. But, as we shall see, this point does raise worries about adequately accounting for all the various aspects of moral disagreement, including disagreement in attitude.

An additional worry specific for the conversational implicature approach is that it cannot do full justice to the practicality of the ethical domain. Conversational implicature is a feature of *conversations*, and so does not appear in private thought. Thus, *at best*, the conversational

[10] I do not share Finlay’s intuitions about cancelability here.

implicature approach can account for the action-guiding nature of ethical claims, but it fails to account for any close connection there may be between inner ethical judgment and motivation. Finally, it seems possible to make ethical claims in private as well as in public. For instance, suppose that, while driving on my own and hearing about Russia's recent invasion of Ukraine, I exclaim, out loud, in frustration, "That is just so *wrong*". Private ethical claims seem to be expressive of motivational states just as much as ethical claims in the contexts of a conversation. But the conversational implicature approach does not explain this.[11]

3.2.4 Strandberg (2011, 2012)

Caj Strandberg has proposed a hybrid account of ethical discourse that is in many respects similar to Finlay's, so I shall only briefly discuss Strandberg's view and generally treat the views together. Strandberg's dual-aspect account of ethical discourse is at once less committal and more specific than Finlay's view. It is less committal in that it is intended to be compatible with *any* sort of cognitivism about ethics. Thus, a minor dialectical advantage of Strandberg's account over Finlay's is that it is not committed to the end-relational semantics for normative sentences Finlay provides. And the view is more specific in that Strandberg provides more detail about the relevant linguistic mechanisms involved (at least, compared to Finlay's work in his 2004 and 2005 papers). In particular, Strandberg's account proposes that ethical claims express motivational states via *generalized*, rather than particular, conversational implicatures. A generalized conversational implicature is a conversational implicature that does not depend on any specific features of the conversation it appears in. Certain phrases or sentence forms are associated with generalized conversational implicatures – this accounts for how the implicature can be generated in a variety of contexts. So for instance, "Sam drank some of the

[11] This criticism is raised in Fletcher (2014, pp. 192-195), and in Bar-On and Chrisman (2009, p. 153).

beer" carries a generalized conversational implicature, where it is the use of 'some' in this sort of sentence structure that generates the implicature. In the moral realm, according to Strandberg, it is *sentence-types*, such as "phi-ing is wrong" that carry implicatures about the speaker motivations, rather than specific ethical *terms* such as 'wrong'.[12]

The dual-aspect account Strandberg proposes is as follows (2012, p. 15):

> A person's utterance of a sentence of a sentence of a type according to which phi-ing has a certain moral characteristic, such as "phi-ing is wrong" conveys two things: (i) the sentence expresses, in virtue of its conventional meaning, the belief that phi-ing has a moral property. (ii) An utterance of this type of sentence carries a generalized conversational implicature, CGI, to the effect that S has a certain action-guiding attitude in relation to phi-ing.

This view constitutes a version of ecumenical cognitivism. It is also an externalist view (on one way of understanding the internalist/externalist distinction), since it allows that one can rationally and sincerely make an ethical judgment while lacking the appropriate action-guiding attitude altogether. In explaining what he takes to be the truth in motivational internalism, Strandberg hypothesizes that as we are raised, we are socialized not to do what is wrong, and to want others to avoid doing what is wrong (2012, p. 108). There is thus a *contingent* connection between ethical judgment and motivation. This socialization, moreover, is what explains the presence of the generalized conversational implicature; for having a certain non-cognitive attitude makes a sentence of the form "phi-ing is wrong" *relevant* to the purposes of moral conversation (including, for instance, persuading others to adopt such an attitude).

[12] Generalized conversational implicatures attach to specific sentence forms and phrases. In this respect, they are similar to conventional implicatures. However, these are two distinct categories; first, generalized conversational implicatures follow the pattern of particular conversational implicatures rather than conventional implicatures, in being nondetachable and cancelable, among other things. And second, generalized conversational implicatures, unlike conventional implicatures, depend on the assumption that speakers are obeying conversational maxims as laid out by Grice.

In addition to agreeing with Finlay concerning cancellability, Strandberg notes in further support of his dual-aspect account that, *unlike* conventional implicatures, conversational implicatures do not ‘project’ through embedded contexts. (Barker, recall, found the need to introduce a new take on validity because conventional implicatures do project through embedding.) If ethical terms express motivational states through conventional implicature, it would turn out that they do so even when appearing in embedded contexts, so that “If T is good, then R is good” would turn out to express (something like) approval, in direct contrast to what is supposed to be a main lesson of the Frege-Geach problem. Strandberg takes a more conventional line when it comes to the Frege-Geach point, and holds that an utterance of an ethical sentence in an embedded context does *not* express a non-cognitive attitude. If this more conventional approach is correct, the dual-aspect account would then have an advantage over conventional implicature views, since conversational implicatures do not project through embedding. (Consider, for instance: “If Sam had some of the beer, then we should order more” – an assertion of this conditional would *not* standardly convey that the speaker believed Sam did not have all of the beer).

Since Strandberg’s proposal is similar to Finlay’s, it is subject to many of the same worries, as well as the problems confronting the implicature approach in general. However, it is worth considering the possibility that *generalized* conversational implicatures, being associated with certain sentence and phrase forms, may also appear in the absence of a conversation (though such appearances could only be derivative). This might contribute to an implicature-based explanation of the connection between ethical judgment and motivation in purely private, inner cases of ethical judgment. If this is right, it would constitute an advantage of Strandberg’s view over Finlay’s.

3.2.5 Boisvert (2008, 2014)

Boisvert's 'expressive-assertivism', like the other views canvassed so far, is a version of ecumenical cognitivism. The ambition of the view, like other hybrid views, is to provide an explanation of how ethical claims express non-cognitive mental states while avoiding the Frege-Geach problem and other worries for pure expressivism. The main innovation we see with Boisvert's approach is that, in contrast to Barker, Copp, Finlay, and Strandberg, Boisvert's 'expressive-assertivism' employs speech-act theory, rather than Gricean implicature, in articulating a hybrid view. This is initially quite a promising idea, but Boisvert does not execute it in a way that allows the view to gain a real advantage over the implicature approach, as I shall argue below.

The initial promise of the speech-act model resides in the following point. The notion of implicature concerns propositional content that an utterance conveys but that does not give the utterance's truth-condition. In the ethical context, the idea is that ethical claims somehow convey *that* the speaker is in a certain non-cognitive attitude. Speech-act theory, by contrast, does not directly focus on the literal conventional meanings of utterances (understood truth-conditionally or otherwise), but instead with features of the acts performed in making such utterances (their illocutionary force). Thus, once we take up the framework of speech act theory for thinking about how motivational states get expressed, we can explain such expression in terms of features of the *acts* performed in making ethical claims, *rather than* in terms in terms of the content of the claims. The significance of this point is that it allows us to treat separately two aspects of ethical claims that traditional expressivists run together: the semantic content of ethical claims vs. the mental states speakers must be in to sincerely make those claims. As we shall see in this next chapter, this philosophical division of labor is independently plausible and key to preserving the expressivist insight while avoiding the Frege-Geach problem.

For our purposes here it is important to note that Boisvert does not go in this direction. On the contrary, Boisvert assumes that there is a close connection between the analysis of illocutionary force and the project of giving a semantic theory of a language. As he puts it, the 'conventional functions' (i.e. speech acts) speakers use moral predicates to perform are significant "since it is these conventional functions that play a role in providing a correct semantic theory for a language" (2008, p. 176).[13] This puts his view within the same framework as the implicature approach, for it forces the analysis of the expressive dimension of ethical claims into the service of providing an account of (some aspect of) their conventional linguistic meaning. The point I would like to highlight here is that this is not obligatory, and I would suggest that developing a plausible hybrid theory requires distinguishing the projects of providing a correct semantic theory for a language from the project of analyzing the acts speakers perform using language.[14] This represents the most fundamental difference between Boisvert's approach and my own to be presented in Chapter 4. Nevertheless, in other respects, Boisvert's view may appear superficially similar to the hybrid illocutionary act theory I shall propose, and so it will be worthwhile to examine some of the details of Boisvert's view and identify what is valuable in it.

[13] Boisvert explicitly sets aside consideration of Bar-On and Chrisman's (2009) ethical neo-expressivism since that view, unlike his own, does not hold "that the expressive content of an utterance of a moral sentence is a part of that sentence's conventional meaning" (Boisvert, 2008, fn. 8).

[14] One might wonder, in the spirit of Davidsonian radical interpretation, whether we would have to know which utterances count as assertions, which as questions, and so on, to begin developing a semantic theory of some never-before-encountered language. Perhaps. This still would not undercut the point that the analysis of illocutionary force is fundamentally a different project from the analysis of semantic meaning (see Davidson, 1979). The relevant idea for our purposes is Davidson's thesis of the 'autonomy of linguistic meaning', according to which "there cannot be a form of speech which, solely by dint of its conventional meaning, can be used only for a given purpose, such as making an assertion or asking a question" (1979, p. 13).

Searle (1979) proposes a taxonomy of speech acts that divides them into five basic types: assertive, directive, commissive, declarative, and expressive. Relevant to Boisvert's project are the categories of assertive, expressive, and (possibly) directive speech act. Most fundamentally, a speech act is an act that one can (but need not) perform in speech (or in thought) by saying (aloud or to oneself) that one is performing that act, as in: "I *command* you to fire!".[15] Boisvert understands assertive acts as acts that "describe the world as being a certain way"[16], expressive acts as acts that "express a certain mental state (as opposed to expressing *that* one is in that state, which would be an assertive)", and directive acts as acts in which one "directs one's hearer to do something" (2008, p. 174). The central idea of expressive-assertivism is that ethical claims are at once assertive and expressive acts. They describe some moral way that the world is, and they express a motivational state of the speaker.

The main features of Boisvert's expressive-assertivism are captured in the following principles (2008, p. 171):

- Dual-Use Principle: If a speaker correctly and literally[17] utters a basic ethical sentence S, then the speaker performs a direct expressive and a direct assertive illocutionary act.
- Extensionality Principle: If a speaker correctly and literally utters a sentence with an ethical predicate in an extensional context, then the speaker performs a direct expressive illocutionary act.[18]

[15] See Austin (1962) and Searle (1979) for classic articulations of speech act theory.

[16] I take issue with this description of assertive acts, at least because it unnecessarily introduces descriptivist assumptions about their function. A more neutral starting point might be to say that assertives present some proposition as being true. Additionally, it seems to me essential to mention the effect that speech acts are to have on their *recipients* in characterizing their function – in the case of assertives, the function should be to bring about a certain *belief* in the hearer.

[17] Boisvert uses 'correct and literal' to rule out cases otherwise problematic to the connection he forges between semantic meaning and speech act, including (for instance) cases where one might recite an ethical sentence as part of a play.

[18] Notice the differences between this and the Dual-Use Principle. In the Extensionality Principle (EP), Boisvert does not specify that the sentences must be a 'basic' ethical sentence, so EP applies to complex sentences containing

- Generality Principle: If a speaker correctly and literally utters a basic or complex ethical sentence, the speaker performs a direct expressive illocutionary act of expressing some conative attitude towards things of a certain kind – things having the property denoted by the ethical predicate(s) used in the sentence.[19]

To give a rough gloss on the view, the idea is that a correct and literal utterance of "Tormenting the cat is wrong" is a direct assertive act that describes tormenting the cat as having some moral characteristic, F. So, the asserted part of the claim is: "Tormenting the cat is F". And the claim is also a direct expressive act, giving voice to a con-attitude towards things that are F. So, the expressive content of the claim is: "Boo for things that are F!"

We can further categorize Boisvert's view by noting that it is a fully realist, objectivist, cognitivist proposal: ethical predicates, on this view, denote moral properties with the same metaphysical status as non-moral properties in a way that is not speaker-relative. Moreover, the view is externalist (on a conventional understanding of the externalist/internalist distinction), as it holds that the mental states expressed by "Tormenting the cat is wrong" – belief that tormenting the cat is F, and disapproval of things that are F – are distinct and do not necessarily accompany each other.

An interesting feature of Boisvert's view is that it, like Barker's implicature theory, holds that moral attitude expression 'projects through' embedding in *extensional* contexts (e.g., conditionals), but not *intensional* ones (e.g., attitude ascriptions). This is the point of the

ethical terms, such as conditionals. But EP is limited to extensional contexts, so this excludes attitude ascriptions involving ethical terms, since those are intensional contexts. This is all to set up Boisvert's idea that moral attitude expression projects through extensional contexts but not intensional contexts. This pattern of projection is crucial for explaining how carrying out a moral inference can generate commitment to having a certain motivational attitude (a problem Schroeder raises for other hybrid theories), while avoiding implausible predictions about moral attitude ascriptions (to be described below).

[19] Thus, Boisvert – like Barker, Copp, and Ridge (as we shall see), takes the motivationally-charged attitudes involved in moral judgment to be directed at a general property of things, rather than to particular acts, policies, or persons.

extensionality principle. So, even a correct and literal utterance of "If torturing the cat is wrong, getting your little brother to torture it is wrong" would express the speaker's disapproval of things with a certain characteristic F, Boisvert contends, this is not problematic because the attitude expressed is directed *not* at torturing the cat, but instead at the property, F, picked out by 'wrong'. So, someone who correctly and literally utters "If torturing the cat is wrong, then getting your little brother to torture the cat is wrong" would express a con-attitude towards whatever has some feature F picked out by 'wrong', rather than an attitude towards torturing cats, or getting others to torture cats. Boisvert emphasizes this point in describing how view avoids Frege-Geach type worries; the 'expressive content' of ethical sentences, says Boisvert, is not compositional, while the 'descriptive content' is. Thus, the validity of good moral arguments turns entirely on the logical relations between their descriptive contents and has nothing to do with their expressive meaning.

A main advantage Boisvert identifies for expressive-assertivism concerns how it handles attitude ascriptions. This is important to consider, as this topic is a major theme of Schroeder's comprehensive critique of nearly all existing hybrid theories – and Boisvert's account handles the issue the best, at least according to himself and Schroeder. To see the problem, consider the following sentences, imagining that Jeremy endorses a utilitarian moral perspective, while Immanuel endorses a Kantian moral perspective:

a. Jeremy: "Donating to charity is right."
b. Immanuel: "Jeremy believes donating to charity is right."
c. Jeremy: "Donating to charity is rarely done."
d. Immanuel: "Jeremy believes donating to charity is rarely done."

And for later comparison:

e. Jeremy: “Mark is a cheesehead”

f. Immanuel: “Jeremy believes Mark is a cheesehead”

It might seem that Boisvert’s account (and traditional expressivism as well) can only explain what attitude is ascribed to Jeremy in (b) and (d) by implausibly positing an ambiguity in ‘believes that’. In particular, a pure expressivist will hold that in (b), Immanuel attributes a non-cognitive attitude to Jeremy, whereas in (d), Immanuel attributes a representational belief to Jeremy. A hybrid theorist like Boisvert will add that each of (b) and (d) ascribe a belief to Jeremy, but that (b) and not (d) also ascribes a motivationally-charged attitude to Jeremy; the problem arises again, because then it seems that ‘believes that’ is ambiguous, attributing just a belief in some cases, but a psychologically complex combination of belief and motivational state in other cases. Boisvert rejects this line of thought by denying the assumption that ‘believes that’ would be ambiguous if the mental states attributed when a nonethical sentence is used as the complement is of a different type than when an ethical sentence is used as the complement. Boisvert’s proposal instead is that the semantics of ‘believes that’ is such that in (b) and (d), the attitude attributed to the subject has the same content as the complement sentence. Because the contents of the complement sentences differ (in (b): that donating to charity is F, plus approval of F things; in (d): that donating to charity is rarely done), the attitudes attributed will differ. The idea is that ‘believes that’ attributes to the subject all the mental states that would be expressed by a correct and literal utterance of the complement sentence on its own. This removes the ambiguity from ‘believes that’.

In this respect, Boisvert thinks that ethical terms are like pejoratives, which also seem to have both descriptive and ‘expressive’ content. To illustrate: the thought is that in (e), Jeremy expresses the belief that Mark is from Wisconsin, and expresses a derogatory attitude towards people from Wisconsin. And in (f), Immanuel attributes each of those psychological states to Jeremy. Thus, attitude ascriptions that involve a pejorative term attribute a derogatory attitude to the *subject,* rather than expressing a derogatory attitude of the *speaker*.

This result seems at odds with Copp's proposal – a proposal that also takes the analogy with pejoratives seriously. According to Copp, for terms that have 'Frege-coloring' – such as pejoratives – the 'coloring' projects through embedded contexts, resulting in the prediction that in (f) Immanuel expresses *his own* derogatory attitude, rather than attributing such an attitude to Jeremy. While Boisvert agrees that expressive content 'projects through' in *extensional* contexts, he does not say that it does in *intensional* contexts, and so Boisvert does not hold that Immanuel expresses his own derogatory attitude in (f).

Here, I suspect intuitions may not be clear or unified, and that context may play a role in settling whether a speaker expresses their own derogatory attitude, or instead attributes a derogatory attitude to the subject in making an attitude ascription containing a pejorative term. (It may also depend on the specific term in question.) Nevertheless, the way in which expressive-assertivism avoids the problem given above (i.e., of predicting ambiguity in 'believes that') is especially significant because it is in effect an explicit endorsement of what Schroeder (2009, 2014) calls the 'Big Hypothesis', to be discussed below.

In sum: Boisvert's view is similar in many respects to Copp's realist-expressivism, as both endorse forms of externalist, realist, objectivist, ecumenical cognitivism. The main point of contrast concerns their explanatory focus; Copp proceeds by employing conventional implicature, while Boisvert's view is set in speech-act theory. I have raised some concerns, to be elaborated on later, about the way Boisvert executes the speech-act strategy in hybrid theorizing. Additionally, some of Boisvert's predictions about the behavior of ethical terms in intensional contexts seem to involve questionable assumptions about how attitude ascriptions and how pejoratives work, as we shall see in Schroeder's criticisms discussed below.

3.2.6 Ridge (2006, 2014)

Ridge proposes his 'ecumenical expressivism' as a view that, like other hybrid theories, hopes to retain what is plausible in traditional expressivism while avoiding its problems, most notably the Frege-Geach problem. However Ridge sets his view apart from the alternatives considered so far by arguing that his is the 'true heir' to the mantle of expressivism. While the other views surveyed should be understood as ecumenical variations on *cognitivism*, Ridge hopes to stay true to the core commitments of expressivism, including its anti-realist commitments. Whereas ecumenical cognitivists are generally concerned to use expressivist tools to account for the distinctive practicality of ethical thought and discourse, Ridge finds that the expressivist perspective also has advantages in its analysis of Moorean Open Question phenomena, explaining supervenience, and in accounting for the intractability and pervasiveness of moral disagreement (2006, pp. 305-310; 2014, Chapter 2). In short, Ridge finds certain arguments against various versions of moral realism convincing, and he wants to improve upon *irrealist* expressivism, addressing its most serious problems by conceding some truth to ethical cognitivism.

In effect, then, it is dissatisfaction with leading realist theories, together with a concern to explain the practicality of the ethical domain – the two traditional motivations for *pure* expressivism – that lead Ridge to his ecumenical expressivism. We shall now have occasion to see whether the *ecumenical* expressivist approach does better than traditional expressivism.

One of the main problems for traditional expressivism, of course, is the Frege-Geach problem, and it is the central project of Ridge's (2006) paper to grapple with this problem head-on (in his 2014, he takes up various other problems for pure expressivism, including, for instance, accounting for moral *truth*). As Ridge sees it, one of the main tasks for expressivists in addressing the Frege-Geach is to explain the features of good moral arguments, including (see also 2.5.2):

> *Validity*: provide a general account of validity (for both moral and non-moral arguments), on which good moral arguments will turn out valid, and bad moral arguments will not.
>
> *Inconsistency-Generation*: explain why it would be *logically* (and not merely practically or pragmatically) inconsistent for someone to accept the premises of a valid moral argument yet reject its conclusion.
>
> *Inference-Licensing*: explain how it is that someone could rationally come to accept the conclusion through accepting the premises and drawing the inference.

Ecumenical expressivism, it is maintained, explains all these features.

The basic structure of Ridge's ecumenical expressivism is as follows. Like traditional expressivists, Ridge starts from within a broadly mentalist understanding of meaning, and proposes a bifurcation between the sorts of mental states expressed by ordinary descriptive claims and the sort of mental state expressed by ethical claims. As we saw in Chapter 2, traditional expressivists sought to introduce greater complexity in the mental states they say constitute ethical judgment than their emotivist forebearers had suggested. Rather than likening moral judgments to inner 'boos' and 'hurrays', Gibbard, for instance, proceeded in terms of a complex theory of planning states, relying on the notion of a hyperdecided planner to help fix content. The hope was that by building more logical complexity into the models of the non-cognitive states constituting ethical judgment, we could build an expressivist semantics that shares the advantages of standard truth-conditional approaches.

Ridge gives up on the idea of specifying logically complex but purely non-cognitive states to do this important semantic work. Instead, he suggests that ethical judgments are not *wholly* non-cognitive; they incorporate a representational belief in addition to a non-cognitive 'normative perspective'. Ridge's hope is that by introducing a genuinely representational belief as an element of his analysis of ethical judgment, he can then rely on the logical and epistemic features of that belief to address the Frege-Geach and related problems.

By introducing a genuinely representational belief into the analysis, Ridge's account does not simply collapse into cognitivism, or even ecumenical cognitivism, as one might expect. This is because he denies that the truth-conditions of ethical claims are guaranteed to be given by the representational beliefs they express. Instead, what representational belief one would express with a sincere utterance of an ethical sentence is determined in part by one's normative perspective: "a set of relatively stable self-governing policies about which standards to reject or accept" (2014, p. 115). Since different agents can have different and conflicting normative perspectives, different agents can express different representational beliefs by uttering the same ethical sentence. But the view does not collapse into subjectivism or relativism, because it says that when assessing the truth of normative claims, one is to be guided by one's *own* normative perspective, not necessarily the normative perspective of the individual making the claim, as subjectivism would require (Ridge 2014, p. 147). To illustrate: When Jeremy says "Torture is wrong", we can say that Jeremy expresses a utilitarian normative perspective recommending actions insofar as they maximize happiness, and the belief that torture fails to maximize happiness. When Immanuel says "Torture is not wrong", we can say that Immanuel expresses a Kantian normative perspective recommending only those actions whose maxim can be willed as universal law, and a belief that it is not the case that torture cannot be willed as universal law.

Notice that Jeremy and Immanuel's normative perspectives recommend certain actions insofar as they exhibit some property (e.g. maximizing happiness). This reflects an aspect of Ridge's earlier (2006) construal of ethical judgment, where he specifies that moral attitudes are keyed to properties of actions. According to Ridge, uttering a sentence of the form "There is moral reason to X" would serve to express "(a) an attitude of approval of a certain kind towards actions insofar as they have a certain property [this reflects an aspect of the judgers normative perspective] and (b) a belief that X has that property" (2006, p. 315). So for instance, S's utterance of "There is moral reason not to eat meat" expresses an attitude of approval towards actions insofar as they have a certain property, F, and the belief that not eating meat has that property. The property in question will be whatever property it is that guides the speakers attitudes of approval or disapproval generally (though the speaker need not have a clear conception of just which property this is). And the reference to this property in the belief expressed is supposed to work something like anaphoric reference, so that the belief expressed is that eating meat has *that* property, whatever it is, that guides the speaker's attitudes of approval and disapproval in general (Ridge 2006, p. 313).[20] Since normative judgments "involve broadly desire-like states" (Ridge 2014, p. 113), and are an essential component of ethical judgment, Ridge's view is also committed to a form of internalism, unlike the other ecumenical cognitivist theories considered so far.[21]

The basic strategy Ridge takes to address the Frege-Geach problem is to derive the logical complexity needed for an adequate response from the content of the representational

[20] As Schroeder notes however, the claim that the belief expressed is anaphoric in this way cannot be taken as *literally* true, since anaphor requires a syntactic antecedent, which is absent in this case (2009, p. 293).

[21] The variety of internalism Ridge accepts is what he calls "capacity judgment internalism", according to which first person normative judgments are necessarily capable of motivating without the help of any independent desire (2014, p. 49).

beliefs expressed by the moral utterances in question (2014, p. 144). Ridge takes a two-pronged strategy to account for the validity of good moral arguments, reflecting his interest in remaining neutral on the question of *truth* for normative judgments. Ridge prefers to keep open the option of following the early emotivists in holding that ethical claims are not truth-apt, while also being able to talk of ethical truth, in either a deflationary sense *or* in terms of a robust truth property (2014, Chapter 7). The first prong is to articulate for an understanding of validity, not in terms of truth-preservation, but in terms of consistency in belief; this would allow Ridge to maintain good moral arguments are valid even if ethical claims turn out not to be truth-apt, so long as they express (in part) representational beliefs (see Ridge 2006, p. 326; 2014, p. 156). The second prong: on the assumption that ethical claims *are* truth-apt, of course, Ridge can then adopt a more standard understanding of validity in terms of truth-preservation. In any event, the overall idea is that the representational beliefs partly constituting ethical judgment have the needed logical features to contribute to valid arguments on either approach to validity considered here.

Concerning the inference-licensing and inconsistency-generating features: Suppose that Jeremy considers the following argument:

1. If lying is wrong, getting your little brother to lie is wrong.
2. Lying is wrong.
3. Therefore, getting your little brother to lie is wrong.

Jeremy’s state of accepting premise 2 is constituted by his utilitarian normative perspective, and the representational belief that lying fails to maximize happiness. In coming to accept premise 1, Jeremy will (again) have a utilitarian normative perspective, and the representational belief that if lying fails to maximize happiness, then getting your little brother to lie fails to maximize happiness. Jeremy’s accepting the conclusion consists in his (again) having a utilitarian normative perspective, and the representational belief that getting your little brother to lie fails to maximize happiness. It would be inconsistent for Jeremy to accept premise (1) and (2) yet deny (3), for that would commit him to believing both that getting your little brother to lie fails to maximize happiness, and that getting your little brother to lie does not fail to maximize happiness; thus, the inconsistency-generating feature is preserved. If Jeremy accepts (1) and (2), and yet neglects to accept (3), even when explicitly considering it, Jeremy seems to be acting epistemically irresponsibly on Ridge's account. For then Jeremy would be failing to believe something that he knows to be a consequence of other things he believes; thus, the inference-licensing feature is preserved (Ridge 2014, pp. 166-168).

It is significant on Ridge's proposal that two speakers might utter the same moral sentence, and yet express representational beliefs with different contents, because the speakers have different normative perspectives. So, when Jeremy utters "Lying is wrong", the belief expressed would be that lying has the property of failing to maximize utility, and when Immanuel utters "Lying is wrong", the belief expressed would be that lying fails to treat rational agents as ends in themselves and not as mere means. This would appear to raise objections of the sort articulated by Boisvert (and further pressed in Schroeder, 2009) when it comes to attitude ascriptions. Consider the dilemma; when Immanuel utters "Jeremy believes lying is not wrong", what is the representational content of the belief ascribed? If that content adverts to utilitarian standards, then it appears Immanuel could not disagree with Jeremy by continuing: "But lying is wrong". Yet if the content ascribed adverts to Kantian standards, then it appears to misrepresent Jeremy's belief.

Ridge's response to this problem is to maintain that Immanuel's utterance "Jeremy believes lying is not wrong" attributes to Jeremy a normative perspective/representational belief pair (Ridge 2014, Chapter 5, §6). However, Ridge does not maintain that the content of the belief Immanuel ascribes to Jeremy is guaranteed to be fixed by the content of whatever representational belief *Jeremy's* utterance of "lying is not wrong" expresses. Instead, the idea is that the representational content of the belief Immanuel ascribes to Jeremy is *itself* the subject of normative debate between Immanuel and Jeremy. When Immanuel assesses Jeremy's assertion of "lying is not wrong", this is relative to *his own* (i.e. Immanuel's) normative perspective, rather than adopting Jeremy's (Ridge 2014, p. 147). Thus, Jeremy and Immanuel have a substantive disagreement concerning how to specify the content of the representational belief expressed by an utterance of "lying is not wrong". Crucial here is Ridge's idea that there is more to a proposition than its representational content; given this, the proposition expressed by Immanuel's utterance is just that <Jeremy believes that lying is not wrong>, so Immanuel can correctly report Jeremy's belief. But at the same time, Immanuel and Jeremy would disagree about how to best specify the representational content of that proposition. This is what their disagreement consists in.

This view is further supported by Ridge's account of truth (2014, Chapter 7). Ridge argues that for a proposition to be true is for its representational content to have whatever is required for truth – be it correspondence to reality, or being accepted at the end of all inquiry, etc.[22] – on any admissible specification of the proposition's representational content, where the notion of 'admissible specification' is itself a normative one. Different agents may have different perspectives on what counts as an admissible specification of representational content. Jeremy

[22] Incidentally, this is how Ridge purports to show that ecumenical expressivism is not confined to holding either that ethical sentences are not truth-apt at all or that they are only truth-apt in a minimal sense. In this respect, Ridge's (2014) *meta*semantic ecumenical expressivism is supposed to be superior to earlier versions of the view. Ridge argues that his view is compatible with *any* view of truth, since we can simply substitute any view about the nature of truth into the account.

and Immanuel disagree about whether the proposition <lying is wrong> is true because, given their respective normative perspectives, they assign that proposition different representational content. Their disagreement is in part a disagreement about the representational content of the proposition that lying is wrong.

Ridge's stance on what's at issue in the disagreement between Immanuel and Jeremy appears at first to confuse moral disagreement with a disagreement about language. For according to Ridge, the disagreement between Jeremy and Immanuel is at least in part over how to construe the representational content of the proposition that lying is wrong – seemingly a meta-linguistic disagreement. This seems a strange result, since we would ordinarily describe Immanuel and Jeremy as having a *moral* disagreement, not a meta-linguistic one. However, Ridge's point seems to be that these are not separate issues; that in this case, the meta-linguistic disagreement just *is* a moral disagreement.

This is a complex and ingenious quasi-realist response to the problem of attitude ascriptions raised by Boisvert and Schroeder. Despite the technical adequacy of Ridge's account in responding to the various problems just discussed, however, it seems Ridge's view is still subject to some of the problems for quasi-realism outlined in Chapter 2. In particular, as I discussed in 2.5.4, Ridge's account still seems to face a problem when it comes to accounting for the possibility of fundamental moral error. When it comes to the non-cognitive attitude, or what Ridge (2014) calls the 'normative perspective' involved in ethical judgment, ecumenical expressivism has no new resources to offer that pure quasi-realist expressivism does not already have. So, if ecumenical expressivism is going to gain any purchase on the problem of fundamental moral error, it must be found in the representational belief involved in moral judgment. This may look promising at first. For it might seem that ecumenical expressivism could explain the possibility of fundamental moral error in terms of recognizing the possibility that the representational belief component of one's ethical judgment may be mistaken even when there are no possible improvements that can be made to one's epistemic position (by one's own lights). For Ridge maintains that the representational belief involved in ethical judgment is just like ordinary representational beliefs in straightforwardly factual domains; thus, given the possibility of fundamental error in *other* domains, there should be no special problem of fundamental error in the moral domain.

However, as we saw at the end of Chapter 2, it seems to me that this solution is only superficial. For it does not really explain fundamental *moral* error – it accounts for error in representational belief, but *not* error in normative perspective, which is what is required to explain fundamental moral error (2.5.4).[23] So, consider again Jeremy's ethical judgment that lying is not wrong, which combines a utilitarian normative perspective with the representational belief that lying does not fail to maximize happiness. Ecumenical expressivism can account for error, and indeed fundamental error, because it is open to Jeremy to consider the possibility that his belief that lying does not fail to maximize happiness might be mistaken, and might continue to be mistaken now matter how much his epistemic position improves by his own lights. However, this is not to recognize the possibility of fundamental *moral* error. In order to do *that*, Jeremy would have to be able to consider the possibility that his utilitarian normative perspective itself is mistaken, and that it might be mistaken no matter how much Jeremy's epistemic position would improve by his own lights. Again, here, ecumenical expressivism lacks the resources to do any better than pure quasi-realist expressivism.

3.2.7 Eriksson (2009, 2014)

[23] Bykvist and Olson (2009) also press this objection against Ridge's ecumenical expressivism.

Eriksson has proposed a Hare-inspired version of ecumenical expressivism as a competitor to Ridge's view. In their core details Eriksson's and Ridge's views are quite similar, so I shall not spend much time treating them separately, except to indicate some points of contrast.

The main features of Eriksson's view are that:

- Ethical terms have both an evaluative and a descriptive meaning.
- The evaluative meaning is *primary* and *constant*; the descriptive meaning is *secondary* and *non-constant*.
- The evaluative attitude expressed is directed at the *subject* of the evaluative judgment, rather than at some general characteristic.

Eriksson's view qualifies as ecumenical expressivist, rather than ecumenical cognitivist, because it takes the 'evaluative meaning' of ethical terms to provide their primary semantic function. Now, Ridge's ecumenical expressivism holds that different speakers give voice to different evaluative attitudes using the same ethical terms, because the evaluative attitude expressed is a function of the speakers normative perspective. By contrast, Eriksson holds that different speakers express the *same* type of evaluative attitude when using the same ethical terms; for instance, any speaker's use of 'good' will express a general attitude of commendation. In turn, the *descriptive* meaning of ethical terms is held to be non-constant, because different judgers will commend things on different grounds, and the descriptive content of 'good' consists in the grounds on which a particular judge calls things 'good'.

Now, an initial worry I see for Eriksson's Hare-inspired ecumenical expressivism is that it, like Boisvert's expressive-assertivism, appears to fuse a speech-act analysis with a semantic one. For 'commending', 'condemning', and the like, each can be understood as acts that we perform with the use of language, rather than elements of the meaning of linguistic items themselves. But Eriksson seems to endorse treating commending/condemning/etc. as giving the literal meanings of ethical terms. In accordance with traditional expressivism, he holds that the meanings of ethical sentences are to be given in terms of the mental states they express, and on Eriksson's view, the relevant mental states are states of commendation/condemnation/etc., where it is then maintained that these states have both an evaluative and a descriptive component.[24] I set this aside for now, since these issues will occupy us in Chapter 4.

Eriksson's strategy for addressing the Frege-Geach problem is slightly different from Ridge's. The idea is to use the descriptive meaning of ethical terms to account for the validity of moral arguments. But where Ridge (like Barker) finds it necessary to introduce a non-standard definition of validity to allow for the possibility that ethical sentences are not truth-apt, Eriksson is happy to keep a standard understanding of validity in terms of truth-preservation. The idea then is that, holding a standard for judging fixed, it will be logically impossible for the descriptive contents of the premises of a valid moral argument to be true while its conclusion is false.

However, this approach to explaining validity in terms of truth-preservation requires that the descriptive contents of ethical sentences give their truth-conditions. In turn, it will turn out that ethical sentences have subjectivist, or perhaps culturally-relative, truth-conditions. Eriksson concedes this point, and reports that he is "not sure how worrying this is", since it would still be the case on his view that the main purpose of ethical claims is not to *report* the speaker's attitudes, but to guide action (2009, fn. 44). Given that this is the point of making ethical claims, Eriksson takes it that one could grant that the descriptive content of an ethical claim M supplies

[24] Notice the slippage here between thinking of commending as an *act*, and thinking of commending as a *mental state*. Making a careful distinction here is key to seeing the advantages of the neo-expressivist account to come.

its truth-condition, *and* one could recognize this content *as* M's truth-condition, yet coherently still *reject* M as true, because one does not endorse the standard by which M is judged. To illustrate: According to Eriksson, Immanuel could recognize that Jeremy's ethical claim "lying is wrong" has, as its truth condition, that lying fails to maximize happiness, and Immanuel could accept that it is true that lying fails to maximize happiness, yet Immanuel could coherently still reject "lying is wrong" as true, because Immanuel judges by a different (non-utilitarian) standard. This strikes me courting incoherence. What could possibly be meant by 'truth-condition' such that one could rationally accept that M's truth-condition is that p, and accept that p is true, yet reject M as true? At the very least, further explanation is required on this point.

In sum: Generally, Eriksson's view is similar to Ridge's. Each endorses a version of ecumenical *expressivism*. Each holds that there is a close, but defeasible connection between moral judgment and motivation – Ridge, in terms of 'capacity judgment internalism' (2014, pp. 50-51), and Eriksson, in terms of a disposition of moral judgment (2009, fn. 46). And each endorses an expressivist, mentalist approach to (meta)semantics. Eriksson's view departs from Ridge's in some of its details, including its position on the truth-conditions of ethical sentences, and on the nature of the non-cognitive attitude and representational beliefs involved in moral judgment. Some of the general worries for Ridge's ecumenical expressivism may carry over to Eriksson's, such as the problem of fundamental moral error. But additionally, Eriksson's carries the questionable commitment to the idea that recognizing a sentence's truth-condition *as* its truth-condition *and* as true is not sufficient for accepting the sentence as true.

3.2.8 Schroeder's criticisms of hybrid theories (2009, 2014)

This chapter so far has primarily focused on direct engagement with prominent hybrid theories, at once summarizing and assessing each view, with the overarching goal of making clear how the hybrid view I discuss in the next chapter is distinct from and more plausible than any of the existing hybrid theories on offer. It has not generally been my concern to consider critical responses to hybrid metaethics in general, and indeed I think the basic hybrid strategy is the right idea for explaining at once the continuities between ethics and other domains as well as what is distinctive about ethical thought and discourse.

However, it is necessary for me to consider the criticisms proposed in Schroeder's "Hybrid Expressivism: Virtues and Vices" (2009) and "The Truth in Hybrid Semantics" (2014). This is because, first, Schroeder has posed an influential, yet I think mistaken, taxonomy of existing hybrid views and the logical space of hybrid metaethics. The very questions Schroeder uses to construct this taxonomy obscure the possibility of the hybrid theory I want to propose, so this taxonomy must be corrected before I can present that theory. And second, the assumptions leading Schroeder to set the issues out in the way he does are by no means idiosyncratic: they are shared and even explicitly endorsed by some of the theorists Schroeder discusses, and they continue to shape subsequent developments in the hybrid literature. Challenging these assumptions is thus essential not only to establishing the theory I propose as a theoretical option; it will also open up new ways forward in subsequent theorizing.

Schroeder's overall thesis in his comprehensive "Hybrid Expressivism: Virtues and Vices" (2009) is that insofar as hybrid theories gain an advantage over pure expressivism, they risk losing any advantages over pure cognitivism, *unless* they commit to a controversial thesis – what Schroeder dubs "the Big Hypothesis" – about how moral attitude ascriptions work. Schroeder's strategy is to present four questions intended to taxonomize existing hybrid views, argue that the most plausible hybrid theory answers "no" to all four questions, and then to argue that such theories still only get an advantage over pure cognitivism by accepting the controversial “Big Hypothesis”. My primary interest will be in the taxonomy Schroeder offers; as we shall see, once we reject this taxonomy in the correct way, the possibility of a plausible hybrid theory that does not endorse the Big Hypothesis becomes available.

I survey Schroeder’s argument as it is presented, on his terms, before criticizing the adequacy of this taxonomical proposal.

Schroeder taxonomizes hybrid theories as follows:

Schroeder’s Four Questions (2009, p. 261)

1. Do different sentences containing the word ‘wrong’ express different desire-like states?
2. Do different speakers express different desire-like states with the same sentence?
3. Does a given sentence have a different descriptive content for different speakers?
4. Does the descriptive content of a sentence depend on the desire-like state it expresses?

With these four questions, Schroeder categorizes various metaethical positions views as follows (some are hybrid, some are pure cognitivist views that seek to do justice to expressivist intuitions, and some are pure expressivist that seek to do justice to cognitivist intuitions):

	Barker, Finlay, Copp[25]	Jackson	Ridge	Boisvert	Gibbard	Eriksson[26]
Q1	No	Yes	No	No	Yes	Yes
Q2	Yes	No	Yes	No	No	No
Q3	Yes	Yes	Yes	No	No	Yes

[25] We could also add Strandberg (2011, 2012) to this box, since Strandberg’s view is in its essentials very similar to Finlay’s.

[26] Schroeder does not consider Eriksson’s view, as Eriksson’s “Homage to Hare” (2009) was published after Schroeder’s taxonomy. However, Eriksson helpfully indicates how he would answer each of Schroeder’s questions. (This, of course, also confirms that Eriksson accepts Schroeder’s way of setting up the issues, which I shall question).

Q4	No	Yes	Yes	No	Yes	No

Fig. 1 Schroeder's taxonomy of hybrid theories

I now briefly summarize why Schroeder thinks a plausible hybrid theory must answer "No" to all four questions, as Boisvert's view does, and why he thinks such a view must accept the Big Hypothesis to gain a real advantage over pure cognitivism. Some of Schroeder's arguments have already been alluded to, and some of his criticisms anticipated by various hybrid theories, so I will aim to be brief with this review.

Why we must answer "no" to Q1 (pp. 268-275)

Simplifying greatly, Schroeder argues that a "no" answer is required to explain why accepting the premises of a valid moral argument commits one to having the desire-like state expressed by the conclusion. If we answered "Yes" instead, then it is possible that the desire-like attitude expressed by the conclusion is different from the desire-like attitude expressed in the premises, and so it would remain mysterious why accepting the premises would commit one to having the attitude expressed by the conclusion. Consider again the argument:

1. If lying is wrong, then getting your brother to lie is wrong.
2. Lying is wrong.
3. Therefore, getting your brother to lie is wrong.

If we answer "Yes" to Q1, a schematized hybrid theory would construe the mental states expressed by each line in the above argument as follows:

1'. *Belief* (if lying has property F, then getting your brother to lie is F); *Attitude* (A)

2' *Belief* (Lying is F); *Attitude* (B)

3' *Belief* (Getting your brother to lie has F); *Attitude* (*C*)

Even though the contents of the beliefs expressed might be used to explain the validity of the argument, it is unclear why someone who accepted the premises and inferred the conclusion from them would be rationally committed to holding attitude (C). But on the assumption that having attitude (C) is partly constitutive of making the ethical judgment that getting your brother to lie is wrong, it is not clear how drawing the inference above by manipulating the belief contents involved rationally commits one to forming that ethical judgment.

Why we must answer "No" to Q2 (pp. 278-284)

The problem for a "yes" answer to Q2 amounts to the difficulty of explaining what the expression-relation could be. To illustrate what a "yes" answer to Q2 amounts to: the idea would be that *which* desire-like mental state is expressed by an utterance of a moral sentence depends on the moral perspective of the speaker. So, if Jeremy utters "lying is wrong", that sentence expresses disapproval towards things insofar as they fail to maximize utility, whereas if Immanuel utters "lying is wrong", his utterance expresses disapproval of things insofar as they fail to treat rational agents as ends in themselves.

This generates a problem on the assumption – an assumption Schroeder finds in the views he is criticizing – that the expression-relation is an *evidential* notion, such that to express a mental state is *to intentionally indicate or convey to an audience that one is in that mental state*.[27] The problem, however, is that if the desire-like state a speaker supposedly expresses in

[27] Gibbard, for instance, understands 'expression' in Gricean terms, such that one expresses a mental state by "uttering words conventionally intended to get [one's audience] to think that [one] is in this state of mind" (1990, p. 86). And Ridge (2006, 2014) and Eriksson (2009, 2014) each follow the account of expression proposed by Davis, which is evidentialist and invokes Gricean intentions to convey one's mental state: according to Davis, S expresses (for instance) the belief that p by doing some action A intending for A to provide an indication that S believes that p (2002, Chapter 2). And the notion of *implicature*, as well, seems to be evidential at least in one sense, since an utterance carrying the implicature that the speaker is in M *conveys that* the speaker is in M -- so all implicature views will likely accept that ethical claims express motivational states in the sense that they provide evidence that the speaker is in some motivational state. It is unclear to me at this stage how Boisvert understands the expression relation.

making an ethical claim is idiosyncratic to that speaker, as a "yes" answer to Q2 predicts, there is no way for utterance to be informative for her audience as regards her motivational state. Either the audience is already aware of the speaker's normative perspective, in which case the utterance is not informative on this score, or the audience is not already aware of it, in which case the utterance of a simple moral sentence can provide no evidence to the audience exactly what idiosyncratic moral perspective the speaker has. Yet, if what it is to express a state is, by definition, to intentionally give evidence that one is in that state, then a "Yes" to Q2 is incompatible with holding that ethical claims express desire-like states. Schroeder concludes that in order for ethical claims to be informative to audiences about the speaker's mental states, they must express the same type of desire-like state for each speaker – so a "no" answer to Q2 is required.

Why we must answer "No" to Q3 (pp. 284-292)

Essentially, a "yes" answer to Q3 amounts to a contextualist stance on the descriptive content of ethical claims. The basic problem Schroeder identifies for this stance is a problem about attitude ascriptions already identified by Boisvert, though Schroeder adds an additional twist that raises some questions about how "true" behaves in ethical discourse. To illustrate this twist, consider the following dialogue, along with Schroeder's gloss on what descriptive content is expressed given a "yes" answer to Q3:

> Jeremy: "Lying to save someone's life is not wrong." → *Lying to save someone's life does not fail to maximize happiness.*
>
> Immanuel: "Jeremy said lying to save someone's life is not wrong." → *Jeremy said that _______.*

Immanuel: “But lying even to save someone’s life *is* wrong.” → *Lying even to save someone’s life fails to treat a rational agent as an end in themselves*.

Immanuel: “So what Jeremy said is false.” → *What Jeremy said is false*.

The dilemma, then, is how to fill in the blank so that Immanuel accurately reports the content of Jeremy’s claim, but also counts as disagreeing with that claim. If the blank is filled in with “lying to save someone’s life does not fail to maximize happiness”, Immanuel correctly reports Jeremy’s belief, but then Immanuel’s cannot rationally infer that what Jeremy said is false. And if the blank is filled with “lying to save someone’s life treats rational agents as ends in themselves”, then the inference is good, but Immanuel has not correctly described Jeremy’s belief. So, Schroeder concludes, it is only by answering “no” to Q3, and so holding that the descriptive content of ethical claims is constant across speakers, that we can explain how Immanuel could both accurately report Jeremy’s belief and rationally infer that what Jeremy said is false.

Why we must answer “no” to Q4 (pp. 292-297)

Schroeder’s discussion of Q4 is primarily intended to target Ridge’s ecumenical expressivism, since according to Schroeder’s reading of Ridge, it is the “no” answer to Q4 that sets Ridge’s account apart from ecumenical cognitivism. As Ridge would put it, his view is not cognitivist because he holds that the representational belief expressed by a moral utterance is not guaranteed to be true just when the utterance itself is true. The content of the representational belief expressed by, say, “there is reason not to eat meat” is a function of the normative perspective of the judger (e.g. a utilitarian perspective, a Kantian perspective, etc.), where this belief could be true even if the utterance is not; for instance, when the judger holds a mistaken normative perspective. Thus, on Ridge’s account, the content of the representational belief expressed by an ethical judgment is related to the normative perspective expressed; hence a "yes" to Q4.

Schroeder's complex complaint against a "yes" answer to Q4 is that it makes it difficult to explain how most thinkers could regard even their own moral inferences as rational. According to Ridge, the validity of a good moral argument is grounded in the fact that for any possible thinker, that thinker's normative perspective will determine the representational content of the beliefs expressed by the premises and conclusion of the argument in such a way that any particular thinkers' accepting the premises and rejecting the conclusion would guarantee that the thinker would have inconsistent beliefs. But according to Schroeder, in order for the *thinker* to regard her own inference as rational, she would have to be in a position to *recognize* that the contents of the beliefs expressed in her accepting the premises and denying the conclusion would have this property. As Schroeder puts it: "If you do not realize this, or if you believe that different sentences express different states of approval, then you will not be in a position to see that accepting the premises commits you to the conclusion, and so it will be perfectly rational for you to accept the premises and deny the conclusion" (2009, pp. 295-296).

It should be noted that I find Schroeder's argument for a "No" answer to Q4 questionable on its own terms. This is because the argument seems to turn on an implausibly strong conception of what is required for an agent to appropriately be able to consider the inferences she makes as rational. Schroeder assumes that to find a moral argument (and presumably any argument) valid, one would have to understand the rule by which moral (or any) sentences express beliefs (2009, pp. 294-295). As it happens, for non-moral sentences, the relevant rule is such that, ignoring contextual factors, a given sentence expresses the same belief no matter where it appears in the argument. But for moral sentences, on Ridge's account, the rule by which a sentence containing ‘wrong’ gets its descriptive content is by ‘anaphoric’ reference to that property, whatever it is, that guides the speaker’s approval and disapproval generally. And only given Ridge’s ecumenical expressivism will it be the case that the property so picked out is guaranteed to be the same for any particular thinker. Schroeder concludes, speakers could only find moral inferences rational if they were aware of and endorsed Ridge’s ecumenical expressivism. But again, it seems far too much to demand of ordinary reasoners that they not only *implicitly* rely on some linguistic rules in carrying out an inference, but that they must also be explicitly aware of those rules to regard their inferences as rational. But I set this issue aside for now, to complete the exposition of Schroeder’s complex argument.

Having purported to establish that a plausible hybrid theory must answer “no” to all four questions, Schroeder proceeds to argue that such a theory must also accept a controversial hypothesis in order to gain an advantage over pure cognitivism. Gaining an advantage over pure cognitivism requires being able to accommodate features of ethical thought and discourse that pure cognitivism purportedly has trouble doing. Two apparent such features – at least, features that Schroeder understandably thinks expressivists will take seriously – are Moorean Open Question phenomena, and accounting for moral motivation (in particular, explaining why internalism might be true as a matter of conceptual necessity). A view that does not explain these two features, or denies that they are genuine features of the ethical domain, can offer nothing that a pure cognitivist view positing a contingent connection between ethical judgment and motivation could do.

As Schroeder presents it, if the Open Question Argument (OQA) is to be understood as a special phenomenon about ethical terms, it must be distinguished in some way from standard versions of Frege’s puzzle. Schroeder aims to illustrate what he thinks the difference is by considering the following questions, where ‘wrong’ is stipulated to have the same descriptive content as some ordinary descriptive term ‘K’:

1. Max believes stealing is K, but does Max believe that stealing is wrong?
2. Max believes that stealing is wrong, but does Max believe that stealing is wrong?
3. Max believes that stealing is wrong, but does Max believe that stealing is K?

If the OQA were just an instance of Frege’s puzzle, we would expect question 1 and 3 to be open, and question 2 closed. But hybrid theories can maintain that there is something special to the OQA; namely, that only question 1 is open, and question 2 and 3 are closed.

And concerning moral motivation, Schroeder points out that for hybrid theories to have any advantage over pure (externalist) cognitivist theories, they should want to explain why motivational internalism should turn out to be a conceptual truth.

In order to get these explanatory advantages (assuming the OQA and motivational internalism are genuine phenomena to be explained), Schroeder contends, hybrid theories would have to accept what he calls the “Big Hypothesis” (2009, p. 301):

> **Big Hypothesis**: If ‘P’ is a sentence expressing mental states M_1 . . . M_n, then the descriptive content of ‘S believes that P’ is that S is in each of the mental states M_1 . . . M_n

The idea is that if the Big Hypothesis were true, we would have an explanation of why question 3 (and question 2) are closed, but question 1 is open. Thinking that stealing is wrong *just is* thinking that stealing is K and desiring not to do what is K. So one could not think that stealing is wrong and fail to think it is K – so question 3 is closed. Question 2, again, is clearly closed. And Question 1 is open, since thinking that stealing is wrong goes beyond simply thinking that stealing is K, on the Big Hypothesis.

And if the Big Hypothesis were true, motivational internalism would fall out as a conceptual necessity. Hybrid metaethics would tell us that “stealing is wrong” would express a belief that stealing is K and a desire not to do what is K. So, given the Big Hypothesis, someone who believed that stealing is wrong would be guaranteed to have a belief that stealing is K and a desire not to do what is K, where this belief-desire pair on its own is sufficient to motivate not stealing, other things being equal.

Finally, to conclude the argument, Schroeder aims to point out the controversial nature of the Big Hypothesis. In effect, Schroeder’s idea is to target one of the main analogies hybrid theorists themselves have tried to make in support of their views. Many hybrid theorists, especially Copp and Boisvert, have relied on an analogy between ethical terms and pejoratives to illustrate the hybrid view. They have suggested that ethical terms are like pejoratives in that they at once serve to describe and to express a non-cognitive attitude; so, just as “cheesehead” at once describes someone as being from Wisconsin and expresses a derogatory attitude towards people from Wisconsin, “wrong” supposedly describes something as having a certain natural property, and expresses disapproval of things having that property. What Schroeder aims to show is that this analogy is not helpful for the hybrid theorist, because hybrid theories need to accept the Big Hypothesis about ethical terms, yet pejoratives (the proposed *model* for ethical terms) do not obey the Big Hypothesis. This is because an utterance of “Mark believes that Jerry is a cheesehead” does not attribute a derogatory attitude to *Mark,* but instead expresses the *speaker's* own derogatory attitude towards people from Wisconsin, according to Schroeder.

Schroeder (2014) clarifies that the problem is not simply that of modelling the analysis of ethical terms on that of pejoratives; after all, Schroeder suggests that the conventional implicature generated by “but” *does* follow the Big Hypothesis, in that “Mark believes that

LeBron is big but fast" attributes belief in a contrast between being big and being fast to Mark, and so "but" may be a better model for hybrid theories (2014, p. 268). The deeper problem, it seems, is that of explaining certain moral inferences involving the word 'true'.

The main point is that: one could accept "What Caroline believes is true" without having any particular non-cognitive attitude, and one could accept "Caroline believes stealing is wrong" again without having any particular non-cognitive attitude – after all, if "wrong" works like "but", "Caroline believes stealing is wrong" attributes an attitude to Caroline, rather than expressing the speaker's own attitude. But if one then went on to infer "Therefore, stealing is wrong", one would be committed to having a particular non-cognitive attitude. Yet *where,* asks Schroeder, does this commitment come from?[28] It seems that the commitment could only be generated if 'true' functioned in such a way that asserting "what Caroline believes is true" committed one *not only* to believing whatever Caroline believes, but *also* to having whatever non-cognitive attitudes Caroline has. At the very least, however, it seems quite strange to think that 'true' works this way.[29]

So, in sum: Schroeder (2009) has argued that hybrid theorists must accept the Big Hypothesis to gain an advantage over pure cognitivism. Schroeder (2014) then adds to this that *additionally*, hybrid theorists must endorse an unusual position on the semantic behavior of 'true' to maintain this advantage.

3.3 Towards a new taxonomy

[28] Note that implicit in this question is the assumption that the commitment must be generated by the conventional linguistic rules for the use of some *term* that appears in the inference. We shall soon be in a position to see this assumption is not obligatory.

[29] This paragraph represents my simplifying gloss on Schroeder's complicated argument in "The Truth in Hybrid Semantics" (2014).

In this section, I criticize both Schroeder's taxonomy and his overall argument for skepticism about hybrid theorizing. However, my intention is not to defend any particular hybrid view that Schroeder's arguments target; indeed, these views seem largely to accept the assumptions behind the criticisms, and so by their own lights should accept the restrictions and commitments that are laid down in Schroeder's arguments.

As we shall see, the crucial assumptions to be identified in Schroeder's taxonomy are assumptions rejected by an alternative *neo-expressivist* approach. Thus, identifying these assumptions at once makes clear the possibility of a view obscured by Schroeder's taxonomy, and undermines the force of Schroeder's arguments against such a view. Insofar as Schroeder's arguments, along with the other critical arguments considered so far in this chapter hang on these assumptions, the result is that the ethical neo-expressivist view developed in Bar-On and Chrisman (2009) is better placed than any of the competing views discussed in this chapter.[30]

3.3.1 The Lockean assumption

The central assumption, present throughout Schroeder's corpus, and widely shared in the literature discussing the expression relation between ethical claims and motivational states, is the following 'Lockean' assumption about meaning. (This topic has also been discussed in Chapter 2 under the label 'the mentalist assumption'. The main difference is that the mentalist assumption is explicitly about first-order semantics, whereas the Lockean assumption is broadened to cover certain expressivist *meta*semantic views, such as the one espoused in Ridge, 2014).

[30] The radical departure that ethical neo-expressivism takes from competing metaethical views seems not to be appreciated in the literature; when ethical neo-expressivism is mentioned, it is treated alongside hybrid theories and taken to share in their advantages and drawbacks. And all of this despite the fact that strictly speaking, ethical neo-expressivism is not even a hybrid view, and that it departs from hybrid views in ways that disarm the objections discussed so far in this chapter. I hope to remedy this oversight in this dissertation, in part by introducing the relevant neo-expressivist notions in the course of responding to Schroeder.

> **The Lockean Assumption**: (i) Linguistic devices (including sentences, words, etc.), as used in context, express mental states (such as beliefs, desires, etc.), where (ii) this expression figures (in some way) in the explanation of the meaning of the linguistic device.

I do not mean to ascribe exactly this view to Locke. But the idea captured in this assumption traces back to some of John Locke's remarks on meaning in *An Essay Concerning Human Understanding*. A crucial passage here, worth quoting in full (Book III.i.1, emphasis added):

> God, having designed man for a sociable creature, made him not only with an inclination and under a necessity to have fellowship with those under his own kind, but furnished him also with language, which was to be the great instrument and common tie of society. Man, therefore, had by nature his organs so fashioned, as to be fit to frame articulate sounds, which we call words. But this was not enough to produce language; for parrots, and several other birds, will be taught to make articulate sounds distinct enough, which yet by no means are capable of language. Besides articulate sounds, therefore, it was further necessary that he should be able to use these sounds *as signs of internal conceptions; and to make them stand as marks for the ideas within his own mind, whereby they might be made known to others, and the thoughts of men's minds be conveyed from one to another*

And, later (Book III.ii.1),

> The use, then, of words, is to be sensible marks of ideas; and the ideas they stand for are their proper and immediate signification

Grice (1989), as well, may be credited for the proliferation of the Lockean assumption, insofar as his influential work gives us a picture of language according to which its primary function is the conveyance of thought from one creature to another, through the operation of complex communicative intentions. Ideas roughly of this sort also permeate the work of Gibbard, who remarks: "that *words* express *judgments* will, of course, be accepted by almost anyone" (1990, p. 84, emphasis added), and maintains that expressing a mental state consists in uttering words conventionally designed to get one's audience to think that one is in that state (1990, pp.

85-86). Another influential thinker in this regard is Davis, who defends at great length an ideationalist account of meaning in *Meaning, Expression, and Thought* (2002). Of special significance for our purposes is Davis' notion of *expression*, explicitly adopted by Ridge (2006, p. 303; 2014, pp. 107-111) and Eriksson (2009, pp. 17-18 and fn. 35; 2014, pp. 159-162). The Lockean assumption connecting significant linguistic items to mental states is widespread, especially so in expressivist discussion of the meanings of ethical terms. Indeed, some (especially Schroeder) seem to take it to be *definitional* of expressivism that it contends that the semantic meanings of ethical sentences are given by the distinctive non-cognitive attitudes those sentences are said to express.

But it is not only the pure expressivist, or even the ecumenical expressivist, who accepts the Lockean assumption in thinking about the meanings of ethical sentences. Some ecumenical cognitivists work with a variation of the Lockean idea, too. Copp and Barker, for instance, maintain that it the expression of some motivational state *is* an aspect of the conventional, linguistic meaning of ethical terms, though, of course, this expression does not contribute to the truth-conditions of sentences in which ethical terms appear (in extensional contexts at least). And Boisvert, as we have seen, holds that the 'expressive content' of sentences contributes to their meanings.

The only views so far considered that seem not to be committed to some version of the Lockean assumption, then, are the conversational implicature views of Finlay and Strandberg.[31] This should perhaps be no surprise, as Finlay and Strandberg each describe their views as versions of pure cognitivism that employ *conversational* implicature to explain some aspect of

[31] At least, such commitment is not a *necessary* component of their views. Nevertheless, Strandberg puts the canonical statement of his view in a way strongly suggestive of the Lockean assumption, writing that a moral "sentence expresses, in virtue of its conventional meaning, the belief that phi-ing has a moral property" (Strandberg 2012, p. 101).

the apparent practicality of ethical thought and discourse. Given that conversational implicatures are generated by norms governing conversation, rather than the conventional meanings of the linguistic devices used in conversation, there seems to be no commitment to (ii) of the Lockean thesis for Finlay or Strandberg. (Still, this very feature is also a *drawback* of their views, since it implausibly predicts that we cannot express our motivational states outside the context of a conversation).

To see how the Lockean assumption permeates Schroeder's presentation of the literature on hybrid theories, we need look no further than the questions Schroeder uses to taxonomize the various views. He asks, for instance: "Do different *sentences* containing the word 'wrong' express different *desire-like states*?", and "Does the descriptive content of a sentence depend on the desire-like state *it* [i.e., the sentence] expresses?" (2009, p. 261, emphasis added). And we see the same assumption in the articulation of the Big Hypothesis: "If 'P' is a *sentence* expressing *mental states* $M_1 \ldots M_n$, then the descriptive content of 'S believes that P' is that S is in each of the mental states $M_1 \ldots M_n$" (2009, p. 301, emphasis added).

As has already been emphasized, endorsing a mentalist semantics seemingly deprives the expressivist of all the resources and successes of a truth-conditional semantics that assigns abstract objects having certain logical and compositional properties as meanings. The core advantage of holding that sentences express propositions is that doing so allows us to explain, among other things, (a) how we can make good translations of sentences into other languages, (b) how the meaning of a sentence could be constant in both asserted and unasserted contexts, (c) how we articulate the content of various attitude ascriptions, etc. (see Bar-On, Chrisman, and Sias 2014, pp. 226-231 for discussion). Given this picture, the proposition expressed by an utterance of the sentence "Lying is wrong" is what that sentence contributes to the meaning of, say “If lying is wrong, then getting your brother to lie is wrong”, and is the content of the belief we ascribe to someone when we say that they believe lying is wrong, and is what a good translation of the sentence into another language means, etc.

In sum: The Lockean assumption appears to be at odds with the compelling idea that sentences express propositions, yet the Lockean assumption has also seemed crucial to gaining whatever advantages there are to be had by adopting an expressivist theory of ethical thought and discourse. Thus, expressivists are pressed to come up with an alternative mentalist semantics to do the work of standard truth-conditional semantics, as we saw in Chapter 2.

Before presenting the neo-expressivist alternative, it will be important to consider another assumption going in the opposite direction as the Lockean assumption. This shall be important to consider, because it can appear as though we might be forced to choose between these two assumptions, and so some of the support from the Lockean assumption may derive from a perceived need to avoid this alternative assumption.

> **The naturalist representationalist assumption**: (i) The meaning of a sentence is the proposition it expresses, where (ii) propositions are inherently representational, in that they describe some way that the natural world might be.

On the naturalist representationalist assumption, the meaning of referring terms is the natural-worldly item referred to, and the meaning of a sentence is, ultimately, the possible natural-worldly state of affairs it describes. The naturalistic component of this picture drives a reductionist project, wherein theorizing must be done to identify plausible referents and states of affairs to stand as the meanings for terms and sentences that resist easy naturalistic construals. For instance, if the only objects we admit into our ontology are those recognized by our best science, what in the world could be the referent of concepts such as causality, number, intentionality, and, of course, moral rightness and wrongness? In each case, we have what Price describes as a ‘placement problem’ (see also 2.4.3): “If all reality is ultimately natural reality, how are we to ‘place’ moral facts, mathematical facts, meaning facts, and so on?” (2013, p. 6). More precisely, what we have are placement problems in their *linguistic* guise, or the problem of specifying what a term *means* in terms of what it is *about*, where there are not any obvious naturalistic candidates for certain terms to be about. As Price insightfully points out, the linguistic version of the placement problem trades on a representationalist assumption about language; the assumption “that the *linguistic* items in question ‘stand for’ or ‘represent’ something *non-linguistic* . . . This assumption grounds our shift in focus from the *term* ‘X’ or *concept X*, to its assumed *object*, X” (2013, p. 9).

The naturalist representationalist assumption, like the Lockean assumption, has a long pedigree. One motivation for pure expressivism is the perceived intractability of the moral placement problem. Expressivism dissolves the placement problem by denying the naturalist assumption, at least when it comes to ethical thought and discourse. Some of the central arguments put forth on behalf of pure expressivism, including Moore’s Open Question Argument, Horgan and Timmons’ Moral Twin Earth argument (1991), as well as various arguments centering around motivational internalism and direction of fit, and the degree and depth of actual and possible moral disagreement, each can be understood as aiming to establish that, given the naturalist representationalist assumption, the placement problem in the moral realm is unsolvable.

However, what has not been as well appreciated until more recently is that it is possible to accept (i) of the naturalist representationalist assumption, and thereby gain the advantages of the idea that sentences express propositions, while abandoning (ii). This is especially clear in

Ridge (2014), where Ridge grants (i), accepting that ethical sentences have truth-conditions and express propositions, but he denies (ii) by treating propositions (following Soames) as *cognitive event types*, where the cognitive event type that constitutes the sort of proposition expressed by an ethical claim is a *normative judgment* combining a normative perspective and a representational belief, as per Ridge's ecumenical expressivism. This view treats normative propositions as entities that are not inherently representational, for they only represent some state of affairs in virtue of the normative perspective of the judger (see Ridge 2014, pp. 124-131).[32]

By accepting (i) of the naturalist representationalist assumption, Ridge's ecumenical expressivism makes substantial progress over pure expressivism in handling almost all of the major problems for expressivism. Ridge then combines acceptance of (i) – the idea that the meaning of the sentence is a proposition – with a metasemantic reading of the Lockean assumption. This combination forces him to accept a view like Soames', that sees propositions as mental event types. In order to hold that the meaning of a sentence is the proposition it expresses, *and* that sentences express mental states, where this expression plays a role in the explanation of the sentence's meaning, Ridge *has* to accept that propositions are fundamentally mental entities. And he is indeed clear on this point, noting that his view reverses the traditional order of explanation: thoughts do not derive their content from the propositions they are related to; rather, propositions derive their contents from the thought types that constitute them (Ridge 2014, p. 126).

[32] Chrisman, as well, argues that we can accept (i) but not (ii) of the naturalist assumption, divorcing the *semantic* project of articulating the theorems governing the semantic units of a language, from the *metasemantic* project of explaining "the psychological, sociological, and ontological underpinnings of meaningful use of linguistic signs" (2016, p. 14), including the question of whether "ought-statements mean what they do in virtue of how they describe the world as being, what motivational attitude they express, or the quasi-logical role they play in a particular kind of reasoning" (2016, pp. 14-15).

Ridge's view thus upsets traditional ways of distinguishing cognitivism and expressivism, in that he combines one element of the naturalist representationalist assumption with a metasemantic understanding of the Lockean assumption. In the next section, I discuss the neo-expressivist stance on the two assumptions discussed in this section. Significantly, neo-expressivism also combines certain elements of the Lockean and the representationalist stance, but in a fundamentally different way from Ridge. As we shall see, neo-expressivism accepts (i) but not (ii) of the naturalistic representationalist assumption, but does not undertake commitment to the Lockean assumption, while providing an alternative explanation of the social aspects of expressive behavior that made the Lockean assumption plausible in the first place.

3.3.2 The Neo-Expressivist diagnosis

We can diagnose the apparent impasse between the Lockean and the representationalist assumptions as arising from the fact that the notion of *expression* is being used for two very different purposes. On the one hand, there is reason to think that sentences express propositions, as just discussed. This is needed to capture the formal logical, compositional properties of language. In this sense, 'expresses' is roughly equivalent to 'means that', and is an essentially semantic notion. But on the other hand, the core expressivist notion arises from the intuitive thought that language is essentially *social*, and that the significance of much of what we do with words resides in the fact that we use them to give voice to our mental states. In this sense, 'expression' concerns a relation between minded creatures and the vehicles they use to give voice to their mental states.

Having presented the issues at this level of generality, we are now in a position to appreciate the contribution of *neo-expressivism*.[33] The core insight offered by neo-expressivism is the following distinction between two kinds of expression:

> *S-expression*: Expression in the *semantic* sense. This is a two-place relation between a significant linguistic item and its conventional semantic meaning. This includes how *sentences* express *propositions*.
>
> *A-expression*: Expression in the *act* sense. This is a three-place relation between a person (or a minded creature more generally), a mental state, and an expressive vehicle. In an expressive act, a person S gives voice to mental state M through the use of an expressive vehicle E.

Some further comments on the distinction: Semantic expression is a conventional relation, such that sentences (in context) express propositions through being conventional representations of them. The s-expression relation is fully compatible with standard truth-conditional approaches to the meaning of linguistic items in natural language, and also with other semantic accounts, such as possible world semantics assigning possible worlds as the representational content of sentences. Beyond this, however, neo-expressivism makes few substantive assumptions about the nature of the s-expression relation. For one thing, neo-expressivism is not committed to there being meaning-giving paraphrastic analyses for many items of a language. That is; we can for the most part (ignoring certain contextual parameters) specify the proposition expressed by a sentence disquotationally, so that “snow is white” expresses the proposition that snow is white. (This is not, of course, to endorse a disquotational theory of truth, or to foreclose the possibility of illuminating ‘deep’ semantic analysis in some

[33] The following exposition draws from Bar-On (2004, 2012, 2015); Bar-On and Chrisman (2009); Bar-On, Chrisman, and Sias (2014).

cases). This ‘semantic innocence’ is not specific to the ethical domain, either. As Bar-On, Chrisman, and Sias put it: “A long history of failures in areas other than ethics to give paraphrases of sentences containing simpler terms and analyses of atomic concepts should make us leery of any attempt to paraphrase sentences containing ethical terms in other terms (normative or not)” (2014, p. 228). Thus, neo-expressivism does not endorse (ii) of the naturalist representationalist assumption, as this is the component which requires finding meaning-giving paraphrastic semantic analyses in terms of naturalistically respectable items.

The notion of a-expression is not purely a linguistic phenomenon, and in more ways than one. First, many non-linguistic creatures are equally as capable of a-expression as humans are, and humans of course are also capable of non-linguistic a-expression. We see a spectrum of expressive behavior, running from animal alarm calls, through to groans, hugs, smiles, etc. Not all cases of a-expression require the use of a linguistically articulate expressive vehicle. Second, a-expression is something that can happen in *thought,* as well as out loud. As one thinks to oneself privately “He’s such a *jerk*”, one can be said to express one’s mind, *in* one’s mind. (This already promises an advantage for ethical neo-expressivism over conversational implicature theories, which predict that the expression of a motivational state is dependent on the existence of a conversational context).

Significantly, given the neo-expressivist framework, we can dispense with the Lockean assumption, *while still* explaining the aspects of thought and discourse that give the assumption its superficial plausibility. Any plausibility accruing to the Lockean assumption derives from the social-communicative functions of expressive behavior. And surely, in discourse, we do communicate our mental states to each other. We should not dismiss the idea that in linguistic communication, we do express our beliefs, desires, and so on. Where the Lockean assumption goes wrong is in reading this social-communicative function into the semantic analysis of language generally. That this is a misstep should be clear from the fact that non-linguistic creatures are just as much capable of performing expressive acts as linguistic creatures; this shows that competence with certain formal linguistic rules associated with a language is not necessary for expressing one's mind in the sense of a-expression.

Now, this is not to deny that there may be significant explanatory connections between the notions of a-expression and s-expression; for instance, it might seem that a-expression is prior to s-expression, and may play an important role in the explanation of the evolutionary origins and individual development of a capacity for language (Bar-On, 2019; see also Chapter 5). However, such an explanation does not elide the a-/s-expression distinction but presupposes it. At any rate, even if the notion of a-expression plays an essential role in the explanation of the origins of semantic expression, it would be a mistake to define the semantic contents of significant linguistic expressions in terms of mental states, at least given that the relata of the s-expression relation are different from the relata of the a-expression distinction (for one, s-expression is a two-place relation while a-expression is three-place, holding between a minded creature, mental state, and expressive vehicle).

In the next chapter I shall consider directly the positive neo-expressivist account of ethical thought and discourse that I favor. What I wish to do here is to use the a-/s-expression distinction to uncover and reject the crucial assumptions in Schroeder's taxonomy. Simply put, the complaint is this: *Sentences* don't express mental states; *people* do, using expressive vehicles such as sentences. Thus, in stark contrast to the canonical statement of expressivism, neo-expressivism is not committed to holding that the semantic content of ethical sentences is given by the distinctive non-cognitive mental states those sentences express. Still, neo-expressivism captures the spirit of the expressivist position, in maintaining that in *acts* of making ethical claims using ethical sentences, *speakers* give voice to motivational states.

Once this point is recognized, we must reject Schroeder's taxonomy of hybrid theories, since the conflation of the a-/s- expression distinction is built into the very formulation of his taxonomical questions. Additionally, it remains to be seen to what extent (if any) Schroeder's criticisms against hybrid views (and expressivism in general) have any bearing at all on the ethical neo-expressivist position to be elaborated in the next Chapter.

3.3.3 Problems with Schroeder's taxonomy

Although I reject the letter of Schroeder's taxonomy, his questions do nevertheless track some significant distinctions between existing hybrid theories. Therefore, in developing a new taxonomy, I begin by offering attempted paraphrases of Schroeder's questions, as follows:

> Q1*. Does the desire-like state a person S expresses in making an ethical claim of the form "X is wrong" have, specifically, X as its intentional object, or is it directed towards a more general characteristic?
>
> Q2*. Do competent users of an ethical term express the same attitudes when making ethical claims by uttering sentences containing those terms?
>
> Q3*. Does a sentence of the form "X is wrong" express a proposition that ascribes a different property to X depending on who the speaker/thinker is? (That is, do ethical terms have an indexical character)?
>
> Q4*. Does a sentence of the form "X is wrong" express a proposition that ascribes a different property to X depending on what desire-like state the speaker/thinker is in?

On this way of understanding the questions, it turns out that questions Q1* and Q2* concern the nature of ethical judgment, considered as a mental state, and Q3* and Q4* concern the semantics of ethical sentences, and particularly, whether ethical terms are like indexicals. Our next question is whether, on this reconstrual of the questions, any of Schroeder's arguments have bearing.

Schroeder's argument for a "No" answer to Q1, recall, was that a "No" answer is needed to explain why someone who accepts the premises of a moral argument and comes to accept the conclusion must be committed to having the desire-like mental state 'expressed by' the conclusion. But now we should be able to see that any 'commitment' to being in a certain desire-like state that comes from drawing the inference is not derived from the semantic features of the inference in the first place. Schroeder assumes that such a commitment must come from linguistic conventions governing the use of some terms that appears in the argument. This presupposes that the commitment is an aspect of the s-expressed content of the sentences in the argument. However, it is open to us to say that any commitment to having a desire-like state that comes with making the inference will come from the fact that S, having drawn the relevant inference, then makes an *ethical claim* ('claim' is understood in the *act* sense – a claim*ing*) in affirming the conclusion. That is, it is a feature of the *act* of making an ethical claim that accounts for S's expressing a desire-like state, not the semantic features of the premise or conclusion. So, a "No" answer to Q1* is not obligatory.

Concerning Q2*: the problem Schroeder raised for Q2 was one about the expression-relation; in particular, how S's utterance of an ethical sentence M could possibly inform an audience about S's desire-like state, on the assumption that the relevant desire-like state types vary across individuals (e.g., if S's attitude of disapproval was governed by utilitarian standards, while T's was governed by Kantian standards). In its reformulation as Q2*, the question seems to rest on an issue about how to individuate mental state types, and what elements of one's mental states are revealed in one’s expressive acts. Most simply, we might just say that S, in making the ethical claim that phi-ing is wrong, expresses her motivation not to phi other things being equal. Schroeder seems to assume (perhaps following the views he critiques) that in making a simple ethical claim that phi-ing is wrong, speakers express more than just a motivation not to phi – they express a (possibly idiosyncratic) attitude of disapproval of things insofar as they have some property. But it is unclear why neo-expressivism would have to hold this; after all, the idea that ethical claims express an attitude towards a general characteristic of things seems questionable in the first place, as noted earlier in the discussion of Barker’s view. The idea is driven by a perceived need to explain how making a moral inference could lead to a commitment to having a certain desire-like state in accepting the conclusion (see discussion of Q1* above). Given that neo-expressivism denies that any such commitment is an aspect of the semantic meaning of ethical terms, there is no reason for neo-expressivism to accept the idea that speakers express a (possibly idiosyncratic) attitude towards general characteristics of things, rather than directly expressing a basic motivationally-charged attitude directed at specific acts, policies, persons, etc.

Concerning Q3*: Schroeder contends that Q3 must be answered “No” in order to avoid a certain problem about attitude ascriptions, inference, and the behavior of the truth predicate. But since Q3 concerns the semantic analysis of ethical sentences, there are really no *special* problems here as far as the neo-expressivist framework is concerned, that wouldn’t be problems for an indexicalist semantics for ethical sentences in the first place. Neo-expressivism aims to remain semantically neutral, and so avoids positing a contextual parameter in ethical terms, given that they do not appear to exhibit one on the surface.

And finally, concerning Q4*: Schroeder's argument is that a "yes" answer would mean that most thinkers could not regard their own moral inferences as rational, since in order to do so they would need to realize that the premises and conclusion of their inference 'expressed' the same desire-like state – which they would only realize if they knew that ecumenical expressivism were true. The problem Schroeder sketches here is predicated on the Lockean assumption. It holds that the semantic content of the premises and conclusion have the logical features they do only because of the mental states they are said to express. Only given this assumption does it turn out that realizing that an argument is valid requires knowledge that the premises and conclusion express a certain mental state in the way predicted by ecumenical expressivism. In the neo-expressivist framework, however, the logical features of a valid moral argument are independent of the sorts of expressive acts agents might employ the sentences involved in the premises or conclusion to perform. Accordingly, it is challenging to re-articulate Schroeder's worry in any compelling way in the neo-expressivist framework. And again, we should not require agents to have sophisticated knowledge of the correct semantic analysis of moral sentences for them to coherently make moral inferences; a disquotational specification of the contents of the premises and conclusion should suffice for ordinary reasoners to grasp its validity.

And finally, the Big Hypothesis itself that Schroeder proposes is not even well-formed given the neo-expressivist framework. Here is an attempted re-articulation:

> **Big Hypothesis***: If a person S expresses mental states $M_1 \ldots M_n$ by uttering the sentence 'P', then the sentence 'S believes that P', uttered in a context where S's claim 'P' is salient, semantically expresses the proposition that S is in $M_1 \ldots M_n$

Put this way, however, the Big Hypothesis seems no longer to be a thesis about the semantics of 'P', but rather an observation about how we report on the mental states given voice in expressive acts. Understood this way, the Big Hypothesis* is not very plausible (even less plausible than Schroeder thinks his own version is), and neither is it clear what work it might do in gaining an advantage for hybrid views over pure cognitivism. Regarding its plausibility: Suppose S were to utter the sentence “You will report for duty at 6:00am sharp tomorrow” in issuing a command to U. Given that this is a command, it seems to express S’s desire that U reports for duty at 6:00am exactly the next day, since this is a sincerity condition for issuing such a command. But the sentence ‘S believes that U will report for duty at 6:00am sharp tomorrow” semantically ascribes to S only a *belief*, and not a desire. The underlying problem is that which mental states S expresses in uttering a sentence ‘P’ is partly a matter of what kind of act S uses ‘P’ to perform, and is not strictly a matter of the semantic content of ‘P’, as it would have to be for the Big Hypothesis* to be true.

Regarding the need for the Big Hypothesis* for any hybrid view that claims an advantage over pure cognitivism: In the next chapter, we shall consider a hybrid version of ethical neo-expressivism that has an advantage over pure cognitivism when it comes to accounting for moral motivation. Given that this advantage does not require any assumption like the Big Hypothesis, it is not true (as Schroeder claims) that the Big Hypothesis is a necessary commitment of any hybrid theory claiming an advantage over cognitivism.

The above discussion of the Big Hypothesis* also gives us a clue about how to address Schroeder’s worry about the behavior of ‘true’. For, our puzzle now is no longer how the sentence “What Caroline said is true” could possibly express a non-cognitive attitude, given that Caroline said “stealing is wrong”. Instead, the question is: why is it that S’s act of *asserting* “What Caroline said is true” commits S to having a certain non-cognitive attitude (in addition to a belief), if what Caroline said was “Stealing is wrong”? And this question does not seem as puzzling, at least on the assumption that to assert “What Caroline said is true” is to commit oneself to asserting that P, if what Caroline said was that P, for any proposition P. Our puzzle then becomes the more tractable one of explaining how agents express motivational states in making ethical assertions. This is a main topic for the next chapter.

3.3.4 Conclusion: A new taxonomy

We are now in a position to set out a new approach to taxonomizing the various hybrid theories. My concern is specifically with distinguishing between various hybrid theories in terms of the features that qualify them as hybrid, rather than on various other aspects of the views. Thus, I shall set aside details about the specific nature of the desire-like state, or whether the semantics of ethical sentences is contextualist, etc.

For any given hybrid theory we can ask:

- Qa: Is it constitutive of ethical judgment that it has both a belief and a desire-like component?
- Qb: In order to properly issue an ethical claim/assertion, must one be in both a belief and a desire-like state?
- Qc: What linguistic/psychological mechanism(s) explain the relation between ethical claim/assertion and the relevant belief?
- Qd: What linguistic/psychological mechanism(s) explain the relation between ethical claim/assertion and the relevant desire-like states?
- Qe: Is the content of the relevant belief guaranteed to be true just in case the related ethical claim/assertion is?
- Qf: Are the motivationally-charged attitudes directed at particular acts, policies, persons, etc., or to a general characteristic of things?

	Barker/Copp	Finlay/Strandberg	Boisvert	Ridge/Eriksson
Qa	No	No	?[34]	Yes
Qb	Yes	No (cancellability)	Yes	Yes
Qc	Norm on assertion?	Norm on assertion?[35]	Norm on assertion	Accountability-expression
Qd	Conventional Implicature	Conversational implicature	Norm on assertion	Accountability-expression
Qe	Yes	Yes	Yes	No
Qf	General characteristic	General Characteristic	General Characteristic	General Characteristic/Particular Thing[36]

Fig. 2 Taxonomy of hybrid views

Questions Qa and Qb concern the relation between ethical thought, discourse, and motivation. Answering "Yes" to Qa amounts to endorsing motivational internalism, as conventionally construed. Question Qb concerns another sort of internalist connection, to be discussed further in the next chapter. A "yes" answer to Qb amounts to a kind of internalist position that Copp has described as 'discourse internalism', and which may or may not

[34] Boisvert (2008) does not discuss the nature of ethical judgment in enough detail to see how he would answer this question. And since his discussion is limited to 'correct and literal' ethical claims, it is unclear whether one way to *in*correctly make an ethical claim is to do so while harboring a relevant belief but not the relevant motivational state.
[35] It is somewhat unclear just how Barker, Copp, Finlay, and Strandberg take ethical claims to express beliefs. Each of these theorists assume that the expression of belief here is unproblematic, and exactly parallel to how ordinary descriptive claims express beliefs -- but they do not specify how it is that ordinary claims do this. At some points, it seems that they take ordinary claims to *semantically* express beliefs, thereby endorsing a global mentalist semantics (see e.g. Strandberg, 2012, p. 15). But to be charitable, I will interpret them as holding that beliefs are a-expressed by ethical claims.
[36] This is one of the main points of contrast between Ridge and Eriksson – Eriksson emphasizes that on his view, ethical claims express pro-/con-attitudes towards particular things, rather than towards general characteristics.

accompany ordinary motivational internalism. Qc is an oft-neglected, but I think significant question for hybrid theories to answer. It is often left un- or under-discussed by ecumenical cognitivist views exactly how they hold that ethical claims express *belief*. Naturally, the focus is on the innovative aspects of ecumenical cognitivism (how they explain the connection to motivation), but I think once it is clarified how these theorists think ordinary claims express beliefs, a question arises about why ethical claims do not express motivational states in just the same way, as the hybrid view I shall propose holds.

Qd is significant for some of the reasons already presented in section 3.3. As Schroeder has pointed out, expressivists are not always clear about what expression-relation they have in mind. There are a number of proposals that have been considered – including Schroeder's 'assertibility-expressivism', Davis's 'accountability-expressivism', and ecumenical cognitivists employ conventional or conversational implicature to do the job. The neo-expressivist framework presents an alternative to these accounts of the expression-relation, pointing out that we are dealing with not one, but *two* expression-relations. Additionally, the neo-expressivist framework is at least in one respect a more genuine representative of the original expressivist idea, insofar as it recognizes that in acts of expression, we directly give voice to our mental states, rather than *indicating*, or *conveying that* we are in those mental states. Even supposedly 'expressivist' views end up relying on conceptions of the expression-relation that smuggle in the *subjectivist* idea that ethical claims indicate *that* one is in a certain motivational state, rather than simply expressing that state. The result of this, as we shall see in the next chapter, is that the neo-expressivist framework is better placed than competitors to accommodate Stevenson's idea of disagreement in *attitude*.

Qe distinguishes between ecumenical cognitivism and ecumenical expressivism. And Qf concerns the nature of the desire-like attitudes the various theories take ethical claims to express. As discussed above, many hybrid theories endorse the idea that agents express attitudes towards a general characteristic of things in order to address the perceived need to explain how going through a moral inference can commit one to having a certain desire-like attitude involved in accepting the conclusion. However, given the neo-expressivist picture, this is not necessary, for any such commitment is not a semantic or logical feature of moral arguments in the first place, but is a commitment generated by the act of affirming the conclusion of the moral argument as an ethical claim. (I discuss the nature of moral inference on the neo-expressivist account in the next chapter). I take this to be an advantage of the neo-expressivist view, since the idea that speakers primarily express attitudes towards general characteristics in making ethical claims, rather than attitudes towards the specific objects of those claims, seems unmotivated independently of the perceived problem concerning moral inference above.

I turn in the next chapter to explain the positive ethical neo-expresivist view and its advantages over the hybrid theories so far considered, and then to develop a *hybrid* version of ethical neo-expressivism.

Chapter 4

A Hybrid Theory of Ethical Discourse

4.1 Introduction

The previous chapter summarized prominent hybrid theories of ethical thought and discourse, discussed Schroeder's taxonomy of and challenge to these theories, and then introduced a neo-expressivist approach not easily categorizable in this taxonomy. In this chapter, I begin (in Section 4.2) by considering a direct application of the neo-expressivist framework to ethical thought and discourse, developed in Bar-On and Chrisman (2009) and Bar-On, Chrisman, and Sias (2014).

Although ethical neo-expressivism may superficially appear to be a hybrid theory, officially, the view is *not* hybrid, because it is only committed to holding that ethical judgments are constituted by non-cognitive states. In this chapter, I argue (in Section 4.3) that there is good reason to prefer an explicitly hybrid version of ethical neo-expressivism. This hybrid ethical neo-expressivism shares in all the advantages of its non-hybrid counterpart and additionally provides an adequate account of the epistemic continuities between ethics and other domains. Moreover, it (like non-hybrid ethical neo-expressivism) is a fairly minimal theory, in that it is able to remain neutral on a surprising number of metaethical questions, such as questions about the truthmakers for moral truths, or the correct semantic analysis of ethical sentences.

Starting in this chapter and continuing into Chapter 5, I also develop hybrid ethical neo-expressivism beyond its core commitments to provide a more robust metaethical theory. In the second half of this chapter (4.5-4.10) I extend hybrid ethical neo-expressivism to give an account of ethical *assertion*, understood as a speech act, and in the next chapter I employ Millikan's biosemantic framework (which I see as compatible with and complementary to the neo-expressivist framework) to develop a substantive account of the proper function of ethical judgment that explains and vindicates the surface features of the ethical domain as presented in Chapter 1.

4.2 Ethical Neo-Expressivism

Ethical neo-expressivism, as articulated in Bar-On and Chrisman (2009) and Bar-On, Chrisman, and Sias (2014), represents an expressivist theory that departs from traditional expressivism in radical ways. Yet this radical divergence does not yet seem to have been fully appreciated in the literature. When ethical neo-expressivism is cited, it primarily seems to be in the context of discussion about hybrid metaethical theories, notwithstanding the point that ethical neo-expressivism is, again, officially *not* a hybrid view (see Bar-On, Chrisman, and Sias 2014).[1] Emphasizing the ways in which ethical neo-expressivism diverges from more traditional ways of thinking of expressivism in metaethics, I hope, will remedy this oversight and also reveal the attractions of the view.

As discussed in Chapter 3, and reproduced here, a central, defining idea of *neo-expressivism* is that we must take care not to conflate two distinct kinds of 'expression'. Following Sellars (1969), the neo-expressivist marks a distinction between expression in the *semantic* sense, and expression in the *action* sense (Bar-On 2004, p. 216):[2]

> *S-expression*: Expression in the *semantic* sense. This is a two-place relation between a significant linguistic item and its conventional semantic meaning. This includes how *sentences* express *propositions*.

[1] Schroeder (2009), Strandberg (2012), Fletcher (2014), Finlay (2010), and Toppinen (2017), for example, each discuss ethical neo-expressivism in the context of hybrid metaethics.

[2] Neo-expressivism also acknowledges a third kind of expression, as well: *Causal expression* (c-expression): "an *utterance* or piece of behavior expresses an underlying state by being the culmination of a causal process beginning with that state" (Bar-On 2004, p. 216). Bar-On does not discuss causal expression as much as act and semantic expression, which are more relevant to her concerns in ethical and avowal expressivism.

A-expression: Expression in the *act* sense. This is a three-place relation between a person (or a minded creature more generally), a mental state, and an expressive vehicle. In an expressive act, a person S gives voice to mental state M through the use of an expressive vehicle E.[3]

Along with this s-/a-expression distinction, ethical neo-expressivism also makes a distinction between expressive *acts* and the *products* of those acts (Bar-On 2004, p. 251). An expressive act is a bit of intentional behavior, an event with certain causes and effects. The *product* of a linguistically articulate expressive act, e.g., a sentence, is a token of a type that has certain semantic and logical features.

It is important to note that the same type of expressive act can be performed using different expressive vehicles. One can express one's pain, for instance, by crying, exclaiming "Ouch!", or by wincing, groaning, uttering "I am in pain", or "This hurts", and so on. In each instance, one directly gives voice to one's mental state, allowing others to recognize one's pain.[4] Each of the expressive vehicles with which one might give voice to one's pain would be the products of the relevant expressive act. Considered as products, these vehicles have various features, and some of them may have semantic and logical features. Cries, winces, groans, and the like, of course do not *semantically* express anything. But sentences (in context), such as "I am in pain", or "This is painful", *do* semantically express propositions – namely, the proposition

[3] Note that while expressive acts are intentional on Bar-On's account, they do not require having the intention *to express one's mental state*. They are 'intentional' in the sense of being (to some extent) under the voluntary control of the agent, as opposed to being strictly 'forced out' of one, in the way that an experience of embarrassment can manifest itself in one's blushing face.

[4] Note that on Bar-On's view, a-expression is a factive notion, since one cannot express one's mental state M unless one is in M. The factivity of a-expression separates Bar-On's view from other *evidentialist* accounts of expression discussed in the metaethical literature (e.g. Gibbard 1990, pp. 84-86). Still, Bar-On's account makes room for expressive *failures* – as cases where someone expresses *a* mental state, but not *her* mental state (see, *inter alia*, Bar-On 2004, pp. 280-281). These can be explained as cases in which one uses – infelicitously (albeit unintentionally) – an expressive vehicle designed to express a *different* state from the state one is in. As noted in Chapter 2 fn. 11, this feature of Bar-On's view may be helpful for traditional expressivists to adopt as well.

that the speaker/thinker is in pain, or that some demonstratively identified object is painful, respectively – in virtue of being conventional linguistic representations of those propositions.

Neo-expressivism aims to remain neutral on many questions about propositional content and semantic expression. I think the a-/s-expression distinction itself helps enable neo-expressivism to remain neutral on these issues. In Chapter 2, I traced the difficulties traditional expressivism has in accounting for the semantic continuities to the traditional expressivist combination of the mentalist assumption and a non-cognitive theory of ethical judgment. This combination has the effect of obscuring the a-/s-expression distinction. Given the a-/s-expression distinction, this combination seems to commit a category mistake, taking one of the relata of the a-expression relation (mental states) and connecting it to one of the relata of the s-expression relation (semantic content). But once the a-/s-expression distinction is made clear, we can divide theoretical labor, separating the semantic analysis of ethical sentences from the analysis of how we express our ethical judgments and what their nature is. With these issues separated, neo-expressivists can remain relatively neutral on the question of the precise semantic contents of particular ethical sentences, and they are not committed to providing a mentalist (meta)semantic alternative to truth-conditional and possible-worlds semantics. It is in principle open to the neo-expressivist to accept any of a variety of general semantic theories, and any of a variety of semantic analysis of ethical sentences.

Propositional content, we can say, is whatever (i) gets preserved in a good translation from one language into another, (ii) explains the fact that declarative sentence tokens have the same content in various contexts, such as when asserted, when embedded in a conditional, etc., and (iii) articulates the contents of various attitude ascriptions (Bar-On, Chrisman, and Sias, 2014, pp. 226-227). But beyond this, ethical neo-expressivism makes no substantive assumptions about the nature of propositions in general, or the semantics of ethical sentences in particular.[5] So, for instance, the view proposes no substantive paraphrastic analysis of ethical sentences, as in: "'Torture is wrong' semantically expresses the proposition that torture fails to maximize utility".

Instead, at least when it comes to atomic ethical sentences, neo-expressivism specifies the propositions semantically expressed disquotationally (after taking into account obvious contextual parameters, such as indexicals). This is of course not to endorse a disquotationalist account of truth, either. The idea is just that there is nothing in the surface syntactic features of "Torture is wrong" that requires specifying its content as anything other than the proposition that torture is wrong. Neither does the neo-expressivist view make any particular commitments concerning moral realism. Although ethical sentences are said to express propositions, it is a further theoretical commitment to think that propositions are always intrinsically representational, such that there must be objective worldly ethical facts and properties, if any ethical propositions are to be true. To think that would be to endorse the naturalist representationalist assumption identified in Chapter 3. Whether that representationalist idea about propositions is true is a matter for further meta-semantic theorizing.[6] To illustrate the neutrality here: Ethical neo-expressivism is strictly speaking compatible with moral realism and with a realist construal of the truth-conditions of ethical sentences. A neo-expressivist *could* (but does not have to) endorse the naturalist representationalist assumption, provide a reductive

[5] Schroeder (2013, p. 410) identifies two roles for propositions and considers the benefits to noncognitivists of distinguishing between two corresponding kinds of proposition. In one sense, propositions are whatever constitute "the objects of attitudes like belief, desire, and assertion, and the bearers of truth and falsity". But in another sense, propositions "play a role in carving up the world at its joints, are associated with metaphysical commitment, and are the appropriate objects of excluded middle". Neo-expressivism is interested in the first role for propositions Schroeder identifies.

[6] Indeed, this is a point that Michael Ridge (2014) capitalizes on in articulating a hybrid version of quasi-realist expressivism that he contends is compatible with truth-conditional semantics and with thinking that ethical sentences have propositional content.

analysis of the contents of ethical terms in terms of naturalistic properties (this would be an account of what ethical sentences s-express), and still think that in acts of making ethical claims, speakers a-express motivationally-charged mental states.

Ethical neo-expressivism holds that ethical sentences, like any other declarative sentence, semantically express propositions whose content we can generally specify disquotationally. So, "Torture is wrong" we can say semantically expresses the proposition that torture is wrong. Ethical sentences are, of course, apt expressive vehicles for making ethical claims (considered as acts), in the same way that linguistically articulate psychological self-ascriptions ("I'm annoyed") are apt expressive vehicles for making avowals. The ethical neo-expressivist captures one of the core ideas of traditional expressivism – that a central function of ethical claims is to express non-cognitive states – in terms of a-expression, rather than s-expression. The idea is that in making an ethical claim, an agent gives voice to a certain non-cognitive attitude, using an ethical sentence as an expressive vehicle. It is for this reason that ethical neo-expressivism can be understood to avoid the mentalist and Lockean assumptions described in Chapters 2 and 3. For, the account rejects the idea that the literal semantic *meaning* of an ethical sentence is given by the attitudes of the agent who uses it. This, to my mind, firmly separates ethical neo-expressivism from just about any other expressivist view, and indeed even from any cognitivist views that also accept the Lockean assumption.

One way to diagnose the shortcomings of traditional expressivism is that it in effect attempts to invoke *one* expression relation for the purpose of explaining at once (i) how speakers give voice to distinctive non-cognitive attitudes and (ii) how ethical sentences have certain semantic and logical continuities with other ordinary sentences. Ethical neo-expressivism recognizes that (i) and (ii) require separate treatments– the former in terms of a-expression, and the latter in terms of s-expression.

Ethical neo-expressivism, then, simply avoids standard versions of the Frege-Geach problem. This should not be surprising, since the neo-expressivist view that ethical sentences semantically express propositions is motivated by the very sorts of considerations that drive the Frege-Geach problem in the first place. “Lying is wrong” has the same meaning in asserted and unasserted contexts because in each context, the sentence s-expresses the same proposition. The same goes for embedding in conditionals, and so on. Whatever propositions are, they are at least truth-evaluable and make systematic contributions to the truth-conditions of logically complex constructions in which they appear. This allows us to explain the validity of good moral arguments in just the way a standard cognitivist would. All this, while maintaining that ethical claims (considered as acts) express non-cognitive states, and that such states are what constitute ethical judgments.

To my mind, the plausibility of the a-/s-expression distinction itself, and the ease with which this distinction makes available an expressivist explanation of moral motivation that completely avoids standard Frege-Geach worries, make ethical neo-expressivism an attractive approach to accounting for the features of ethical thought and discourse, as laid out in Chapter 1. The various logico-semantic continuities between ethics and other domains, emphasized in supporting cognitivist treatments of the ethical domain, are accounted for by the fact that ethical sentences express propositions in the very same way that any ordinary declarative sentence does. And the distinctive practicality of ethics (including the connection between ethical claims and motivation, and the prescriptive, action-guiding nature of ethical claims [more on the latter in

Chapter 5]) is explained by the fact that, in making ethical claims, speakers a-express certain motivationally-charged states.[7]

We have just seen how ethical neo-expressivism departs from traditional expressivism and from various versions of ecumenical expressivism that still endorse a version of the Lockean assumption, either as a semantic or a metasemantic thesis. Given that ethical neo-expressivism holds that ethical sentences semantically express propositions, some have taken the view to be similar in substantial respects to ecumenical *cognitivism*. However, it would be a mistake to think of ethical neo-expressivism as an ecumenical cognitivist view, because it does not endorse representationalist assumptions about ethical propositions, nor does it maintain that ethical claims express representational beliefs. Ethical neo-expressivism does not easily fit into either expressivism or cognitivism on standard ways of understanding these approaches. Yet it is not officially hybrid theory, either. For this reason it is difficult to accurately place ethical neo-expressivism in standard taxonomies of metaethical options (see, for instance the difficulties locating it in Schroeder's taxonomy of hybrid views discussed in Chapter 3). Below, I contrast ethical neo-expressivism with what may seem to be the most similar position in the literature; implicature-based versions of ecumenical cognitivism.

Neo-Expressivism vs. Implicature-based Ecumenical Cognitivism

As discussed in the previous chapter, several ecumenical cognitivists – Barker, Copp, Finlay, and Strandberg – employ Grice's notion of *implicature* to explain how we can express certain motivational states in making ethical claims. Let us call such views instances of

[7] Indeed, ethical neo-expressivism maintains that it is essential to an act's being an act of making an ethical claim that one expresses a non-cognitive state in making it. In taking on this commitment, ethical neo-expressivism endorses motivational internalism. However, ethical neo-expressivism also allows for failure of motivation; when this occurs, one can be said to express *a* motivational state, though not *one's own* motivational state, since one is not *in* that state. This reconciles internalist intuitions with the possibility of akratic or amoralist individuals. More on this in 4.9.

Implicature-Based Ecumenical Cognitivism (IEC) We can now clarify both how ethical neo-expressivism differs from IEC and identify a further problem for IEC.[8]

The difference between ethical neo-expressivism and IEC can be found in the contrasts between a-expression and implicature. Let us consider what the relation between a mental state M and speaker U is, on the implicature approach. Implicature content is propositional, so that when U's utterance S carries an implicature, it is always an implicature *that p*.[9] So, *unless* IEC defines propositional content in terms of mental states, it cannot be that the implicature content of U's moral utterance is *itself* U's mental state M. And of course, IEC should not want to define propositional content in terms of mental states anyway, on pain of endorsing an expressivist mentalist semantics these views want to avoid. So, what could the implicature content of moral utterances be, such that IEC can explain a connection between ethical claims and motivation? The only serious candidate, it seems, is that the implicature content is the proposition *that U is in M*, where M is a certain motivationally-charged state. IEC thus has to hold that when U utters a moral sentence S, S carries the implicated propositional content that U is in M. (Recall that Barker explicitly presents his view in this way – he holds that expressing a mental state amounts to *conveying that* one is in it.)

This point makes IEC more closely aligned with *subjectivism* than with expressivism, on one way of understanding that contrast (see 2.3). The familiar difference between simple subjectivism and expressivism is that to *express* a mental state is not the same as *saying that one is in it*. Pure expressivists maintain (or should maintain if they are to merit the label 'expressivist') that ethical claims directly *express* motivationally-charged non-cognitive attitudes, whereas simple subjectivism maintains that ethical claims assert *that* the speaker is in

[8] The following point is originally presented in Bar-On and Chrisman (2009); Bar-On, Chrisman, and Sias (2014).
[9] See Chapter 2.3, on implicature accounts of the expression relation.

such an attitude. Now, of course, the implicature content of an utterance does not give its truth-condition. So IEC is no simple subjectivism. But still, IEC is committed to holding that an ethical claim conveys the proposition that the speaker has a non-cognitive attitude – but to convey that one has a non-cognitive attitude is not the same as expressing that attitude. For this reason, we should perhaps call implicature ecumenical cognitivism *neo-subjectivism*, as suggested by Bar-On and Chrisman (2009, pp. 150-158). By contrast, according to ethical neo-expressivism, agents do not convey *that* they have a motivational state in making ethical claims; they directly express their motivational state itself.

Why should we prefer ethical neo-expressivism to IEC? We have already seen some of the difficulties for IEC in Chapter 3; one reason to prefer ethical neo-expressivism is that it accomplishes much of the same work as these views, while avoiding the problems generated by the implicature model. Another reason has to do with the contrast between expressivism and subjectivism. Simple subjectivism, we have seen, faces the problem of lost disagreement, where expressivism does not. Now IEC is no simple subjectivism, and IEC can be seen to have a simple response to the problem of lost disagreement. Suppose S utters “Torture is wrong”, and U utters “Torture is not wrong”. S’s assertion, according to IEC, carries the implicature that S has some non-cognitive attitude – let’s say, disapproval of torture. U’s utterance would carry the implicature that U does not disapprove of torture. Of course, there is no disagreement here, since these implicature contents are consistent with each other. But IEC can say that S and U still disagree, because apart from the implicature content of their utterances, their utterances also have standard (non-implicature) contents that *are* inconsistent. Because IEC takes a cognitivist approach when it comes to the literal meanings of ethical claims, IEC can capture moral disagreement in just the same way as standard cognitivists would. (Such a response is at least an option for versions of IEC that, like Copp's, hold that ethical terms do not have speaker-relative denotations).

A variation on the problem of disagreement re-emerges for IEC, however a problem of lost disagreement in *attitude*. On a simple account of disagreement, disagreement occurs when two logically incompatible propositions are claimed to be true – call this *logical* disagreement. Now, Stevenson (1937) introduced the notion of disagreement in attitude to explain how moral disagreement is possible given that his non-cognitivist approach cannot explain moral disagreement as logical disagreement. It seems to me, however, that the notion of disagreement in attitude is an explanatorily useful one even independently of its contribution to an expressivist account of moral disagreement. We can consider a variation on Stevenson's original example (1937, p. 27):

Alex: "Let's go to see a movie tonight"

Bill: "No, I don't care for any movies out now. Let's go see the symphony play".

As Stevenson notes, "[t]his is disagreement in a perfectly conventional sense" (*ibid.*). But it does not seem well captured by logical disagreement. At least, it is not obvious what propositions Alex and Bill might be asserting that are logically inconsistent. Rather than saying that Alex believes some proposition to be true which Bill does not here, it seems more natural to say that Alex and Bill disagree in another sense: they take up *practically* inconsistent attitudes towards the prospect of going to see a movie. Alex and Bill, each committed to spending their evening together, prefer to do different things, and they cannot do both of them together at the same time. In voicing their disagreement as in the dialogue above, Alex and Bill each attempt to influence the other to take up their own attitude. I would suggest, with Stevenson, that this sort of disagreement intuitively also occurs in ethical discourse. Consider again the following case, where Jeremy and Immanuel make the ethical claims:

Jeremy: “Torture is wrong”

Immanuel: “Torture is not wrong”

Set aside the question of whether Jeremy and Immanuel are in a logical disagreement for the moment. Intuitively, they do at least disagree in attitude. I would suggest that this is related to the action-guiding feature of ethical thought and discourse – ethical claims are supposed to have some bearing on the actions of oneself and others (1.3.2). An explanation and vindication of this feature will come in Chapter 5, but granting at least that the action-guiding feature of ethical thought and discourse is not merely apparent, appealing to this feature can help to explain part of what is at stake in ethical disagreement. Jeremy and Immanuel, in voicing their disagreement, can be naturally understood as each trying to effect a change in the others’ motivationally-charged attitudes towards torture, so as to alter their actions or dispositions towards acting in certain ways. Jeremy does not (just) want Immanuel to verbally concede that torture is wrong; Jeremy (also) would expect Immanuel not to torture nor be disposed to do so, and perhaps also to have a certain negative affective reaction to the prospect of torturing, and so on. At the same time, it may be that Jeremy and Immanuel *also* are in a logical disagreement, as cognitivism predicts. But it seems to me that something important would be lost in the analysis of moral disagreement if we *only* recognized logical disagreement and did not leave room also for disagreement in attitude. In short: The notion of disagreement in attitude is not just a post-hoc move by emotivists; it names a genuine phenomena in moral discourse that is distinct from logical disagreement.

IEC can explain any logical disagreements involved in the ethical case, in terms of the propositional contents of ethical claims voiced in the disagreement. But it does not explain the disagreement in attitude. This is because the expressivist-leaning concession of IEC is that ethical claims carry the implicature *that* the speaker is in some motivationally-charged state. So Jeremy's utterance ("Torture is wrong") carries the implicature that Jeremy (say) disapproves of torture, and Immanuel's utterance carries the implicature that Immanuel does not disapprove of torture. But there is no disagreement, either in attitude or logically, between these implicature contents. And while IEC does predict logical disagreement at the level of the *asserted* (not implicature) contents of Jeremy and Immanuel's claims, it does not explain the disagreement in *attitude* at that level either.

By contrast, on the neo-expressivist account, agents directly express their mental states, rather than conveying *that* they are in those states. So neo-expressivism can explain the disagreement in attitude involved in moral disagreements; it is the states expressed that conflict, not self-ascriptions of those states. And since neo-expressivism also holds that ethical claims semantically express propositions, neo-expressivism can explain the sense in which moral disagreements are also logical disagreements. This is one advantage of ethical neo-expressivism over IEC.

4.3 Ethical Neo-Expressivism Meets the Wishful Thinking Problem

As discussed above, ethical neo-expressivism is ideally placed as an expressivist view to avoid the Frege-Geach problem altogether. However, there is another simple and powerful objection to expressivist theories, independent of the Frege-Geach problem, which takes on a specifically epistemic character. This is the wishful thinking problem, originally presented by Dorr (2002). Although ethical neo-expressivism goes farther than competing theores, I shall argue that the official ethical neo-expressivist view does not adequately address the Wishful Thinking problem. Fortunately, only a simple modification to ethical neo-expressivism is needed to address the problem. I argue that ethical neo-expressivism should 'go hybrid' by holding that an ethical claim does not simply a-express a non-cognitive state, but also a cognitive one.

Dorr's wishful thinking problem does not target the semantic or logical features of ethical sentences as construed by traditional expressivists. Instead, it targets the traditional expressivist's claim that ethical judgments are constituted only by non-cognitive, desire-like mental states. The wishful thinking problem concerns what expressivist can say about the rationality of drawing inferences using arguments with a moral premise and a non-moral conclusion.

Consider Dorr's example (2002, p. 97):

1. If lying is wrong, then the souls of liars will be punished in the afterlife.
2. Lying is wrong.
3. Therefore, the souls of liars will be punished in the afterlife.

This is clearly a valid modus ponens argument. Let us grant that expressivists are able to explain why the argument is logically valid. But there is a further, general feature of valid arguments that Dorr thinks expressivism cannot account for: namely, that for valid arguments, it is generally possible for reasoners to rationally come to accept the conclusion by accepting each of the premises and reasoning from them to the conclusion.[10] Of course, this isn't always the case: sometimes, one can validly reason to a clearly absurd conclusion it is rational to reject; when this is so, it is rational to reject one of the premises leading to that conclusion rather than accepting the conclusion. But clearly not all arguments with moral premises and non-moral

[10] This is roughly what Ridge and Schroeder describe as the inference-licensing of good moral arguments (see discussion in 2.5.2).

conclusions are like this. Sometimes, it is rationally permissible to accept a non-moral conclusion from an argument containing a moral premise.

The problem of wishful thinking is that expressivists are apparently unable to explain how it could be rational for someone to accept (1) and (2) and reason from them to accept (3). For coming to accept (2), according to traditional expressivism, means coming to have a non-cognitive attitude of (say) disapproval towards lying, rather than coming to have a belief. In general, it is not rational to change one's beliefs in response to a change only in one's purely *non-cognitive* attitudes. That would be like engaging in wishful thinking, where one's beliefs are irrationally responsive to one's wishes and desires, rather than to one's evidence. In general, a change in one's wishes or desires by itself provides no evidence relevant to one's beliefs.

How well does ethical neo-expressivism fare when it comes to the wishful thinking problem? Initially, things look promising. Recall the distinction between expressive acts and expressive products. As it turns out, many of the terms used to discuss cognitivism and non-cognitivism in ethics – such as 'assertion', 'claim', 'judgment', 'belief', 'desire', and so on – all exhibit an act/product ambiguity: there is an ambiguity between the act of assert*ing* and what gets assert*ed*, and likewise between claim*ing*/claim*ed*, judg*ing*/judg*ed*, believ*ing*/believ*ed*, desir*ing*/desir*ed*. Inference too exhibits the act/product ambiguity: an ambiguity between inferr*ing* – a mental process or act wherein one transitions from accept*ing* the premises to accept*ing* the conclusion, on the one hand, and an inference considered as product of the mental act of inferring – a logical relationship that holds between a set of propositions, on the other. So, an ethical neo-expressivist might say, the rationality of the inference from (1) and (2) to (3) is a matter of the logical relationships between the propositions s-expressed by the premises and conclusion, and that this suffices to account for the logical rationality of the inference.

While I think the neo-expressivist response just canvassed is correct as far as it goes, it still seems to me that there is a puzzle to be answered. Dorr's original challenge can be reformulated. After recognizing that arguments containing moral premises and non-moral conclusions can be valid in virtue of the logical relationship between the propositions s-expressed by the premises and conclusion, we should still wonder how to make sense of the rationality of inferr*ing* (considered as mental act) (3) from (1) and (2). For noting the act/product distinction as it applies to inference does not remove the need to make sense of inference when considered as a mental act.

The problem restated, is this: supposing that ethical judgments are constituted solely by non-cognitive attitudes (as per non-hybrid ethical neo-expressivism), and supposing that to accept a moral premise such as (2) in the process of making an inference amounts to forming an ethical judgment, how could it be rational to come to accept – and so believe – a non-moral conclusion such as (3)? Behind this problem is the compelling idea that it can only be rational to adjust one's doxastic states in light of changes in one's evidence and other doxastic states – not in light of a change in one's non-cognitive attitudes alone (that would be like wishful thinking). So, if accepting (1) and (2) make it rational to believe (3), this can only be if accepting (1) and (2) consist in taking up a belief in the contents of (1) and (2). Additionally, it seems that for one to rationally draw the inference from (1) and (2) to (3), one must take the truth of (1) and (2) to rationally support believing (3). (This is in accordance with a 'taking' condition on inference requiring that one reflectively views the conclusions one draws in an inference as supported by the premises of that inference. See Boghossian, 2014). One could not rationally take one's acceptance of (1) and (2) to make it rational to accept (3) unless one's acceptance of (1) and (2) involved having a doxastic state of belief towards the propositions semantically expressed by those premises.

Before presenting a simple modification to ethical neo-expressivism that can handle this problem, let us examine a possible response that does not require any modification. In restating the problem, I assumed that to accept (2) in the process of making an inference would amount to forming a non-cognitive attitude on the ethical neo-expressivist view. But ethical neo-expressivism can resist this idea. It can be maintained that accepting (2) *in making the inference* consists in forming a belief with the content of the proposition s-expressed. That is, when one accepts (2) in carrying out the inference, one is *not* making an *ethical judgment*. Instead, one forms a perfectly ordinary, non-ethical belief, that happens to have ethical content.

This distinction between an ordinary belief with ethical content and an ethical belief initially strikes as obscure, but I think there is a subtle and important issue to navigate here. To elaborate; the idea would be that, in keeping with the neo-expressivist a-/s-expression distinction, we cannot always read off the sort of mental state a person expresses in uttering a sentence S from the propositional content of S alone. What mental state the agent a-expresses is in addition a matter of the kind of expressive act performed. As Bar-On argues in the case of avowals, there is intuitively an expressive difference between uttering “I am angry at my father” just after I have gotten out of a heated argument with my father, and uttering “I am angry at my father” at the end of an extensive therapy session. In the first case, I directly a-express my anger, rather than reporting on it on the basis of some evidence, which is what I do in the second case. The first case, but not the second, is an act of *avowing*. Each utterance is about my anger. But only the first properly expresses my anger. The second only reports on it. What makes something an act of avowing is not the use simply of a linguistically articulate psychological self-ascription;

additionally, one must *speak from* one's mental state and not just report on it. Perhaps a similar strategy can taken for the ethical case. In uttering (or thinking to myself) "Lying is wrong" in the course of drawing an inference, I might not be making an ethical judgment, but simply forming a belief with the content that lying is wrong. Whereas if I utter "Lying is wrong" after catching my friend lying to me, I would intuitively be expressing something like disapproval of lying, and so making an ethical judgment.[11]

Not all linguistically articulate psychological self-ascriptions are avowals in Bar-On's sense, as the anger example reveals. Perhaps not all claims of the form "X is F" where F is an ethical predicate ('wrong', 'good', etc.) are ethical claims, either. These claims may have an ethical issue as their subject matter, but to get to count as ethical claims proper, the ethical neo-expressivist might contend, they must be used in the course of an expressive act that would give voice to a motivationally-charged non-cognitive state. Given this, it seems possible as well that one could claim (or affirm in thought) "Lying is wrong" and not thereby express a non-cognitive attitude – instead, one would express a belief, in the process of drawing an inference. As just discussed, whether an utterance of an ethical sentence counts as an act of making an ethical claim, and so expressing a non-cognitive state, depends on features of the *act*, rather than the semantic content of the claim alone (as product). In the context of acts of drawing inferences, tokenings of ethical sentences will not be ethical judgments, but instead express doxastic states, on this line of response.

Ethical neo-expressivism thus seems to have the resources to handle the wishful thinking problem, while still maintaining the expressivist idea that, in making ethical claims, agents give

[11] Note that this is all compatible with motivational internalism, a thesis Bar-On, Chrisman, and Sias accept. Since, on the present proposal, forming the relevant moral belief in the process of making an inference is *not* to make an ethical judgment, the internalist thesis is not violated.

voice to ethical judgments that are constituted by non-cognitive states. However, I think there remains a further consideration, loosely related to the wishful thinking problem, that can motivate the idea that ethical judgments must be in part constituted by doxastic states. This concerns the relationship between the non-cognitive state a-expressed by S in making the ethical claim that lying is wrong, and the doxastic attitude S forms towards the proposition that lying is wrong in the process of drawing an inference involving that proposition. Presumably, S's mental states in these cases will not be unrelated to each other. If S were to accept that lying is wrong in drawing the inference, we would expect S *also* to have formed the ethical judgment that lying is wrong and so to also have a certain non-cognitive state, other things being equal. And likewise, if S were to come to rationally *reject* believing that lying is wrong, we would again expect S not to hold the moral judgment that lying is wrong, and so we would not expect S to be at least somewhat motivated to avoid lying, other things equal. This suggests, I think, that ethical judgments, even if partly constituted by non-cognitive attitudes, are also responsive to evidence, reasons, and the like, in the same way that ordinary doxastic states are. The best explanation of the responsiveness of ethical judgments to evidence and rational considerations is that such judgments are partly constituted by doxastic attitudes.

Now, officially, ethical neo-expressivism remains neutral on the exact nature of the mental states involved in ethical judgment. The view's primary positive commitment is just that ethical judgments are constituted at least in part by non-cognitive states; the view is not also committed to holding that ethical judgments are *not* constituted by beliefs. Still, as Bar-On, Chrisman, and Sias note, it is at least *open* to ethical neo-expressivism to say that ethical judgments are also partly constituted by belief states. Nothing in the view prevents this. Considerations of simplicity, and possibly a concern to remain consistent with a Humean separation of beliefs and desires, speak in favor of taking ethical judgments to only consist of non-cognitive attitudes (Bar-On, Chrisman, and Sias 2014, pp. 244-245). But if the argument in the previous paragraph succeeds, ethical neo-expressivists can no longer remain neutral on this question and should instead hold that ethical judgments consist of both doxastic and non-cognitive states. That is, we have reason to 'go hybrid' with ethical neo-expressivism. Doing so results in the view I now endorse and label 'hybrid ethical neo-expressivism' (HENE). In the remainder of this section, I aim to defend the plausibility of this HENE by dispelling some initial worries.

4.4 Hybrid ethical Neo-Expressivism: Objections

Worry 1: First, one might think that HENE carries a commitment to either moral realism or error theory, if one thought that the beliefs involved in ethical judgment represent some moral way the world might be.

Response: If this worry held any weight, it would also apply to the *non*-hybrid ethical neo-expressivist response to the modified wishful thinking problem discussed above. That response required holding that agents can form ordinary beliefs in moral propositions in carrying out certain inferences. Regardless, hybrid (and non-hybrid) ethical neo-expressivism can avoid the commitment to realism or error theory anyway. Simply accepting that ethical sentences semantically express propositions does not carry any realist commitments. Any such commitment would require further representationalist assumptions about propositional content. So too, simply admitting that ethical judgment is partly constituted by a belief does not carry realist commitments without further representationalist assumptions about the nature of belief. All that is required for our purposes is that beliefs are the sorts of mental states that are generally rationally responsive to reasons and evidence, and can figure in inferences (understood as mental processes). The further claim that beliefs necessarily aim to represent some naturalistically specifiable features of the world must be argued for, not simply assumed.

Worry 2: One might think that hybrid ethical neo-expressivism is incompatible with a prevalent Humean psychological picture according to which doxastic states are ‘distinct existences’ from desire-like states. (Indeed, this worry seems to inform Bar-On, Chrisman, and Sias’ neutral stance on whether ethical claims express beliefs.) This Humean picture figures into explanations of action and motivation; at the level of folk psychology, motivation to act is generated by the combination of a desire with a belief about how to satisfy that desire (see 1.4.1). By introducing a cognitive element to ethical neo-expressivism HENE seems to violate the Humean division between beliefs and desire-like states.

Response: It is unclear why hybrid ethical neo-expressivism would be any worse off in this regard than any other hybrid metaethical theory, since hybrid theories contend that ethical judgments or ethical claims somehow involve both doxastic and desire-like states. Moreover, it does not seem that hybrid theories generally have a problem here. In order to accommodate a broadly Humean picture, hybrid theorists only have to maintain that though ethical judgment is constituted both by doxastic and desire-like states, these states are themselves distinct components of the judgment. Hybrid theorists can hold that it is the combination of non-cognitive attitude and belief in ethical judgment that explains moral motivation.

Even if it is true that hybrid ethical neo-expressivism is incompatible with the Humean psychological picture, it is not clear that this would be a significant problem. While it is certainly true that in sophisticated mature human cognition there is a sensible distinction to be made between beliefs and desires, it is not clear that the Humean requirement that no mental state could have both a belief-like world-to-mind fit and a desire-like mind-to-world fit at once is anything more than philosophical dogma. In the next chapter, I argue that basic ethical judgments have both descriptive and directive directions of fit at once; thus, the overall view defended in this dissertation rejects the Humean picture of psychology anyway. Still, it seems to me that there is room for a hybrid version of ethical neo-expressivism not committed to rejecting Humeanism.

Taking stock: As we have seen, ethical neo-expressivism makes use of an independently plausible distinction between two senses of ‘expression’ to explain (i) how speakers can give voice to certain non-cognitive states in making ethical claims, and (ii) the various logical and semantic continuities between ethical sentences and other sentences. This distinction allows ethical neo-expressivism to completely avoid standard Frege-Geach problems that face other expressivist theories. I take this to represent a major advantage of the view. However, I argued that there are also *epistemic* continuities between ethical judgments and other judgments that need to be accounted for. The best way to account for these continuities while retaining an expressivist explanation of moral motivation, I propose, is to endorse a hybrid version of ethical neo-expressivism. This hybrid ethical neo-expressivism, like standard ethical neo-expressivism, is a minimal view that can be further developed in different directions.

We have now reached a turning point in the dissertation. Whereas Chapters 1, 2, and 3 were devoted to setting out some desiderata on a satisfying account of ethical thought and discourse and showing how various proposals in the literature fail to meet these desiderata, I have now presented the first central positive proposal of the dissertation: HENE. The remainder of this chapter is devoted to developing a more substantive version of HENE that takes on several commitments going beyond what the minimal version so far considered strictly requires.

These further developments, I shall argue, afford a more comprehensive explanation and justification of the various features of ethical thought and discourse.

4.5 Expression, Claims, Assertions

In the remainder of this chapter, I develop an account of ethical assertion against the background of HENE. Developing this account will be aided if we have some subtle but important distinctions in the philosophy of language concerning expression and speech acts in mind. With the relevant distinctions in hand, I can then provide a multi-faceted treatment of the variety of things we can use ethical sentences to do. The following distinctions will concern us here:

- A distinction between the act of *claiming* that P and the act of *asserting* that P.
- A distinction between a-expressing M and performing a speech act that has M as one of its sincerity conditions.
- A distinction between making a non-ethical claim that has ethical content and making a genuinely ethical claim (see 4.3).

Concerning the first distinction: The idea is that asserting that P is one way, but not the only way, of claiming that P. What makes the difference between asserting that P and claiming (without asserting) that P has to do with the conventions (or lack thereof) governing the act. To illustrate the difference, it will be useful to examine how neo-expressivism about *avowals* (=spontaneous, linguistically articulate self-ascriptions of occurrent mental states made from a first-person perspective) answers the question “Are avowals assertions?”.

According to avowal neo-expressivism (Bar-On 2004), in avowing a mental state M, one a-expresses M, using a linguistic vehicle that s-expresses the proposition that one is M. Bar-On notes that “on a suitably *weak* understanding of assertion, we can at least agree that one can express one’s feeling, thought, etc., even when making an assertion” (2004, p. 247). But “if we understand assertion as a specific kind of speech-act, with a relatively well-defined point or purpose and felicity conditions, on a par with making a request, issuing a command, asking a question, then we may insist that at least some acts of expressing one’s feeling, thought, etc. in language are not acts of making an assertion, issuing a statement, or delivering a report” (ibid.). This is because assertion, qua speech act, is constituted by certain social conventions or rules (discussed in more detail below), whereas a-expression does not essentially depend on social or even linguistic conventions.[12] That is, contrary to what is often assumed, expressive acts are not just a category of speech act. Ordinary avowals (as in “I would really like a glass of water right now”) are expressive acts that use declarative sentences as linguistic vehicles, yet are not assertions (understood in a strict sense). So, according to neo-expressivism, one can a-express a mental state – including belief – using a declarative sentence, without thereby making an assertion.

If this seems surprising, I suspect that is because ‘assertion’ is often used in a loose sense just to mean an affirmative utterance of a declarative sentence that expresses a belief in the content of that sentence. To keep our discussion precise, let us understand ‘claim’ as follows:

Claim: S claims that P when S utters a declarative sentence that semantically expresses the proposition that P, in an act of putting P forward as true.

Since assertions are (among other things) acts of putting forward the asserted content as true, to assert is (at least) to make a claim. However, not all claims are assertions. This is because one can make a claim without invoking the norms constitutive of assertion, including norms that

[12] Consider that crying, laughing, smiling, etc., are all bits of behavior that are plausibly suited by natural selection to a-express mental states (of sadness, joy, happiness), but are not best understood to do so via social convention. See also Ekmann’s work on the universality of certain facial expressions as expressive of emotion (see e.g. Ekmann and W.V. Friesen, 1971).

require one to be able to answer challenges to one's claim, or to provide justification for thinking the claim is true, etc. Instead, one may directly a-express one's belief in making a claim, without thereby making an assertion. For instance, according to Bar-On, when one avows being in pain, one both a-expresses one's pain and the belief that one is in pain.[13] Yet avowals on this view are not assertions because, among other things, avowals are *properly* epistemically baseless, whereas an assertion is only properly made if one is able to provide epistemic justification for the assertion.[14] As another intuitive example of claiming without asserting, consider a case wherein one looks out the window and is surprised to see that it is raining. One may remark "It's raining" just to oneself, thereby a-expressing one's belief that it is raining. But one would not be best understood as *asserting* that it is raining – for one thing, the illocutionary point of assertion is to get one's audience to accept what one asserts, yet in this case, one does not even have to have an audience.

Having just argued that expressive acts and speech acts comprise two separate pragmatic categories, it is important to clarify that this is not to say that a particular act could not at once be a speech act *and* an expressive act. It is just that such an act is not expressive *in virtue* of being a speech act of a certain kind, nor is it a speech act *in virtue* of being expressive in a certain way. Expressive 'force' and speech act force are orthogonal to each other, and may both be aspects of

[13] This is the Dual-Expression Thesis (Bar-On 2004, p. 307), also discussed in 2.3.

[14] According to Green (2017), for instance, a fidelity condition on the speech act of assertion is that one should be able to provide "strong justification" for one's assertions when challenged. That (at least some) basic self-beliefs are epistemically baseless in some sense is a common starting point in the self-knowledge literature, shared by authors with very different treatments of self-knowledge, including Bar-On (2004), Wright (1998), and Coliva (2016a). Wright, for instance, claims "the demand that somebody produces reasons or corroborating evidence for such a claim about themselves – 'How can you tell?' – is always inappropriate" (1998, p. 14).

a single utterance.[15] (We are now proceeding to the second distinction to be discussed in this section)

Arguably, S's assertively uttering that P also counts as expressing belief that P, since asserting that P is one way to claim that P, and claiming that P is an apt way to express belief that P. Asserting that P does not preclude one from also a-expressing belief that P. The point I wish to make here is just that S does not a-express belief that P simply *in virtue of* the fact that having that belief is a sincerity condition on making that assertion. The relation between a speech act type and its sincerity condition is established by linguistic convention; the relation between an expressive act and the mental state expressed is more basic, found for instance in the communicative behavior of non-linguistic creatures. It is common to say that speech acts *express* the mental states that are their sincerity conditions.[16] In the metaethical literature, too, it is not unusual to say that ethical assertions *express* beliefs (if one is a cognitivist) or that they express motivationally-charged attitudes (if one is a non-cognitivist).[17] I am urging caution about these ways of speaking; they are not totally inaccurate, but the reality is a bit more complicated than they suggest.

The distinction between expressive and speech act force opens up some new options for hybrid and pure theories, recast in terms of speech act theory. For instance, some conceivable positions might be that:

[15] I get this idea about expressive vs. speech act force from conversation with Nadja-Mira Yolcu. According to Yolcu and Freitag (unpublished manuscript), "[l]inguistic acts usually have both an *illocutionary* and an *expressive* dimension" (7).

[16] See for instance Searle (1979, p. 5): "The psychological state expressed in the performance of the illocutionary act is the *sincerity condition* of the act".

[17] For instance: as we have seen, Copp accepts that, in addition to expressing beliefs, "in making moral assertions, we express certain characteristic conative attitudes and motivational stances" (2001, p. 1). And Schroeder, even in a piece devoted to articulating an adequate account of *expression* for expressivist, ends up arguing that the most suitable notion is that of "*assertability* expression", according to which assertions express those mental states that assertors must have for the assertion to be appropriate (2008, pp. 108-111).

- Ethical claims, insofar as they are assertions, require that the asserter have well-justified belief (or possibly knowledge) to count as sincere. As expressive acts, they just express representational belief. (*Pure cognitivism*).
- Ethical claims, insofar as they are assertions, require just that the asserter have well-justified belief (or possibly knowledge) to count as sincere. As expressive acts, they express a motivationally-charged non-cognitive state. (*Ecumenical cognitivism*).
- Ethical claims, insofar as they are assertions, require just that the asserter have an appropriate motivationally-charged non-cognitive state to be sincere. As expressive acts, they express a representational belief. (*Ecumenical expressivism*).
- Ethical claims, insofar as they are assertions, require just that the asserter have an appropriate motivationally-charged non-cognitive state to be sincere. As expressive acts, they express just a motivationally-charged non-cognitive state. (*Pure non-cognitivism*).

An additional way in which we can use ethical sentences is in making *ordinary* (that is, *non-ethical*) claims and assertions (this is the third distinction listed above). As discussed in section 4.3, one might make use of an ethical sentence to make an ordinary, non-ethical claim, albeit one with semantic content that concerns an ethical matter. There is a sense in which any declarative utterance of a basic ethical sentence (i.e., a sentence make central use of an ethical predicate, as in "Torture is wrong") can be considered an ethical claim; this would be so, for instance, if we individuated domains just by subject matter, keyed to the domain's distinctive vocabulary. But in 1.2, I argued against this principle for domain individuation in favor of a functionalist approach. So we need to determine what function or purpose an utterance of a basic ethical sentence is supposed to serve before we can conclude that the utterance is an ethical claim. I will have more to say about the function of ethical claims in Chapter 5.

HENE, as described above, provides a hybrid account of ethical claims in general, with the idea that we can use such claims to a-express moral judgments comprised by moral beliefs and motivationally-charged non-cognitive states. In what follows, I aim to articulate and defend a compatible hybrid account of ethical *assertion* as a category of ethical claim.

In sum: this section has distinguished between the categories of *claim* and *assertion*, distinguished between the *sincerity conditions* on speech acts and what mental states such acts can *express*, and distinguished between *ethical claims* and *non-ethical claims with ethical content*. Whether an utterance of an ethical sentence constitutes an ethical claim or a non-ethical claim with ethical content depends on the function of that utterance. In supporting HENE, I endorse thinking that ethical claims a-express both motivationally-charged states and beliefs. In the next section, I shall begin arguing that the possession of an appropriate motivationally-charged state and of moral belief are also sincerity conditions on ethical assertion.

4.6 Speech Acts, Assertion, and Lying

In this section, I shall begin arguing that some ethical claims are assertive speech acts. The argument is that (i) it is sufficient for a claim's being an assertion that the claim can be used to fully lie, rather than to merely mislead, and (ii) intuitively, one can fully lie in making an ethical claim. I then argue that ethical assertion forms a distinctive species of assertion, as it is subject to a further sort of insincerity condition ordinary descriptive assertions are not: one who makes an ethical claim while lacking any motivation whatsoever to act in accordance with that claim is being *hypocritical*.

First, it will be necessary to provide some further description of the relevant aspects of speech act theory. A speech act is an act that one can perform in speech (or in thought)[18] by

[18] The term 'speech act' is somewhat misleading, given that one can make speech acts in thought. A better term might be 'illocutionary act', but I shall persist with 'speech act' because that phrase is more familiar, and because I

saying (aloud or to oneself) that one is performing that act.[19] For instance, one can order someone to close a door by saying that one does so ("I order you to close the door"). The *illocutionary force* of an utterance – what act that utterance is being used to accomplish (e.g., *ordering*, by saying "shut the door!") – can be distinguished from the *locutionary content* of the utterance (i.e., the propositional content *that the door is shut*). Illocutionary force and locutionary content can also be distinguished from the *perlocutionary effect* of an utterance – the effect it has on its audience (in our example, that one's addressee forms the intention to shut the door).

A *sincerity condition* on a speech act is a mental state that the issuer of the act must be in if the act is to be properly performed. Importantly, though an insincerely performed speech act is in a sense defective, it may nevertheless still succeed as the kind of act it is. I can succeed in ordering you to close the door by uttering "close the door!" even if I do not wish for you to close the door at all. (If you were to discover I did not wish this, it would be natural to ask: "then why did you tell me to do it?"; but notice that the question still presupposes that I *did* in fact tell you to close the door.)

One way to categorize families of speech acts is according to the ways they are capable of being performed insincerely.[20] If A *promises* to phi in uttering "I will phi", A's utterance will be insincere if he does not intend to phi. But A might utter the very same sentence ("I will phi") instead with the illocutionary force of a *prediction*, rather than a promise. In that case, A's utterance would be insincere if he did not *believe* that he would phi. We can classify speech acts according to the sorts of normative failures to which they are subject. This is a main point of

want to avoid suggesting that the illocutionary force of an utterance is of primary importance. As we shall see in the next chapter, the perlocutionary effects (discussed shortly) of speech acts are also significant.

[19] See Austin's classic (1962) and Searle (1972).

[20] Searle (1979, pp. 4-5) discusses this under the heading of what mental state the relevant speech act expresses. But see above discussion of the distinction between a-expression and speech acts.

Stainton (2016), who identifies ‘full-on stating’ (for our purposes, we can identify this with asserting) as a distinctive kind of speech act by identifying a kind of normative failing specific to it - namely, it’s being ‘lie-prone’ (p. 400). As the ‘I will phi’ example shows, and as Stainton likewise recognizes, investigating the possible normative failings of an utterance can be more revealing of its illocutionary force than that utterance’s syntactic form is. The fact that the very same sentence can be used to express different mental states on different occasions of use suggests that the difference must be accounted for in *act-theoretic* terms, rather than in terms of the *semantic meanings* of the words used alone.[21]

Another way to categorize families of speech acts is according to their purpose.[22] I will follow Millikan in identifying the function of speech acts in terms of their effects on hearers (what Austin called their perlocutionary effects): on this picture, the function of the assertive family of speech acts is to produce (true) beliefs in hearers, whereas the function of the directive family (including commands) is to bring about the relevant fulfilment-condition (Millikan, 1984, Chapter 3; 2004; 2005, pp. 171-173). (I discuss this idea in more detail in Chapter 5). Now, it seems to me that there is an indirect connection between the sincerity conditions on a speech act and the function of that act-type: part of the explanation for why a speech act type H has mental state M as its sincerity condition is that H could not perform its function unless those H-ing were in M often enough. For instance, the speech act of promising could not perform the function of assuring one’s addressee that one will do what one has promised if those making promises never intended to do what they promised. If we failed to intend to do what we promised enough of the time, the act of promising would lose its point, and so would eventually die out – hearers would

[21] See Davidson (1979).

[22] See, e.g., Searle (1979, p. 2). As is discussed shortly, I prefer to think of the purpose of language devices in terms of Millikan’s notion of ‘proper function’ (Millikan, 1984).

no longer find it useful to believe that speakers will do what they 'promise' to do; and speakers will no longer find it useful to 'promise' because this would fail to give hearers any assurance about what they will do.[23] So, as a methodological point, I shall take intuitions about when an utterance is insincere as evidence of the type of illocutionary force that utterance has, and I take such evidence to be strengthened if an explanation can be given of why satisfying that particular sincerity condition would be important for the fulfillment of the function of the speech act type in question.

Our central question here is: what is it that we are doing when we make ethical claims, such as when uttering "Poaching elephants for their tusks is morally reprehensible"? I shall now argue that ethical claims are often assertions. This is because (i) it is a sufficient condition for an utterance's being an assertion that it is capable of being used to fully lie, and (ii) ethical claims can be used to fully lie.[24] I discuss (i) and (ii) in order.

There is a conceptual connection between the notion of assertion and lying. In particular, it seems that "one who asserts that P lies if she does not believe that P" (Green 2017, 9).[25] There are a variety of ways for one to be deceptive in making a claim, but not every case of deception is an instance of lying. *Insincerity*, as I shall understand it, consists in a culpable mismatch between what one says, conveys, or expresses (or appears to express), and the mental states one is actually in (one's relevant actual beliefs, intentions, attitudes, and so on). In order for a speaker to *lie*, it is not sufficient simply that they fail to believe some content *conveyed* by their assertion.

[23] This explanation of the function of promising is modeled on Millikan's explanation of the stabilizing proper functions of descriptive and directive representational devices (1984, Ch. 3).

[24] Though I use the notions of lying and hypocrisy in discussing ethical claims, I take no view on whether lying, or hypocrisy, are themselves morally wrong. For the purposes of this paper, lies and hypocrisies are violations of illocutionary-act norms - whether they also violate moral norms is a further question.

[25] See also Dummett (1973, p. 356). And see Stokke (2013) for an argument that lying should be defined as asserting that which one believes to be false.

Susan, in uttering "I broke a finger yesterday", may thereby lead Bob to think it was her own finger she broke (this is an implicature of Susan's utterance). Susan's utterance would be deceptive, or misleading, if she meant for Bob to come to believe she broke her own finger when it was in fact Alex's finger she broke. But in such a case, Susan has not, strictly speaking, lied to Bob, because she does believe what she strictly stated, namely, that she broke *a finger* (though not her own). This illustrates that in order to lie (and not merely mislead), it must be the *asserted* content of one's claim that one fails to believe.[26] Lying is thus a form of insincerity specific to assertion. This might be explained in terms of the function of assertion for producing (true) hearer beliefs, as follows: If speakers frequently enough assert propositions they do not believe, then, given that most speakers are mostly correct in what they believe, the propositions they assert would frequently fail to be true, and so hearers could not reliably form true beliefs on the basis of what others say, in which case the practice of assertion loses its point.

Can one fully lie in making an ethical claim? Intuitively, it seems that one can. In support of this intuition, consider the following case: Don, a politician, does not believe that there is anything morally wrong with accepting money in exchange for political favors (as far as Don is concerned, he deserves whatever benefits he can get from such exchanges). However, Don knows that in order to secure reelection, he must do everything he can to garner the support of those who oppose political corruption. In attempting to do so, Don claims "Politicians ought not take bribes for political favors". I submit that Don's utterance is not merely misleading, but a full lie; Don does not accept what he says. In fact, it seems to me that it is as plausible to take Don to

[26] As Stainton points out in this regard, our judgments of whether a person S has fully stated that P vs. whether S's utterance merely conveyed that P track our judgments of whether a person could have lied vs. merely misled given S does not believe that P (2016, p. 405).

have lied in this case as it would have been if he had made a non-ethical claim that he did not believe, such as: “all politicians accept bribes”.

One might wonder whether having the intention to get one’s audience to believe the asserted content of one’s utterance is a necessary condition on lying. For present purposes, I take no official position on this matter, though my own intuition is that such an intention is not a necessary condition – so-called ‘bald-faced’ lies not intended to convince an audience at all still seem to me genuine lies.[27] Nevertheless, even on more stringent definitions of lying, it seems clear that some ethical claims can count as lies – for instance, we can easily imagine in the case of Don that he really does intend to convince his audience of what he says.

4.7 Ethical Assertion and Hypocrisy

Consider a variation on the case considered above: imagine now that Don does accept what he says; he thinks politicians really ought not take bribes. Nevertheless, Don routinely accepts bribes, assigns positions of power to his family members, and so on, without hesitation or regret. Ordinarily, we might take this way of acting as evidence that Don did not believe corruption is wrong, but still, in some ways of describing the case it seems initially plausible that Don could act this way while still believing corruption is wrong. Perhaps the temptation to accept bribes overwhelms his better judgment, or perhaps he is an amoralist who accepts certain moral claims are true but does not think he has any reason to act in accordance with those claims. There are other ordinary cases where it seems, prima facie, that a person holds a genuine moral belief which they fail to be properly motivated to act upon. For instance, consider a member of a

[27] My intuitions about lying, then, are basically in accordance with the view put forward by Stokke (2013), who argues that lying should be defined as asserting that which one believes to be false. Stokke does not require that the content of the lie be false, or that the speaker intends for her audience to come to believe the content of her assertion for it to be a lie. Thanks to Paul Bloomfield for pointing me to Stokke’s account and for valuable discussion on this topic.

search committee who, despite believing that men and women are equally capable of doing philosophy, systematically gives preference to male philosophers in his assessment of job candidates. Or consider someone who is convinced by moral arguments for veganism, yet routinely finds themselves consuming animal products when vegan alternatives are readily available.

Still, one might wonder whether the cases described above are genuine possibilities; why not take the actions and underlying motives in these cases to provide evidence that the relevant subject does not genuinely hold the moral belief in question (even when they think they do), or is perhaps being irrational in some way? This is the line of response likely to be taken by defenders of motivational internalism, as discussed in Chapter 1. (This thesis, recall, holds that there is a necessary connection between genuine moral judgment and motivation, such that if one rationally and sincerely forms an ethical judgment, one must be at least minimally motivated to act in accordance with it.)

Setting aside the initial challenges that these sorts of cases pose to the internalist thesis, what is important for our purposes here is to notice that these cases highlight – rather than undercut – a distinctive feature of ethical thought and discourse identified in Chapter 1; namely, that there is at least the expectation of a connection between making an ethical claim and being in a motivationally-charged state (see 1.4.1).

The account I now propose explains this feature as follows: Supposing that Don believes that corruption is morally wrong, then, if he lacks any motivation whatsoever to at least avoid or discourage corruption in the relevant circumstances, Don's claim was insincere. Don violates a sincerity condition on ethical assertion requiring one to be at least minimally motivated to act in accordance with the ethical assertions one makes. I take this sincerity condition to be distinct from the sincerity condition on ethical assertion requiring that assertors believe the content of their assertions, discussed in Section 4.6. Thus, Don's utterance could be insincere in virtue of the fact that he lacks any motivation at all to at least avoid corruption, even if Don nevertheless believes the content of his ethical assertion. In such a case, though Don has not lied, his assertion is insincere in virtue of being an instance of *hypocrisy* – he does not practice what he preaches.

To clarify: my proposal is that there are two sincerity conditions on ethical assertions: first, that one believes the content of one's ethical claim, and second, that one be motivated to act in accordance with that claim. Ethical claims are thus hypocrisy-capable, in addition to being lie-capable. Where liars state that which they believe to be false, hypocrites propose a standard for action to which they do not (or would not in the appropriate circumstances) hold themselves.[28] These two sincerity conditions, of course, mirror HENE's account of the sort of mental state(s) expressed by acts of making ethical claims. But the account of ethical assertion I give here further justifies the sorts of social censure it is appropriate to apply when someone makes a public ethical claim while lacking the required belief or motivation. While expressive failures are possible, they are not always subject to the same kinds of admonishment or criticism as insincere speech acts.

4.8 Ethical Assertion and Moore's Paradox

Now, one might be tempted to think that Don's behavior provides evidence that he does not genuinely judge (believe, accept) that there is anything wrong with corruption. If, contra my proposal, this is the correct explanation, then Don's claim is straightforwardly insincere in the

[28] Though I think it useful to continue talking about hypocrisy as the relevant defect in the case of Don and a wide range of similar cases, it may be that the notion of hypocrisy is not broad enough to capture all the ways an ethical claim can be defective with respect to motivation. We may want to go for something like inauthenticity, or bad faith, or dishonesty, as the relevant defect in some cases. I am open to this possibility, but I shall continue the discussion in terms of hypocrisy.

way lies are, and so there is no need to suppose that his ethical claim can additionally go wrong by being an instance of hypocrisy. In short, one might think ethical claims are lie-capable, but they are not in addition hypocrisy-capable.

However, the above line of thought seems to me to take on an explanatory burden. The default view, I suggest, is one like the view proposed here, which recognizes a difference between making an ethical claim that one does not accept, and making an ethical claim one does accept but which one fails to be motivated to act upon. That this is the default view is suggested by the prima facie possibility of amoralism and akrasia. The view I propose straightforwardly accommodates the difference between sorts of insincerity ethical assertions are subject to.

Here is further support for the claim that there are two distinct ways for ethical assertions to be made insincerely: notice, first, how the connection between assertion and belief can be illustrated through anomalies such as Moore's paradox: imagine that S utters the following:

(R) "It's raining, but I don't believe that it's raining."

The strangeness of asserting such a sentence reflects the fact that when one makes an assertion (that it's raining) sincerely, one believes the content of what one asserts. The added conjunct ("but I don't believe that it's raining") disavows the very state that would be the sincerity condition of an assertion of the first conjunct. Notice that we can generate Moorean paradoxes using ethical assertions, as well:

(M) "Murder is wrong, but I don't believe that murder is wrong."

This bolsters the argument of Section 4.6. If (contra my argument) one could sincerely make an ethical assertion without believing the content of that assertion, then (M) would not be Moore-paradoxical.

But Moorean paradoxes are not limited just to cases concerning belief. We can also construct a motivational analogue of Moore's paradox:

(N) "We ought not consume animal products, but I am in no way inclined to avoid consuming animal products."[29]

Though I take no stance here on the correct analysis of the absurdity of Moore's paradox, I think at the very least we should not limit our analyses so they can only apply to cases involving belief.[30] To assert (N) would be absurd in the same way as asserting (M). In each case, it seems, the assertor is overtly manifesting insincerity, where this is strange because we usually attempt to conceal when we are lying and when we are being hypocritical (with some exceptions, of course: hypocrisies can be overt in the way that bald-faced lies are).

And note that there is no motivational analogue of the paradox for non-moral assertions; there is nothing Moore-paradoxical about:

(R*) "It's raining, but I don't care at all that it is."

The non-paradoxicality of (R*) suggests that the paradoxicality of (N) is generated by a distinctive feature of the ethical domain. What accounts for the paradoxicality of (N) then is distinct from what accounts for standard non-ethical Moore-paradoxical assertions, such as (R).

My suggestion is: first, that (M) is paradoxical in the same way that standard Moore-paradoxical sentences (such as [R]) are, and second, that (N) is paradoxical in a parallel but

[29] See Cholbi (2009) for support. See also Schroeder (2008b) regarding the expressivist's 'parity thesis' that ethical claims express motivational states in the same way that non-ethical assertions express beliefs. Given this parity thesis, we should expect expressivists to predict that (N) is paradoxical in the same way that (M) is. But see Jack Woods (2014) for critical discussion of this expressivist version of Moore's paradox. Woods reports that he finds (N) less paradoxical than standard Moore-paradox sentences, such as "It's raining, but I don't believe it." I must confess I don't share Woods' intuitions. However, at the very least, I do not need to claim that (N) is paradoxical to the same *degree* as ordinary Moore-paradoxes; only that it has the same *kind* of paradoxicality (thanks to Bill Lycan for suggesting this point).

[30] In addition to the motivational analogue to Moore' paradox just discussed, there are other versions as well, including expressive conflicts, such as if one were to utter "this is not painful at all" while grimacing or wincing, or if one were to say "This is so interesting" while yawning. See Bar-On (2004, p. 219 and pp. 375-376).

distinct manner that concerns motivationally-charged states rather than beliefs. The strangeness of standard non-ethical Moore-paradoxical sentences reflects the fact that belief is a sincerity condition on assertion; the strangeness of Moore's paradox applied to ethical claims in (M) reveals that belief is also a sincerity condition on ethical assertion. And the strangeness of the motivational analogue of Moore's paradox in (N) reveals that being in a motivational state is a further sincerity condition on ethical assertions. The presence of these two distinct sorts of Moorean paradoxes for ethical assertion provides further support for the view I endorse here, that ethical assertions can be insincere in two ways, one concerning belief, the other concerning motivation.

In the next chapter, I further support this view about ethical assertions, along with HENE, with a biosemantic account of what I see as the dual role of ethical claims and judgments. Once such an account is given, we will have an explanation of the existence of norms on ethical assertion that would require agents to have the relevant belief and non-cognitive state for their assertion to be sincere, and an explanation for why it would be important for us to be able to express both beliefs and motivational states in making ethical claims.

4.9 Advantages of the view

In the remainder of this chapter, I discuss how the considerations presented so far put pressure on both pure cognitivist theories and pure expressivist theories, thereby supporting hybrid views. I then highlight some additional features of the hybrid account of ethical assertion I have proposed.

The considerations presented in Sections 4.6 and 4.7 of this chapter constitute a challenge to pure cognitivism and pure expressivism. I argued that there is a commonsense, intuitive distinction between two ways for ethical assertions to be made insincerely, a distinction reflected in the complex function of ethical assertion. If we were to grant the pure expressivist's understanding of ethical assertion (along the lines of Schroeder's assertability-expression, for instance), it becomes a puzzle how we are to explain this distinction in ways of being insincere. The expressivist-friendly explanation has it that insincerity in ethical assertion occurs when a speaker makes an ethical assertion while failing to have any motivation whatsoever to act in accordance with it (this is what I have been calling hypocrisy). But this leaves no room for recognizing other ways in which ethical assertions can be made insincerely, including by being lies.

This is not only a problem for pure expressivism. Pure cognitivism predicts that S's ethical assertion that M can be insincere only when S does not believe M (so S has lied), where 'belief' is here understood in a representational sense. This, again, leaves no room for recognizing the kind of insincerity involved in hypocritical ethical assertions. The pure expressivist and the pure cognitivist, then, each face the explanatory burden of explaining away the intuition that there are two kinds of insincerity in ethical assertion. The need to explain this intuitive distinction should make hybrid metaethics our default starting point, rather than a direction we are forced into by dissatisfaction with existing 'pure' theories.

In order to locate HENE and the hybrid account of ethical assertion in relation to other hybrid theories, let us consider how they answer the various taxonomical questions posed at the end of Chapter 3 (Fig. 1).

Question	Answer
Qa: Is it constitutive of ethical judgment that it has both a belief and a desire-like component?	Yes

Qb: In order to properly issue an ethical claim/assertion, must one be in both a belief and a desire-like state?	Yes
Qc: What linguistic/psychological mechanism(s) explain how agents are related to the relevant *beliefs*?	For ethical claims: a-expression. For ethical assertion: sincerity condition
Qd: What linguistic/psychological mechanism(s) explain how agents are related to the relevant desire-like states?	For ethical claims: a-expression. For ethical assertion: sincerity condition
Qe: Is the content of the belief involved guaranteed to be true just in case the ethical claim/assertion is?	Yes
Qf: Are the motivationally-charged attitudes involved directed at particular acts, policies, persons, etc., or to a general characteristic of things?	Neutral

Fig. 1 How HENE answers taxonomical questions

Compared to the hybrid theories considered in Chapter 3 (see Chapter 3 fig. 2), the hybrid proposal defended in this chapter is most like Boisvert's expressive-assertivism. This should be no surprise, since Boisvert's view, like the view defended here, emphasizes features of the *act* of making an ethical assertion. However, as noted in Chapter 3, Boisvert's expressive-assertivism is committed to the Lockean assumption about meaning, which neo-expressivism and the account of ethical assertion I have offered is not.[31]

[31] Morgan (2016) has argued for a hybrid account of ethical assertion that coincides with my account in its central theses, including the rejection of the Lockean assumption, although we reach our views in quite different ways.

Before concluding, I would like to highlight one attractive feature of the view propose in this chapter, regarding the debate over motivational internalism (see 1.4.1).[32] While ethical claims have seemed to some to be necessarily connected to motivation, it has been difficult to specify the putative connection in a suitably weak way so as to avoid implausibly ruling out the very conceptual possibility of amoralism, for instance. HENE (and non-hybrid ethical neo-expressivism) can offer such a connection in terms of a-expression: subjects a-express motivational states in making ethical claims. Failures of motivation can be viewed as cases of expressive failures. In such cases, though one cannot be said to have a-expressed *one's own* motivational state, one has still a-expressed *a* motivational state (or; one's claim is *expressive of* that state). Regarding ethical assertion: the hybrid account of ethical assertion I have proposed holds that, as a matter of speech-act norms, ethical assertions are not properly made unless one is in the relevant motivational state. The view thus posits a defeasible but necessary connection between ethical assertion and motivation, namely, that if one makes an ethical assertion in full propriety one believes what one asserts and is at least minimally motivated to act in accordance with it.[33]

Now, 'propriety' and 'sincerity' seem to be normative notions. If being 'proper' with respect to one's ethical claims and assertions ultimately came down to no more than being motivated to act on the moral considerations one recognizes, then the internalist position articulated above may end up saying no more than that we ought to be motivated to act in accordance with the ethical judgments we make – yet this is a claim that would be accepted by

[32] This feature is also shared by non-hybrid ethical neo-expressivism.

[33] Bar-On, Chrisman, and Sias (2014) pose their internalist view in terms of 'propriety' conditions on ethical claims. One might wonder why they use 'propriety' conditions as opposed to 'sincerity' conditions. This may be because their theory is framed in terms of expressive acts, rather than speech acts, where 'sincerity condition' is a technical term in speech act theory. Since I adopt both HENE and a speech-act account, I can say that ethical claims can have propriety conditions when considered as expressive acts, and sincerity conditions when considered as speech acts.

all parties to the debate on motivational internalism (see Bloomfield, 2001, pp. 156-159 for discussion of this point). So, for the view discussed above to avoid triviality and defend a genuinely internalist view, it must hold that the normative force binding proper ethical claims to motivation must itself be something other than *moral* normativity. I do not think it is implausible to hold such a view, since it seems to me that there are various sources of normativity, and plausibly speech act norms and other communicative norms are of a fundamentally different kind as moral norms, though they often overlap (for instance, lying often not only violates a sincerity condition on assertion, but can also violate one's moral duties). At any rate, a positive argument is needed for thinking that ultimately all normativity is of a piece such that internalist theses that build in conditions of rationality, sincerity, etc., fail to articulate a non-trivial claim.[34]

Internalism, as it is ordinarily construed, is a thesis about ethical *judgment*, rather than ethical claims or assertion. One might wonder then, what qualifies HENE, ENE and the hybrid view of ethical assertion as supporting motivational internalism. In the case of HENE and ENE, the connection is just this: expressive acts, in Bar-On's view, can be performed in speech *or in thought.* Just as one can express one's pain publicly, by uttering "This hurts!", one can also suppress any outward expressions of pain yet still express that pain in one's own mind, by tokening "This hurts!" in thought. So too with ethical claims; as one judges, silently to oneself, that torture is wrong, one can be said to be making an ethical claim in thought. So the internalist account of ethical claims given above carries over to ethical judgments themselves. However, it does not seem so plausible that we can make *assertions* in thought only. For one thing, as noted above, the point of assertion is to bring about a certain belief in one's audience, yet this does not

[34] Cuneo devotes a whole book to defending the view that the existence of moral rights, responsibilities, and obligations are a necessary condition for linguistic communication (2014). I hope to respond to Cuneo's argument on another occasion.

seem to be what goes on when one makes an inner judgment. Still, the account of ethical assertion articulated above can be seen as supporting a version of *discourse internalism* (Copp 2001, fn. 43), according to which fully sincerely making an ethical claim that one morally ought to do something entails that one is at least minimally motivated to do it.

Now, discourse internalism is, strictly speaking, compatible with motivational *externalism* about ethical judgment, since it is silent on the relation between inner ethical judgment and motivation. Still, insofar as an ethical assertion is also an expressive act of making an ethical claim, the motivational internalism endorsed by HENE and ENE is operative. I further support this version of motivational internalism in Chapter 5 by giving an account of the function of ethical claims and judgments according to which our practice of making ethical claims depends on their being accompanied by motivational states in a critical proportion of cases. Thus, borrowing a term from Tresan (2006, 2009), the proposal is a form of *communal* internalism, according to which a community that made assertions with ethical *content* but where there was no correlation between such assertions and motivation whatsoever would not count as engaged in a practice of making genuinely ethical assertions. (See above distinction between ethical claims and non-ethical claims that have ethical content, 4.3).

4.10 Conclusion

In this chapter, I have made two positive proposals: First, I proposed a hybrid version of ethical neo-expressivism (HENE), a view that retains the advantages of ethical neo-expressivism while also adequately addressing a variation on Dorr’s wishful thinking problem. Second, I proposed a complementary hybrid account of ethical assertion. In this conclusion, I anticipate some of the themes of the next chapter by considering the relation between these two positive proposals, and why each is needed.

As discussed in 4.5, the relation between a speech act and its sincerity condition(s) is not the same as between an expressive act and the mental states expressed. Because of this distinction, I found it necessary to treat ethical claims qua expressive acts separately from ethical claims qua speech acts. Having made this distinction, one naturally wonders how the phenomena distinguished are related to each other.

Note that, in the neo-expressivist framework, expressive behavior is a sort of intentional behavior, a capacity for which is (a) shared by social creatures, and (b) that cuts across the linguistic/non-linguistic divide. We can express our mental states not just with articulate linguistic devices, but also with groans, winces, smiles, hugs, and so on. Many of these expressive devices are shared with social non-human species. A brief survey of animal communicative behavior reveals that animals employ an immense array of expressive vehicles to communicate with each other, from monkey alarm calls, to beaver tail slaps, to canine play bows, etc. Clearly, a-expression encapsulates a broad and heterogeneous category of behavior. Moreover, (c) the relation between expressive behavior and the psychological state expressed is immediate, in the sense that suitable placed receivers can *directly* recognize the mental state expressed, without having to engage in, say, inference from behavior. In addition, (d) expressive behavior is *naturally* designed to show one’s mental state, rather than being established by social convention. It is important also to consider the *purpose* of expressive behavior. Plausibly, the point of expressive behavior is (e) to bring about fitting responses in creatures positioned to recognize the state expressed. A monkey’s alarm call for instance, can perhaps be said to have the purpose of getting conspecifics to act appropriately to the danger at which the fear is directed. Taking (a)-(e) altogether, we have it that expressive behavior is a heterogenous class of behavior designed to show one's mental states for the purpose of enabling appropriate action in suitably placed receivers.

Speech acts differ from expressive behavior in that (a') only creatures capable of sophisticated linguistic communication can make them, (b') their existence is language-dependent, (c') the audience of a speech act cannot directly recognize the speaker's mental state simply from recognizing its status as speech act – at best, the performance of the speech act is good evidence that the speaker is in that state,[35] and (d') speech acts constitutively depend on social conventions, rather than natural design. But speech acts are similar to expressive behavior in that (e) their purpose is to get receivers to respond to them in certain ways. What I would like to suggest here (and refine in the next chapter) is that the way that speech acts induce corresponding hearer behavior is slightly different from the way purely expressive acts do so.

The idea is that purely expressive acts induce appropriate responses in receivers via a sort of emotional contagion mechanism, so that for instance the monkey's alarm call spreads fear throughout the group, thereby leading to the group avoiding the dangerous object of that fear. But speech acts induce appropriate responses in hearers in virtue of their entanglement in a system of social norms, a system that generates defeasible normative pressure on receivers to react to speech acts in certain ways, and that requires agents making speech acts to be prepared to back those acts up in certain ways. In effect, along with a capacity for producing speech acts comes a sort of 'normative currency'; speech acts generate (defeasible) reasons to act on the part of receivers, and they generate obligations on the part of agents. A payoff of the capacity to produce speech acts is the ability to exert greater social influence without using costly measures such as physical force, coercion, etc. While the capacity to produce and respond to speech acts

[35] But if, as suggested earlier, a bit of behavior constituting a speech act is *also* an expressive act, receivers can directly recognize the relevant mental state from the expressive behavior.

likely comes later in phylogeny than the capacity for expressive behavior, there is no reason to think that social creatures capable of producing and responding to speech acts would not have use for expressive behavior; indeed, the two might work in conjunction with each other. I take these issues up further in the next chapter, where I propose a biosemantic account of ethical judgment and ethical claims.

Chapter 5

Ethical Affirmation and Proper Function

5.1 Introduction[1]

In the previous chapter, I argued for a hybrid version of ethical neo-expressivism (HENE), according to which ethical sentences semantically express propositions, and when subjects utter ethical sentences in acts of making ethical claims, they express both a belief and a motivationally-charged non-cognitive state. I have argued that this account goes further than any other account considered so far in accounting for the various features of ethical thought and discourse. I then extended the account to provide a hybrid speech act analysis of ethical *assertions*, understood as a category of ethical claim.

In this chapter, I continue to develop the hybrid program of the dissertation by situating it within a broader *biosemantic* framework (owing to Millikan 1984). I shall employ that framework to generate a hybrid account of ethical *judgment* and further bolster the hybrid accounts of ethical claims and assertions given in the previous chapter. Along the way, we shall also be able to consider a tentative explanation of the evolutionary origins of ethical thought and discourse. While neither the biosemantic account provided in this chapter or the hybrid account of ethical assertion in the previous chapter are necessary commitments of HENE, they are amenable to it, and I think the additions are justified by the greater explanatory resources they offer and the nuanced treatment of ethical thought and discourse they provide. I work out an account of ethical judgment that gives it a more nuanced treatment than the 'motivationally-charged non-cognitive state + belief state' placeholder I have been using thus far. Unlike the

[1] Parts of this chapter are based on previously published work (see Johnson, forthcoming).

sophisticated expressivist projects of Gibbard and Blackburn, however, I do not need this account to do important work accounting for the semantic and cognitive continuities. My concerns in this chapter have more to do with exploring the place of ethical thought and discourse in the natural world and their significance in our lives than with solving technical problems with expressivist semantics.

A brief comment is in order concerning the choice to develop my hybrid project in a biosemantic framework. As we shall soon see, Millikan's biosemantics is radically at odds with more standard approaches to semantics and metasemantics taken in metaethics. A recurring theme of this dissertation is that by questioning certain standard assumptions in the literature, the possibility of unconventional yet promising treatments of various aspects of the ethical domain are revealed. This holds true in the present chapter as well. Millikan's biosemantics, despite comprising a prominent and influential theory of representational content (not to mention one that I find fascinating independently of its application here), has received surprisingly little attention in metaethics as a framework for thinking about ethical judgment.[2] The application of biosemantics to ethical judgment in this chapter offers a view that, like HENE, is difficult to categorize among existing metaethical theories. The biosemantic account to be provided, for

[2] As far as I am aware, the only direct applications of Millikan's view to metaethics in the literature are in Bergman (2019, 2021), Bloomfield (2018), Dowell (2016), Harms (2000), Joyce (2001a), Sinclair (2007, 2012), and Wisdom (2017). Bloomfield develops a Millikan-inspired tracking theory in giving a naturalist epistemology for tracking virtue. To my understanding, my own proposal is compatible with this account, though I do not make a substantive commitment to a particular normative theory. Dowell employs the radical semantic externalism of Millikan's biosemantics to disarm the Moral Twin Earth argument against moral descriptivism (due to Horgan and Timmons, 1991). While I am sympathetic to Dowell's conclusion, my main focus here is on ethical judgment, rather than the meaning of moral terms (but I discuss Dowell's argument again in Chapter 6). My own proposal agrees with Sinclair's in treating ethical judgment as a species of 'Pushmi-Pullyu Representation'; I here provide a further development of this basic idea. Recently, Bergman (2019) has also proposed a sophisticated biosemantic account of moral judgment. One difference between our views is that Bergman follows Artiga (2014) in treating *all* representations as 'Pushmi-Pullyu' Representations, whereas in my own view, the distinction between Pushmi-Pullyu Representations and purely descriptive and purely directive representations is important (see Millikan, forthcoming, for a response to Artiga). Harms and Joyce debate to what extent a biosemantic approach to ethics is committed to naturalist moral realism, and Wisdom proposes a proper function theory of moral realism. Again, my concern in this paper is with the proper function of moral *judgment*, rather than moral realism directly.

instance, integrates elements of realism, expressivism and prescriptivism, each of which are ordinarily thought to be competing views.

In the next section, I give an overview of the central features of Millikan's biosemantics. In Section 5.3, I then develop a biosemantic account of ethical judgments and ethical claims, and give several proposals about what they might be supposed to 'track' and what they are supposed to get us to do. Section 5.4 examines some further applications of the biosemantic account, including a discussion of the origins of moral reasoning, and some further nuances concerning the relations between ethical judgment, belief, and motivation. I conclude in section 5.5 by briefly situating the proposal of this chapter in the literature on hybrid metaethics.

I shall be discussing both the function of *self-directed, inner ethical judgment*, such as when one judges silently to oneself that one ought to be more generous, and the function of *other-directed, public, assertive ethical claims,* such as when one affirms to an audience that we ought to be doing more to eradicate global poverty. While I view each of these as contributing to the same distal social coordination function, it will be important later to distinguish the mechanisms by which inner ethical judgments and public ethical claims accomplish this function. I introduce the phrase 'ethical affirmation' as a neutral term to describe both inner ethical judgment and public ethical claim.

In this chapter, I will focus on ethical affirmations explicitly invoking 'ought', as in affirmations taking the form "X ought to phi (in circumstances C)". Although my focus is limited in this way, if it is right to think of 'ought', in its moral sense, as a core ethical concept, in terms of which others might be defined, then much of what I say will apply to ethical affirmations in general – but I make no commitment to this in the present chapter, leaving the extension of the approach for future work. At any rate, my focus will be on exploring an aspect of the meaning of ethical affirmations – their proper function – that does not essentially depend on the particular form that the affirmation takes. An ethical affirmation's particular form is relevant, however, to determining another aspect of its meaning, in terms of what Millikan calls its *semantic mapping function* (not to be confused with 'stabilizing proper function'. More on this below) (See Millikan, 2005b).

5.2 Biosemantics

Broadly speaking, biosemantics aims to provide a naturalistic teleological account of representational content, including both 'inner' representations in thought, and 'outer' representations in public language. The guiding idea of biosemantics is to use a biological model for thinking about representations through investigating their proper functions. The notion of 'proper function' is teleological, referring to what a thing is 'supposed to do' – this is significant for making sense of the possibility of *mis*representation – but it is also naturalistic, as proper function is ultimately explained in terms of the survival value of a device's Normal[3] effects across its actual historical environments (biological, cultural, or even developmental). Millikan's biosemantics can be seen as a sort of all-purpose tool for theorizing about thought and language; it is variously applied to predicates, names, sentences, judgments, concepts[4], speech acts, maps, labels, bee dances, beaver splashes, and so on.

5.2.1 Proper function and normal explanation

The most central notion in biosemantic analysis is that of the *proper function* of a device, which is roughly:

[3] 'Normal' has a technical meaning in Millikan's theory, to be explained shortly (hence the capital 'N').

[4] Or to use Millikan's now preferred term, 'unicepts' (2017).

Proper Function: A proper function of a device *o* is an effect instances of *o* have had, historically, that account for why *o* continues to be reproduced.[5]

Hearts, for example, have a proper function of circulating blood throughout the body; it is this effect which has, historically, accounted for why hearts have continued to be reproduced. Crucially, in giving an explanation of how a device performs its proper function, we do not necessarily examine those effects it *usually*, or *statistically normally* has, but only the effects that account for continued reproduction, which may occur with less than statistical regularity. Thus, the Normal conditions for the proper functioning of a heart will be those conditions hearts have historically been in when they have *actually* circulated blood throughout the body - including, for instance, the heart's being in a body, the presence of a closed circuit of blood vessels leading to the heart, etc. To illustrate that Normal conditions can come apart from statistically average ones, consider: the Normal condition for the performance of a proper function of sperm include that they find an ovum, but this is a condition that only a very small proportion of sperm are ever in (Millikan 1984, p. 34).

In applying the notion of proper function to representational devices, what matters are the effects the representational device in question has that account for *its* (that is, the representational system's) proliferation. This is not the same as considering what effects a representational device has that contribute to the reproductive fitness of the organisms that might produce or consume it. But there is surely some relation between a representation's effects and the purposes of producers and consumers, for it is only if the effects of a representational device are in some way of use to its producers and consumers that they will continue producing and using such representational devices, hence continuing the proliferation of those devices. While Millikan

[5] This is a rough gloss of the full definition given in *Language, Thought, and Other Biological Categories* (1984, Chapters 1 and 2).

recognizes that a token representation can be used by any number of potential interpreters for any number of their various purposes, representations in language usually proliferate because they have *stabilizing* proper functions; effects that serve the cooperative purposes of speaker and hearer, and so tend to keep producers producing them in standard ways, and consumers consuming them in standard ways. Thus, while a potential hearer might use my utterance of "It will rain soon" as a sign that I'm thinking about rain, or that I believe that it will rain, or of the typical volume of my voice, and so on, the effect that my utterance is supposed to have, that accounts for the proliferation of such assertive claims, is that it gets hearers to form the true belief that it will rain soon. It is because speakers often enough assert only what they believe (truly), and because hearers often enough believe what is asserted, that assertoric practices fulfil their cooperative function of producing true hearer beliefs. Thus, that speakers and hearers believe what is asserted is a stabilizing proper function of assertion: this effect is part of what accounts for the continued practice of making assertions.

The emphasis on the effects of representational devices on hearers already indicates a point of contrast between a biosemantic metaethics and other prominent approaches. For instance, a primary focal point, common to both expressivists and their opponents, is the mental states that ethical claims are said to express. This directs our initial focus to what producers are doing in making ethical claims.[6] From the biosemantic perspective, this emphasis on the *production* of ethical judgments can at best only provide half of the story when it comes to their function. For instance, in assessing expressivist theories of ethical thought and discourse, we should take care to ask *why* it might be useful for us to have a domain of discourse suited for

[6] An exception is ethical neo-expressivism, along with HENE. The neo-expressivist view emphasizes the effects that expressive behavior is supposed to have on suitably endowed receivers (see e.g. Bar-On and Chrisman 2009, p. 163).

giving voice to motivationally-charged attitudes. What end might such expressive behavior serve for suitably endowed *receivers*? Now, expressivists have offered an answer to this question: the expression of a motivationally-charged attitude is useful insofar as achieves a social coordination effect (Gibbard, 1990, pp. 64-69). But given the difficulties for traditional and hybrid expressivism discussed in previous chapters, it is worth revisiting the expressivist claim about social coordination from a biosemantic stance. Biosemantics can capture the idea that ethical judgment serves a distal social coordinating function, without commitment to an expressivist mentalist semantics.

5.2.2 Semantic mapping function

Another central element running throughout Millikan’s work on representation is the notion of a *semantic mapping function* (see especially Millikan, 2004, Chapter 6 and 7). On Millikan’s view, no representational system could have the function of representing just one state of affairs. Instead, it is of the essence of a representational system that there are rules for transforming significant articulate elements in a token representation of that system, where such transformations are supposed to map onto corresponding transformations in the state of affairs represented, yielding an indefinite number of possible representations that the system could produce. Just what the transformation rules are is relative to the representational system. So, more than one representational system could produce token representations representing the very same state of affairs, where differences in the degree and kind of articulation in the various representational systems explains why the token representations produced may nevertheless differ in content.

Thus, “It’s raining” and “Rain is falling here now” in their standard uses (see Millikan, 2005b, pp. 63-64) map the same state of affairs (have the same truth-condition), but their semantic mapping functions admit of different sorts of transformations (e.g. “It’s sleeting” is a transformation of the former, and “Rain fell in London yesterday” is a transformation of the latter), corresponding to different possible states of affairs they could represent (e.g. the precipitation of sleet at the place and time of utterance, or the presence of rain in London the day before the utterance). As Millikan illustrates the point: “compare the semantic-mapping function of a bee dance with that of an English sentence having the same truth-condition. Bee dances show by the angle of their axis where there is nectar relative to a line between hive and sun, but there are no transformations of the bee dance [as there are for the English sentence] that would tell about nectar location relative to objects other than hive and the sun, or about the location of anything other than nectar” (2005b, p. 64). In sum: variations in the type and degree of articulateness afforded by a representational system allow for more or less sophistication and variability in the sorts of world affairs that system can represent, but for all that, even quite disparate types of representational system can share in their satisfaction conditions.

Together, proper function and semantic mapping function are each central aspects of meaning. We might describe the former, very roughly, as the ‘force’ or purpose of a representation (by analogy with illocutionary force) and the latter its ‘content’ (Millikan, 2004, pp. 137-138). Accordingly, two token representations may have exactly similar semantic mapping functions, yet different proper functions, depending on what kind of representational family the token is copied from. The linguistic form “You will be here at 6 a.m. sharp tomorrow” has the same content (admits of the same significant transformations) but a different proper function when it is used to make a command, versus when it is used to make a prediction. Likewise, two token representations may serve the same basic purpose, e.g., getting the consumer to arrive at 6 a.m. exactly tomorrow, yet have different content because belonging to systems comprised of different semantic mapping functions: compare "You will be here at 6 a.m. sharp tomorrow" with "I hereby command you to be here at exactly 6 a.m. tomorrow", and with "Be here at 6 a.m. sharp, or else!", each of which admits of different significant transformations corresponding to different range of affairs represented.

My focus in this chapter is on ethical affirmations taking the form "X ought to phi (in circumstances C)". Transformations of this form will correspond to differences in the states of affairs represented; "I ought to be more generous" contrasts with "We ought to be doing more to eradicate global poverty" and "You ought to keep your promises" with respect to the subject and content of moral obligation. The basic form above also admits of significant transformations for "ought", including, for instance, non-moral senses of "ought", but also, e.g. "will", "can", etc., yielding a variety of non-ethical affirmations. Providing a detailed account of exactly what transformations to ethical utterance are supposed to map onto exactly which variations in the environment and/or subsequent behavior would provide an account of the content of ethical claims, and likely involve commitment to some first-order moral theory. For instance, consequentialists and deontologists will make different predictions about what world affairs ethical utterances are supposed to map onto. But crucially, in order for ethical claims and judgments to serve their *proper functions* (discussed below), moral agents do not need to understand the exact operation of the semantic mapping function of particular ethical claims and judgments. My aim in this chapter is not to give a detailed account of the semantic mapping functions of ethical affirmations. Rather, my focus shall be on describing the proper function, or purpose, of ethical affirmations invoking "ought". Because of this focus, I shall be able to remain fairly neutral on the correct semantic analysis for ethical sentences in general.

5.2.3 Directive, descriptive, and Pushmi-Pullyu Representations

According to Millikan, representations are divided into two broad categories, in terms of their proper functions: directive representations and descriptive representations. In speech, directive representations are standardly carried by the imperative mood, and descriptive representations are standardly carried by the indicative mood. Directive representations have fulfilment-conditions; descriptive representations have truth-conditions.

The stabilizing proper function of a directive representation is “to guide the mechanisms that use it so that they produce its satisfaction condition” (Millikan, 2005a, p. 171). Directive representations proliferate because, often enough, it serves the cooperative purposes of their consumers and producers that the consumers bring about the satisfaction condition of the directive. (Just what the satisfaction condition is, will be determined by the semantic mapping function of the representation). Of course, the explanation for why consumers find it beneficial to follow directive representations may be itself disjunctive, with different kinds of purposes being served on different occasions of use – where for our purposes, the relevant purposes can be understood quite broadly, to even include simply avoiding sanctions. Additionally, what producers get out of having the directive representation followed may not be the same thing as what purpose of the consumer is served (1984, pp. 56-57). Nevertheless, again, it is the production of these fulfilment-conditions that historically has kept consumers responding to directives in standard ways, and producers producing directives in standard ways.

In Millikan’s account, the proper function of a descriptive representational device is to adapt consumers to the truth-condition (determined by the semantic mapping function) of the representation, in the performance of the interpreting device’s proper functions (1984, Chapter 6; 2005a, p. 172; 2004, Chapter 6). That is; the effect of a descriptive representation that accounts for why such representations have continued to proliferate, is that they enable their consuming systems to fulfil their (i.e. the consuming systems') proper functions, where this requires sensitivity to the obtaining of certain states of affairs. So for example: A beaver's tail slap aids its consumers' (fellow beavers') purpose of avoiding danger only if often enough the slap corresponds with the presence of danger. Descriptive representations, then, only aid their interpreting mechanisms in accordance with a Normal explanation when they are true.[7] For instance, Normally, my belief that there is an umbrella in the hallway can only help fulfil my desire to stay dry while I go out in the rain if the umbrella is where the belief represents it as being. (Likewise, beaver slaps only aid fellow beavers in accordance with a Normal explanation when there is in fact danger in the vicinity). In language, sentences with an indicative form are standardly descriptive representations, with the function of producing true hearer beliefs. Speakers continue to produce true sentences, and hearers continue to form true beliefs in response to such sentences, only because doing so serves the ends of hearers and of speakers.

There is an important third category of representations: "Pushmi-Pullyu" representations (PPRs) (Millikan, 2004; 2005a, b; 2017). A PPR has, as its proper function, to "mediate the production of a certain kind of behavior such that it varies as a direct function of a certain variation in the environment, thus directly translating the shape of the environment into the shape of a certain kind of conforming action" (2005a, p. 173). Thus, for instance, a hen's food call to her chicks functions to mediate the behavior of the chicks to vary with the location of food in the environment. "[W]here the hen finds food, there the chick will go" (*ibid*.).

[7] Of course, on occasion, false beliefs may aid us (for instance, perhaps some instances of self-deception are beneficial to the one deceived). Though such beliefs may be beneficial, they do not aid the believers *in accordance with a Normal explanation*. For cases of self-deception do not contribute to the explanation of why beliefs continue to proliferate. (Compare: a broken clock is right twice a day; but the occasions where one learns what time it is from a broken clock do not figure in the explanation of the continued manufacturing of clocks).

What is distinctive of PPRs is that they are at once descriptive and directive. They both adapt interpreters to their truth-condition, and direct consumers to produce their satisfaction-condition. Because PPRs are simultaneously directive and descriptive, they simultaneously have two distinct satisfaction-conditions, and therefore PPRs need to at once be members of two different representational systems, one whose semantic mapping rules determines the descriptive content of the PPR, the other whose semantic mapping rules determine the directive content. (The result of this is that although PPRs are like ‘besires’ [see McDowell, 1978, p. 19; 1979, p. 346] in having two directions of fit, they differ from besires in that they have these fits with respect to distinct contents).[8] As Millikan puts it, “[t]he descriptives and directives would be written in different languages, languages that contained the same set of representation forms but with different meanings. (Compare: ‘1101’ means one thing in a decimal system and another in a binary system)” (forthcoming: 7). Each of the descriptive and directive function of a PPR is also a direct, literal function of the PPR: That is, neither is generated by some sort of Gricean implicature, and neither is given theoretical primacy over the other. (This point provides a basis for distinguishing my PPR proposal about ethical affirmations from other hybrid metaethical theories).

Despite having two directions of fit, Millikan proposes that PPRs are cognitively *less* sophisticated than corresponding pure directive and pure descriptive representations representing the same states of affairs. One reason for this is that having a purely descriptive representation requires the ability to store away information for which one has no immediate use – the ability to represent that which is not present in the here-now. And having a purely directive representation requires the ability to represent goals which one may not know how to accomplish – goals that

[8] Thanks to Robert Audi for alerting me to this important difference between PPRs and besires.

outstrip the abilities currently in one's repertoire (2005a, p. 175). Now, in sophisticated human practical reasoning, purely directive and purely descriptive representations can come together in a practical inference to motivate action. However, there may be reason to doubt that non-human animals can form pure descriptive and pure directive representations as required for practical reasoning. PPRs 'compress' the steps of a practical inference into less cognitively demanding states since a creature employing a PPR can use the very same representation as both a directive and a descriptive. This makes PPRs a plausible candidate for theorizing about forms of animal cognition that appear to share surface similarities with practical inference as it occurs in humans. Indeed, Millikan suggests that non-human animal cognition *exclusively* employs PPRs.

We should also expect human thought and language to include the category of PPRs, since there is no reason to think that the more primitive PPRs would have ceased to be useful once a capacity for pure descriptives and directives has come onto the scene. As an example of a PPR in human language, consider representations of social norms and roles, such as "we don't eat peas with our fingers", used to instruct a child on etiquette. Such a claim *at once* describes (how we in fact eat peas) and prescribes (how the child should eat peas) (Millikan 2005a, p. 179). Nevertheless, because humans can form purely descriptive and purely directive representations, we have the cognitive resources needed to 'dissect' these PPRs into their directive and descriptive elements, forming a purely descriptive representation with the same truth-condition as the dissected PPR, and a purely directive representation with the same fulfilment-condition as the dissected PPR (2005a, p. 178).[9] Humans, unlike most animals, are also capable of employing purely descriptive and purely directive representations.

[9] Future research on this topic will consider whether the human capacity to dissect PPRs into discrete descriptive/directive elements might play a role in enabling sophisticated moral inference operating on the descriptive element of ethical affirmation.

5.2.4 Pushmi-Pullyu Representations and metaethics

We are now in a position to articulate a difference between the cognitivist and non-cognitivist metaethical camps in biosemantic terms. A cognitivist will see ethical judgments as a kind of purely descriptive representation; ethical judgments function to adapt the deliberation of agents to the truth-condition of the judgment. Motivational externalists will add that such a judgment is then only contingently related to motivation and subsequent action. Some non-cognitivists will see ethical judgments as a kind of purely directive representation; ethical judgments do not have descriptive content, despite their declarative form, but instead represent how the world is to be made. And an interesting hybrid position is made available by the possibility of treating ethical judgment as a variety of Pushmi-Pullyu Representation.

If it is a function of ethical judgment to promote social coordination, it is clear how this might be explained by taking ethical judgments to be directive representations. The idea would be that ethical judgments have, historically, proliferated just because often enough they do result in compliant consumer behavior. However, I do not think that this treatment of ethical thought and discourse provides the whole story. For ethical judgments are unlike strict commands in that producing compliant consumer behavior, though perhaps necessary, is not sufficient to account for the continued proliferation of ethical judgment. Imagine that we began to only produce ‘ethical’ judgments that were routinely *false* – e.g., judgments such as “one ought to torture one’s family”, “one ought to cause as much unnecessary and undeserved misery as possible” – but where these judgments regularly resulted in compliant consumer behavior. I suspect that were this to occur, these sorts of judgments would eventually cease to be produced (even though they perfectly result in compliant behavior), precisely because they do not track the relevant facts –

the sorts of facts that make it the case that it is morally wrong to cause unnecessary and undeserved pain and suffering, or to torture one's family.

Accordingly, it seems to me that the ethical judgments we humans have made in the past that account for their continued proliferation had the job of tracking certain right- and wrong-making features, in aid of the social coordinating function these judgments perform. Note that this is not intended as an *a priori* conceptual claim about the nature of ethical judgment, but an empirical hypothesis about what in fact keeps us producing them, supported by the range of *actual* historical agreement in ethical judgment. A broad enough range of agreement with others in one's application of ethical terms provides some defeasible evidence that our judgments "are not hunting chimeras but are focused on real-world elements" (Millikan, 2017, p. 81). This point does, however, raise the issue of just how much agreement we really observe in ethical judgment. I take this issue up in Chapter 6. But to anticipate part of that discussion: while I grant the existence of intractable ethical disagreements, there does not seem to be much dispute concerning, for instance, whether the fact that an action causes unnecessary and undeserved misery is a reason not to perform that act. The range of agreement in our moral perspectives is sometimes overlooked, I think, owing to a tendency to devote attention to controversial moral cases. (Still, I shall argue in Chapter 6 that there is a remaining skeptical problem of *deep* moral disagreement). If it is correct to think that ethical judgments are supposed to track real features of the world of some kind, in aid of promoting a certain kind of social coordination, then it seems reasonable to think of ethical claims and judgments as having a descriptive as well as a directive function. In the next section, I develop just such a hybrid proposal according to which ethical judgments are PPRs.

5.3. The stabilizing proper function of ethical judgment

The main proposal I wish to make about ethical affirmation – call it the *Stable Coordination View* – is that it is a distal proper function of ethical affirmation to produce behavior in members of the moral community that coordinates in a limited range of collectively beneficial ways around the morally salient features of social situations members of the moral community find themselves in.[10] A proximal function of a given token ethical affirmation M will be to produce some behavior B in a consumer C that maps onto features F of the situation in such a way as to be collectively beneficial to the moral community. This proposal captures the intuition that ethical thought and discourse is essentially action-guiding. The proposal is also fairly minimal, and can be further developed in various ways depending on one's views about the membership of the moral community, what counts as a collectively beneficial arrangement, what the morally salient features are (i.e. the right/wrong-making features of actions), and exactly what mapping (i.e. what semantic mapping function) from morally salient features to behavior ethical affirmations are supposed to produce. We can state the basic schematic view, to be filled in in various ways, as follows:

5.3.1 The stable coordination view

> *The Proper Function of Ethical Affirmation* (PFEA): to mediate the behavior of consumers so that they vary as a direct function of variations in the morally salient features of a relevant situation.

[10] By 'member of the moral community', I here mean just 'a potential producer or consumer of ethical affirmation, or appropriate target of moral concern'. Many may judge that only mature humans may be producers and consumers of ethical affirmation. I take no official stand on this issue. It may be that certain sophisticated non-human animals are capable of producing and consuming primitive ethical affirmations. I take 'member of the moral community' also to include appropriate targets of moral concern, to account for the intelligibility of ethical thought and discourse concerning obligations to moral patients that are not also moral agents nor (if these categories differ in extension) potential producers or consumers of ethical affirmations (e.g. sentient but cognitively unsophisticated creatures).

Recall that this is intended as a proposal about the proper function, or purpose, of ethical affirmations invoking "ought" in general. It is not intended to supply a semantic analysis of the contents of specific ethical affirmations. Such an analysis would involve stating exactly how transformations in the articulate elements of the ethical affirmation are supposed to map onto exactly which variations in behavior that the consumer is supposed to produce. As indicated earlier, different normative theories will make different predictions about the semantic mapping function of ethical affirmations. Though I do not make a commitment to a particular first-order moral theory, there are two important constraints to consider on such theorizing in the biosemantic framework. First: whatever the morally salient features are, they must have a place in a broadly naturalistic worldview, given biosemantics' naturalistic commitments. And second: we must be able to provide an explanation of why it would have been beneficial for consumers of ethical affirmations to have their behavior guided by those features in a certain way. If the second constraint is not met, we lack an explanation for the continued proliferation of ethical judgment – the crucial ingredient in a biosemantic analysis. In this sense, any proposed first-order normative theory must 'earn its keep' to be considered adequate on the basic PPR proposal above.

It is important to note that consumers themselves need not act in accordance with ethical judgments because they *take* or *understand* doing so to be somehow beneficial for them, and nor do consumers themselves need to understand just what features their ethical affirmations are supposed to track. Acting in accordance with the judgments must *in fact* have aided consumers (and producers) in the performance of their proper functions, but this benefit need not itself be understood by the consumer (or producer) as such. Indeed, it is essential to the PPR proposal that there is no necessary mediation via practical inference from the recognition of some morally salient feature of a situation to corresponding action. Rather, the recognition of the morally salient feature activates the corresponding disposition to act *directly*. When operating properly, it is the registration of the right/wrong-making features of an action that is supposed to get one to/not to perform that action. This explains how the Stable Coordination View can avoid charges of moral ‘fetishism’ (Smith, 1994): one is supposed to wade in to the pond to save the drowning child because one sees that otherwise she would die or be seriously hurt – not because one first desires to do what is right, whatever that is, and then forms the belief that saving the drowning child is right, and then proceeds to save the child as a result of drawing a practical inference. PPRs at once tell what is the case and what to do about it, without room for an intervening inference that, in the case of basic moral affirmations, would involve ‘one thought too many’.

5.3.2 Justification for the stable coordination view

The Stable Coordination View just described, according to which ethical affirmations have the distal function of getting consumers to be directly guided in their actions by the right/wrong-making features of actions available in their situation, supports thinking of ethical affirmations as PPRs. I discuss the motivation for the PPR proposal in more detail in this section, by indicating why I think ethical affirmations do not have a purely directive nor a purely descriptive function, but must be simultaneously directive and descriptive.

There are many ways to secure social coordination, not all of which fall into the ethical domain. Many arrangements that enable social coordination are arbitrary to a greater degree than we expect to find in moral matters. For instance, drivers in the U.S. could have, but did not, establish a convention of driving on the left side of the road. If one thought that ethical judgment played a social coordinating role in the same way as conventions about driving, one would have to hold that what counts as correct ethical behavior can vary greatly across societies. Relativism would soon appear to follow.[11]

However, I would suggest that ethics admits of less variability than more ‘purely’ conventional behavior, and thus provides a more stable, universal basis for coordinating action. Ethical affirmation achieves social coordination in part by anchoring our ways of acting around certain features of situations we might find ourselves in, features which it is beneficial for us to collectively respond to in only a limited range of ways. This may help to explain the intuition that moral obligations transcend legal institutions (which involve a greater degree of arbitrariness), as well as the intuition that morality is in some way objective. Still, ‘objectivity’ may come in degrees. Even if there is widespread agreement across cultures about, say, the value of human life, that stance itself may manifest in quite different specific moral codes. The idea here is that the conformity of action to the content of directive representations widely shared in a community, while possibly necessary for moral social coordination, is not sufficient; the actions themselves must also map on to the morally salient features of the situation.[12] Ethical judgments are descriptive representations (in addition to being directive), because they have as part of their proper function adapting consumers to their truth-conditions.

Representational devices that have a descriptive mapping function are supposed to answer to real variations in the world. What explains the proliferation of devices that descriptively represent is that often enough they do map onto world affairs as they are supposed

[11] Recently, Millikan herself has speculatively suggested that moral truth might be relativistic for the sort of reason discussed above; the effects that account for the proliferation of moral sentences may vary across cultures (2018, pp. 240-241).

[12] It might be complained: what role does the *success* of such tracking play here? This sort of question is pressed by evolutionary debunking arguments, exemplified in Street (2006) and Joyce (2006). I respond to evolutionary debunking arguments in Chapter 6..

to, thereby furthering the ends of representational systems that employ them to navigate their world. Thus, in contending that ethical judgments have a descriptive function, my approach here is committed to there being real variations in the world that ethical judgments are supposed to track. Moreover, Millikan's overall framework is thoroughly naturalistic, and so whatever such real variations there are will need to be *natural* variations in the world. This may suggest that my approach is committed to naturalistic moral realism. While I am sympathetic to naturalistic realism, I do not think this is a necessary commitment of the Stable Coordination View. This is because the Stable Coordination View is not committed to any particular answer to the question of whether the features of social situations that ethical affirmations are supposed to track are themselves *moral* properties, or instead naturalistic but non-moral properties. Indeed, it is open to the Stable Coordination View to say that what qualifies ethical affirmations as ethical is not that they represent moral properties, but that they connect morally salient non-moral properties to action in the way characteristic of moral practice.

At any rate, the naturalistic constraint does not impose an overly heavy burden for a biosemantic theory of ethical affirmation, even if the view is developed as a version of moral realism. This is especially so because, given Millikan's radical semantic externalism, considerations that are often taken to speak against naturalist moral realism (such as the Moral Twin Earth argument) have no application to the biosemantic approach.[13] If social coordination

[13] Biosemantics appears a useful tool for moral realists concerned to address Moorean Open Question considerations through appeal to semantic externalism, as bioesemantics is arguably a more thoroughly externalist account even than the causal-historical picture espoused by Putnam, Kripke and Burge. According to Millikan, "The stabilizing function, the current meaning, of a word rests on what has, as a contingent matter of fact, been holding its usage in place, effecting agreement among users and for users with themselves, despite the use of a variety of alternative recognition techniques" (2010, p. 65). Thus, it is not even the case that all competent users of a natural kind term such as 'water' must share some methods for identifying water – instead, it must just be the case that the properties and distribution of water have been constant enough in our history to afford agreement in our judgments. Millikan's view is radically externalist in that it denies that empirical terms have *any* defining intensions whatsoever. It is for this reason that Millikan rejects intensions as probative for meaning (see e.g., Millikan, 2010).

is a function of ethical judgment, the actual ethical judgments of moral agents must frequently enough track the *same* features, but judgers themselves need not have a clear or unified conception of what these features are. Additionally, while biosemantics carries naturalistic commitments, the view is more permissive in its ontology than one might expect from a strict naturalism, including categories such as “dances, books and musical pieces, diseases, clubs, ceremonies, countries, grocery stores, monetary systems, and so forth” (Millikan, 2017, p. 26). Given this, it at least seems open to hold that there is a place in a Millikanian natural world for real categories of rights, virtues, and the like. (It must be admitted, however, that appeal to the permissiveness of Millikan’s ontology can at best provide only some defeasible license for optimism for the viability of a naturalistic moral realism, by comparison with more stringent naturalistic requirements, as reflected in eliminativist or reductionist approaches to social-cultural categories of being.)

There is another important sense in which the naturalistic constraint here is not overly difficult to meet on the biosemantic approach. It is often assumed that, because biosemantics takes a biological model for thinking about representation in terms of selection pressures,

Given this radical semantic externalism, Descriptivist metaethical views modelled on Millikan’s biosemantics are seemingly immune not only from Moorean Open Question considerations, but also from Horgan and Timmons’ well-known Moral Twin Earth argument (Horgan and Timmons, 1991), as is compellingly argued in Dowell (2016). Putnam’s original insight about ‘water’ turns on the intuition that speakers of Earth English and speakers of Twin Earth English (where what Twin Earthlings call “water” is causally-historically related to XYZ, not H2O) talk past each other in their judgments involving the sign-design “water”. Horgan and Timmons contend that this is not the case for moral predicates such as “good”: We are supposed to have the intuition that Earthlings and Twin Earthlings do disagree (and do not just talk past each other) when an Earthling affirms “X is good” and a Twin Earthling affirms “X is not good”. The point here is that Putnam’s original argument, and Horgan and Timmons’ moral variation, both take linguistic intuitions about disagreement as relevant to understanding semantic function. But this is exactly what is denied by Millikan’s brand of content externalism. As Millikan puts it: “Unlike Putnam in ‘The Meaning of ‘Meaning’’, I cannot use the armchair method of example and counter-example, calling on our a priori intuitions about what is or what could be in the extension of this term or that, for that method assumes exactly what I am aim to disprove” (2010: 52). By the same token, intuitions about what is in the extension of a term cannot be exclusively relied upon to undermine a Millikan-inspired moral Descriptivism, as argued in Dowell (2016). In short, Millikan’s biosemantics has a potential advantage over competing externalist theories of content when it comes to articulating a plausible Descriptivist account of ethical thought and discourse. This is just one of the interesting ramifications of applying Millikan’s biosemantics to the moral domain.

biosemantics in general, and a biosemantic metaethics by extension, must be in the business of explaining how the functions it describes would have arisen in evolution. There are two mistakes to avoid here. First, biosemantics is primarily concerned with describing what various representational devices have been doing in their recent history to keep themselves around, not their evolutionary history. As Millikan puts it, the job of the biosemanticist is "to figure out how these various mechanisms work, how they function. Explaining how they evolved is not part of their job" (forthcoming, p. 2). The biosemantic approach should not be understood to be committed to offering a theory of the evolution of moral thought and discourse.

Second, we should take care to avoid conflating the function of ethical thought and talk (the goal of this chapter) with the function of moral behavior. An account of the function of moral behavior would explain why doing the morally right thing might have contributed to reproductive fitness. An account of the function of ethical thought and discourse should explain, instead, what ethical claims and judgments have been doing to keep themselves around. This does not require speculating on why natural selection would have favored creatures that behaved morally over those that did not. In fact, it could even turn out that in terms of natural selection, moral behavior is *mal*adaptive, and it could still be true that what continues to account for the proliferation of ethical affirmation is that it produces moral behavior. So long as ethical affirmations serve some purpose of the organisms that produce and consume them, even if not the strict biological purposes of those organisms, that purpose can be appealed to in explaining their proliferation. So, the biosemantic proposal does not need to provide an account of why moral behavior would be evolutionarily adaptive, though it may help to supplement it with one. I make some speculations about the evolution of moral behavior in this spirit just below, but it should be understood that this is not essential to the core proposal of the dissertation.

It seems plausible that social creatures capable of some degree of cooperative behavior would find it beneficial, evolutionarily, to be able at least to track pain in conspecifics, and to react in ways to mitigate such pain and its causes. At the very least, it is to be expected that animals that care for their young would be suited to detect behavior in their offspring that indicates pain, distress, etc., and to be moved to react to such states. As Tomasello (2016) notes, sympathy and helping are basic moral emotions and behavior. They represent the first rung on the evolutionary ladder to the full range of human morality, and are a kind of moral behavior we share with certain other social creatures. Emotions such as sympathy seem likely to be PPRs in thought that drive moral behavior; sympathy involves at once a recognition of some situational feature (pain/distress/injury in a community member) and directs appropriate action (comforting/aiding/alleviating/etc.) Correlatively, behavior designed to *express* one's pain to a suitably endowed receiver can be understood as a primitive sort of communicative ethical PPR. Wincing, groaning, moaning, and the like, on this proposal, at once express one's pain, and call for others to help.[14]

The basic idea is that moral emotions and their expressions themselves may constitute a sort of proto-ethical affirmation – 'proto' insofar as they lack the semantic articulation required to qualify as full ethical claims or judgments – but which can contribute to full ethical affirmations both epistemically and motivationally. What semantic articulation adds, in the case of ethical affirmation, includes (1) a common linguistic framework that allows for ethical affirmations to be put into conversation with each other, enabling deliberation over competing

[14] As Bar-On notes, animal expressive signals are 'Janus-faced' in that they at once point inward to the expressed mental state, and outward to the intentional object of that state (2019, fn. 32). It is worth considering whether such signals might also be Janus-faced in the PPR way, pointing to features of the environment and what is to be done about them. An expression of fear, for instance, may (i) point inward to the fear state, (ii) point towards the fearsome object, and (iii) direct others to flee (see Bar-On, 2022) for just such a multi-faceted proposal).

moral considerations (e.g., in moral dilemmas), and (2) the possibility for sophisticated moral reasoning, including explicit inference from moral premises to moral conclusions. Moral emotions, I am suggesting, have the same basic function as ethical affirmations – they identify features of moral concern and direct fitting responses to those features. But it is only when such emotions are given voice using expressive vehicles that are semantically articulate that we have a *full* (not merely proto-) ethical affirmation: a claim or judgment capable of being either true or false, of being negated or denied or agreed with, of being a piece of knowledge, of entering into moral inferences that can extend our moral knowledge, and so on.

To illustrate, consider *moral anger*. Arguably anger's function, as it is significant for ethical affirmation, is to enable individuals to identify injustice and to motivate action to hold those responsible accountable. This contributes to the distal function of ethical affirmation, insofar as the threat of being held accountable for committing an injustice deters such acts. On this view, then, anger has both an *epistemic* function (identifying injustice) and an action-guiding function (accountability). Expressions of anger can come in varying degrees of semantic articulation, and not all such expressions will constitute full ethical affirmations. Anger might be expressed through an inarticulate yell; or through rude gestures; or through avowals (as in "I'm angry that you did that"); or, in a full ethical affirmation, through an identification of the wrong as such, as in "it was wrong of you to treat me that way".

If certain social species employ a kind of primitive ethical PPR, we should also expect to find similar PPRs in human moral thought and discourse.[15] Indeed, it seems that something like this primitive, cognitively undemanding sort of ethical PPR just described is responsible for the

[15] This depends, of course, on the possibility of extending the current proposal of this paper beyond ethical affirmations explicitly involving the concept 'ought'. The proto-ethical affirmations I am speculating about would likely lack significant linguistic articulation, and cannot be understood to explicitly invoke 'ought' thoughts.

immediacy with which we react in situations calling for an urgent moral response: one sees the child drowning in the pond, and one does not need to think about it - the thing to do is to jump in and save the child. One's behavior is to be guided directly by the perceived need for help. As noted earlier, a positive feature of the PPR proposal is that it makes the right predictions about what should provide moral reason to act. We act directly on behalf of the needs of others, rather than having to carry out a practical inference from an antecedent desire to do what is right, whatever that is, and the antecedent belief that saving the child is the right action.

Regarding the directive function of ethical judgment: as already emphasized, it seems that at least part of the proper function of ethical judgment is its effects on the behavior of the consumers of the judgment. In accordance with Millikan's radical semantic externalism, this function of ethical judgment is not simply intended to reflect intuitions – part of the 'data' – concerning ethical thought and discourse. Nor is it a claim about the *concept* of ethical judgment, such that any possible community that made judgments that had the relevant effect on consumers would necessarily count as making ethical judgments. Instead, given the reference to *history* in the notion of proper function, this should be seen as a claim about actual, contingent, historical *fact*.[16] Items belonging to the category of thought and discourse we call ethical judgment have proliferated, I claim, because they do get consumers to behave in (close enough to) the right way enough of the time. Our ethical claims and judgments are tools for moral progress.

5.3.3 First-personal ethical judgment and public ethical claims

[16] See Millikan (2010) for a discussion of the relevance of history through a discussion of Davidson's 'Swampman' thought experiment. Some find it intuitive that 'Swampman' – an exact replica of Donald Davidson created by pure coincidence at the exact moment Davidson himself is disintegrated – has intentional states. But biosemantics must deny this, for there is no historical explanation of the right kind of Swampman's similarity to Davidson. As I interpret Millikan (2010), the use of intuitions about Swampman's 'mental states' as an argument against biosemantics begs the question, since, if biosemantics is correct, what determines meaning is precisely not speaker intuitions/dispositions, but 'local, natural, this-world history', and so the Swampman case "removes the natural planks on which our terms for mental events are resting" (2010, p. 77).

Thus far, I have discussed the notion of ethical *affirmation*, using this to cover both public ethical claims, and inner ethical judgment. In this section I discuss the distinction between ethical affirmations in public discourse and inner thought. While I conceive of each as PPRs, it is necessary to add a slight complication here, as public ethical claims and self-directed ethical judgments have slightly different adapted functions, owing to the fact that they are directed at different consumers.[17]

I propose that a properly functioning public ethical claim (such as "You ought to keep your promise to Joe") guides the behavior of the hearer, through modifying/reinforcing hearer intention/disposition to act. A properly functioning inner, self-directed ethical judgment (such as "I ought to be more forgiving") either brings about, activates, or reinforces a behavior, intention or disposition to act in the judger. The former – the function of public ethical claims – captures the intuition that public moral discourse is *action-guiding*; ethical claims are to direct the behavior of others. And the latter – the function of inner ethical judgment – explains the connection between self-directed moral judgment and *motivation*, for when functioning properly, a self-directed moral judgment modifies one's own behavior by first modifying one's intentions and dispositions to act. I discuss each of these in more detail below, starting with self-directed ethical judgment in thought.

In the case of self-directed ethical judgment, we should expect there to be psychological mechanisms of some kind that have the job of directly linking the recognition of a morally salient feature to subsequent action. I propose that the relevant psychological mechanisms involve the activation of affective, motivationally-charged pro- and con-attitudes, such as moral commendation and condemnation, reactive attitudes like guilt, shame, indignation, and basic

[17] See Millikan (1984, p. 40) on the notion of 'adapted function'.

prosocial emotions such as sympathy. I am now in a position to highlight an attractive feature of the PPR proposal, arising from the proposed psychological mechanisms just discussed – namely, the fact that it yields a version of motivational internalism able to accommodate a surprising range of purported counterexamples to more conventional versions of the internalist thesis.[18]

I've proposed that ethical judgments accomplish their proper function by activating motivationally-charged affective states. If this is right, then there is a necessary connection between *properly functioning* ethical judgment and motivation, since the presence of motivationally-charged states is a Normal condition on ethical judgment. Recall that the Normal conditions on the performance of a device's proper function are not necessarily statistically average conditions and indeed may be quite rare. So, even if it is a necessary condition on the *proper* functioning of an ethical judgment that there be a corresponding motivationally-charged affective state, it is possible that only a very small proportion of *actual* ethical judgments are accompanied by that attitude. Since the normativity of the 'supposed to' here is modelled on biological teleology, rather than *moral* normativity, the view avoids making the vacuous claim that people are motivated to act as they morally ought when they are morally virtuous (see 4.9). Instead, the view proposes that people are motivated to act as they morally ought when their ethical judgments function properly in accordance with a Normal explanation.

In making *other-directed* public ethical claims (e.g., an assertion of "you ought to keep your promises"), I propose that a central mechanism that drives compliant hearer behavior is the expressive communication and transfer of the relevant motivationally-charged attitude from speaker to hearer, through the vehicle of the ethical claim. Such expressive behavior allows suitably placed receivers to at once see one's motivationally-charged affective attitude, and

[18] See also Bedke (2009), whose moral judgment purposivism is very much in the spirit of this proposal, and also draws inspiration from Millikan's biosemantics in articulating it. See also 4.9.

directs receivers' attention to the features of a situation one's attitude is directed at, thereby eliciting an appropriate reaction by activating a corresponding attitude in receivers. The idea, then, is that our moral concern is directed not just by recognizing the contents of others' ethical claims, but also by seeing various features of the affective states others express in the act of making those claims. As emotivists recognized, and as more recent psychological research supports, the expression of affective states is infectious;[19] this may well be an important mechanism for bringing about action that coordinates in the right way around features of moral concern.

5.4 Moral reasoning

In Section 5.3.2, I briefly considered one way in which human moral cognition goes beyond the proto-moral judgments of other social animals; namely, that certain features of human rationality and sociality may ground special moral categories like rights or virtue which it would then be the job of human ethical judgments to track. In this section, I discuss one other way in which I see moral cognition in humans as more sophisticated than the primitive sort of proto-moral cognition of other social animals.

Humans have the capacity to form, in addition to PPRs, *purely* directive and *purely* descriptive representations. When it comes to ethical judgment, this capacity enables us to 'dissect' ethical judgments into their descriptive and directive elements, and to employ these elements in moral deliberation.[20] That we can cognize the truth-conditions of ethical judgments

[19] See Stevenson (1937), who emphasizes that, concerning ethical statements, "[t]heir major use is not to indicate facts, but to *create an influence*" (p. 18). See also the more recent psychological literature on the phenomenon of *emotional contagion* (see, e.g., Dezecache, Eskenazi, and Grèzes (2016); Doherty (1997). The notion of emotional contagion is of special relevance in the present context given that it also figures in some discussions of the evolution of morality (e.g., de Waal, 2012)).

[20] Millikan notes that although human thought and language contain PPRs, there are also "more sophisticated mechanisms by which we moderns may also dissect the relevant [social-coordinating] norms to reveal two faces", the 'two faces' being the two aspects of PPRs, one with a world-to-mind direction of fit, the other with a mind-to-world fit (Millikan, 2005a, p. 178).

as purely descriptive representations (rather than as PPRs) explains how we are able to manipulate such representations in thought in such a way as to allow sophisticated reasoning and deliberation about moral matters.

This capacity to consider the truth-conditional content of ethical claims, for instance, is what enables us to reason from moral premises to conclusion, as in: reasoning from 1 "If lying is wrong, getting your little brother to lie is wrong" and 2 "Lying is wrong" to the conclusion 3 "Getting your little brother to lie is wrong". We can consider the content semantically expressed by premise 2 without making an ethical judgment – this is a precondition on being able to reflectively appreciate the logical relations that hold between the content of that premise and the content s-expressed by premise 1 and the conclusion. A creature capable only of PP representations could not grasp such an inference, since it would not be able to countenance the representational contents involved in the premises and conclusion in isolation from any possibilities for action in the here-now. At the same time, in accordance with my treatment of the Wishful Thinking problem in 4.3, the purpose of *moral* inference, qua mental act (rather than logical relation between propositional contents), is to deliberate on moral matters so as to form an ethical judgment. In sum: the capacity to consider propositional contents in isolation from acts of judging them true/false allows us to see the logical relationships necessary to rationally draw moral inferences, and to thereby form new ethical judgments.

This suggests that a crucial component to moral reasoning is the capacity to manipulate representational content in thought independently of the uses to which such content might be put (e.g. in a directive, descriptive, or PP representation). This, in effect, requires a separation between force and content. In Millikanian terms, this separation amounts to the distinction between the proper function of a representational device, and the significant transformations that device affords – its semantic mapping function.

Individual sentence-elements, taken in isolation from any particular sentence in which they might appear, lack a direct proper function, and have only a *relational* proper function, in Millikan's account - token instances of a word acquire a direct proper function only when adapted to the sentence in which it appears. So too, sentences themselves have relational proper functions with respect to more complex constructions in which they appear as elements. Thus, the stabilizing proper function of a language device S is different from the derived proper function tokens of S have when produced in, say, an instance of "Not S", though there is a systematic connection between the two. The way in which negative representations (e.g., "Not S"), for instance, map onto the world is a function of how their corresponding positive representations map: in particular, where a positive representation R maps a state of affairs T, a negative representation *not R* maps onto a *positive* state of affairs T' incompatible with with T (1984, Chapter 14). This applies equally for directive and for descriptive representations: "There is no mud on the rug" maps a positive state of affairs – the rug being a certain way (say, *clean*) – that is incompatible with there being mud on it. So too, "Don't track mud on the rug" maps a state of affairs wherein the addressee positively intends to do something incompatible with tracking mud on the rug (say, intending to take his boots off before entering) (1984, pp. 224-228). Given a similar analysis for other logical operators (such as 'if . . . then . . .'), the takeaway here is that there can be no special Frege-Geach problem for a biosemantic metaethics. So long as ethical judgments have some propositional content, these representations will afford significant transformations in devices containing them under logical operators, in just the same way as with ordinary non-ethical representational devices. But in order for a creature to be a moral *reasoner*, it must be capable of appreciating the logical connections between the contents of ethical judgments and other semantic contents. And this is something that a creature capable only of PP representation likely cannot do. So, moral reasoning appears to require at least the capacity to form purely descriptive and purely directive representations. At the same time, given that other animals can form *proto*-ethical judgments, we see a continuity between human and non-human moral cognition, pointing the way to an account of how the distinctively human moral capacity might have originated in evolution.

5.5 Conclusion

The biosemantic account given in this chapter supplements HENE and the hybrid account of ethical assertion presented in Chapter 4 by situating those views within a general account of representation and cognition that explains why ethical judgment would have the features those views propose. This biosemantic account, though radically at odds with standard approaches to philosophy of language and mind, especially as these are applied in metaethics, is independently plausible and supports a nuanced treatment of ethical judgment and cognition. Although this biosemantic account of ethical judgment warrants further development and defense, I hope to have made the possibility and attractions of such a view evident in this chapter.

One last issue I would like to briefly consider concerns how to locate the PPR proposal about ethical judgment in relation to other metaethical theories. The biosemantic view of ethical affirmation in this paper is like existing hybrid theories in proposing that ethical affirmations have both a descriptive and a directive function. But it is unclear how objections to other hybrid theories might be extended to the Stable Coordination View. Millikan explicitly denies that PPRs incorporate any kind of implicature. The descriptive and directive functions of a PPR are each literal, with neither taking conceptual priority over the other (2005a, p. 179). Thus, at the very least, the initial difficulties confronting competing implicature approaches in the hybrid literature do not apply to the PPR proposal. Additionally, Schroeder's worries, if one attempts to carry them over to the biosemantic framework, seem to conflate proper function with semantic mapping function. Schroeder assumes that the descriptive and expressive aspects of ethical affirmation that the hybrid theorist posits must figure at some level in the semantic contents of such affirmations (though this assumption is understandable, given that Schroeder's targets generally endorse a mentalist approach to semantics or metasemantics). However, the biosemantic approach taken distinguishes the analysis of the semantic content of an utterance from the analysis of that utterance's purpose, or the type of act it is used to perform.[21] "Stealing is wrong", when accepted as the conclusion of a moral argument, is not supposed to motivate because of some semantic feature of the sentence or of the sentences stating the premises of the argument; rather, it is supposed to motivate because it is produced by mechanisms whose job it is to produce representations that guide the behavior of their consumers to vary according to the morally salient features of some situation.

This concludes the main exposition of my positive views on the function of ethical thought and discourse, consisting of a hybrid version of ethical neo-expressivism, a theory of the speech act of ethical assertion, and the stable coordination view about the function of ethical affirmations. In the next chapter, I turn to two remaining matters on the agenda set out in chapter 1: A discussion of moral disagreement, and a defense of the possibility of objective moral knowledge.

[21] A similar distinction can be found in Bar-On's neo-expressivism, which distinguishes between expression in the 'semantic' sense, a two-place relation between a significant linguistic item and its content, and expression in the 'act' sense, a three-place relation between a minded creature, a mental state, and an expressive vehicle, as discussed in Chapter 4 (see Bar-On 2004, p. 216). I see (hybrid) ethical neo-expressivism and the biosemantic proposal in this chapter as complementary.

Chapter 6

Moral Knowledge

6.1 Introduction

Let us take stock. In Chapter 1, I discussed the features of ethical thought and discourse that an adequate metaethical theory should either explain or explain away. These features are: *truth-aptness*, *embeddability*, and *objectivity* – features shared with other cognitivist domains – and the features distinctive of ethical thought and discourse: the *connection to motivation*, *action-guiding* feature and, possibly, *disagreement*. In Chapter 2, I considered traditional expressivist accounts of ethical thought and discourse. While expressivist theories are well-poised to explain the distinctive features of the ethical domain, they have difficulty accounting for the semantic, logical, and epistemic continuities between ethics and other domains. I suggested that ultimately, traditional expressivism is unable to meet this challenge because it fails to adequately capture the *epistemic* continuities between ethics and other domains, especially concerning the possibility of fundamental moral error. In Chapter 3, I considered various hybrid theories designed from the outset to account for both the cognitivist appearances and the distinctive features of the ethical domain. I argued that existing hybrid theories have implausible semantic, metasemantic, or pragmatic commitments. In Chapter 4, I considered ethical neo-expressivism, and argued that this view improves upon the theories considered so far in the dissertation. I then argued that ethical neo-expressivism still does not adequately capture the epistemic continuities between ethics and other domains.

In the second half of the dissertation, I made some positive proposals of my own. First, I argued that ethical neo-expressivism can be easily modified into a genuinely hybrid view that avoids the difficulty I presented for the non-hybrid version. The resulting hybrid ethical neo-expressivism represents a theory of ethical thought and discourse that, despite making only relatively minimal commitments about the nature of ethical judgment and the analysis of ethical sentences, accounts for many of the cognitivist appearances and distinctive features of the ethical domain better than any competing theory considered so far. I then further developed this proposal into a richer meta-ethical account that gives a more substantive explanation of the nature of ethical judgment and of ethical claims. In Chapter 4, I did this by developing an account of ethical *assertion* to supplement the basic hybrid ethical neo-expressivist proposal. And in Chapter 5, I further supplemented this proposal with a biosemantic theory that provides a framework which justifies the 'cognitive' and 'expressive' elements of ethical judgment in terms of proper function.

Thus far, I have discussed how my overall proposal explains:

- the close connection between ethical judgment and motivation to act,
- the action-guiding nature of public ethical claims,
- the truth-evaluability and embeddability of ethical sentences, along with other semantic and logical continuities, and
- how ethical judgments can enter into relations of evidential support and inference.

Of the features of ethical thought and discourse identified in Chapter 1, this still leaves the possibility of moral knowledge, and the extent and depth of moral disagreement, to be discussed. I take these issues up in this chapter. As we shall see, there is an apparent tension between them; seemingly, the extent and depth of moral disagreement provides reason to think that we cannot have rationally grounded moral knowledge.

In this chapter, I shall argue, first, that the account I have developed over the past two chapters disarms some of the most prominent skeptical challenges to the possibility of moral knowledge. I then, second, consider a remaining skeptical worry generated by the extent and depth of moral disagreement. In order to address this worry, I defend a radical but I think compelling epistemological theory – a version of 'hinge' epistemology – that explains why rationally irresolvable disagreements do not have the skeptical import they are often taken to have. With this account in hand, I shall be able to respond to the argument from disagreement against moral knowledge.

6.2 Skeptical challenges to moral knowledge

In 6.2.1-6.2.3 I articulate several skeptical arguments, and variously argue that they either fail on their own terms, or do not apply specifically to the proposal of this dissertation. In 6.3, I articulate what I take to be the most serious skeptical challenge for my own view, having to do with the skeptical import of the extent and depth of moral disagreement. Fully responding to this argument will require the introduction of theoretical resources that go beyond what is available in the hybrid proposal defended so far.

6.2.1 The argument from internalism

One of the distinctive features of the ethical domain, which it has been a concern of this dissertation to qualify and explain, is that there is an apparent close connection between ethical judgment and motivation to act. Proponents of motivational internalism can argue that there is a simple explanation of this apparent connection, for, according to the internalist idea, motivation is 'internal' to ethical judgment, such that one could not genuinely and rationally judge that something is wrong without being at least somewhat motivated to avoid doing it. In Chapters 4 and 5, I have defended a nonstandard version of the motivational internalist thesis, one that accommodates a wide range of actual failures of moral motivation. However, some take internalism to support an argument against moral realism (see also 1.4.1 for discussion of this argument). Since I intend my view to be compatible with moral realism, if such an argument succeeds, it would show an inconsistency in my view.

Perhaps the most well-known articulation of this challenge is found in Mackie's (1977) (problematically titled) 'argument from queerness', discussed in section 1.4.1. To briefly reiterate, here is the challenge and why I think it is not a serious problem. Mackie assumes from the outset that a version of motivational internalism holds as a matter of conceptual truth. He then contends that internalism could only be true if moral properties somehow had 'to-be-pursuedness' built into them. But, he continues, any such property would be very strange, and if such things existed at all, we could only have knowledge of them through some special faculty of moral perception. Since there does not appear to be any good evidence that we have such a faculty, it seems we are forced to conclude either that there are no moral properties (and so moral realism is false), or that we cannot have moral knowledge (for we have no way to track the moral properties). So we must reject non-skeptical moral realism.

I do not find this argument very persuasive because it relies on the dubious assumption that if motivational internalism is true, this could only be because of some feature of moral *properties*. Mackie assumes that the motivational function of ethical judgment would have to be built into the *metaphysics* of moral properties. Instead, I have argued that it is part of the proper function of inner ethical judgments to bring about, reinforce, or activate a certain motivational state in the judger; ethical judgments continue to be reproduced because they have had this effect in the past. There is no need to think that there is something inadmissibly strange about moral properties in order to explain this. Hybrid ethical neo-expressivism also explains the connection between ethical claims and motivation in terms of the features of *acts* of making ethical claims. Again, we do not need to invoke some special feature of moral properties to explain this feature of acts of making ethical claims.

6.2.2 The argument from no cognitive contact

When it comes to moral knowledge, those who think that there is something peculiar about moral properties have a special challenge. Non-naturalists hold that moral properties are fundamentally unlike naturalistic facts in that, among other things, they do not enter into causal relationships. As a result of this commitment, non-naturalists face the difficulty of explaining how we could ever get to know about such properties. For if moral properties are 'causally inert', they apparently cannot get into the right sort of causal relationships with our cognitive systems to enable knowledge.[1] One might posit a special faculty dedicated to detecting moral properties, but the existence of such a faculty seems to be no less mysterious than the existence of peculiar moral properties themselves.

Now, it is not my concern to respond to the argument from *no* cognitive contact, since that argument only arises given non-naturalist metaphysical commitments which I reject. However, there is a similar argument in the vicinity that *does* appear to apply to naturalist moral realism. A crucial insight of the argument from no cognitive contact is that for ordinary, causally-efficacious matters of fact, knowledge of the relevant facts seems to require that one's cognitive facilities be causally related to the facts in the right way. Indeed, one way to explain the intuition that Gettier cases are not true cases of knowledge is that the subjects' justified true

[1] But non-naturalists do have available a 'companions-in-guilt' type argument; for it seems that *mathematical* truths are also acausal, yet we can have knowledge of them. I am uncertain how powerful this argument really is, especially since the same considerations seem to generate a skeptical puzzle in the case of mathematics, too. Establishing that moral knowledge is no more mysterious than mathematical knowledge is not very helpful if they both remain mysterious.

belief in those cases are not causally connected in the right way to the facts they concern; the beliefs are only *accidentally* true. This point can be leveraged even against naturalistic moral realism, if we can legitimately question whether moral properties are connected to our beliefs about them in the sort of way conducive to knowledge. I turn to such an argument in the next section.

6.2.3 The evolutionary debunking argument

Evolutionary debunking arguments aim to provide evidence against non-skeptical moral realism.[2] The evolutionary debunker argues that evolutionary forces have significantly influenced moral belief in ways conducive to reproductive success rather than to discovering moral truths, and that this fact undercuts warrant for moral belief, leading to a skeptical result (given realist assumptions).[3] In what follows, I summarize the core debunking argument, and then argue that we should reject a crucial premise in the argument linking moral belief contents to natural selection pressures.

Crucial to any evolutionary debunking argument is the reliance upon principles connecting processes of natural selection operative in the situations in which *homo sapiens sapiens* evolved to the contents of widely held moral beliefs. If evolutionary forces are to be said to ‘debunk’ non-skeptical realism, such forces must have a relatively strong influence upon moral belief contents. But debunking arguments have understandably avoided positing a direct connection; for, at the very least, while it may be plausibly supposed that natural selection can

[2] Versions of debunking arguments can be found in Bedke (2009); Greene (2008); Joyce (2001b, 2006); Kitcher (2005); Ruse and Wilson (1986); and Street (2006). See Vavova (2015) for an overview.

[3] Certain moral anti-realist views appear not to face this skeptical worry, for they can insist it is no coincidence that the evolutionary forces operative on moral belief contents would have selected for moral truth – for moral truth, on these versions of anti-realism, is constructed out of our moral attitudes, not discerned by them (Street 2006). Still, one might worry that the strongest debunking arguments pose a challenge not just to moral realism, but to any view on which morality is about ‘more than spreading our seed’ (Vavova 2021, p. 721). If so, moral anti-skeptics have all the more reason to address the debunker’s challenge.

explain the proliferation of certain *traits* within a population (including psychological dispositions and capacities), it seems far less plausible that natural selection could directly explain the proliferation of *beliefs* with specific contents. For now, a vague statement of the connection premise will suffice:

Connection: Evolutionary forces indirectly influenced moral belief contents.

The intuitive idea behind this premise is that certain moral beliefs have widespread currency, not because they successfully track realist moral truth, but because possession of certain evaluative tendencies or psychological dispositions that tend to produce those beliefs contributed to greater reproductive fitness in the environments in which humans evolved. So, for instance, a debunking explanation for the widespread acceptance of the moral belief *that it is good to help others when doing so comes at no cost to oneself* might go as follows:[4] Social creatures that had psychological traits grounding a disposition towards altruistic behavior likely helped group members more often than individuals who lacked this trait; this helping behavior would have provided a greater benefit to the group overall compared to purely self-interested behavior, thereby increasing the groups chance of survival in situations of inter-group competition for resources; this would explain the proliferation of groups with a critical proportion of individuals with altruistic dispositions. Insofar as this altruistic psychological trait is responsible in individuals who possess it for the belief *that it is good to help others when it*

[4] This is, of course, assuming that such a belief is indeed widespread. Readers who disagree are encouraged to substitute whatever moral belief they think *is* widespread, and charitably construct a debunking explanation for that fact following the model provided in the main text. If one thought that *no* single moral belief has widespread currency, there is of course no fact to be explained here. But I suspect that a lack of consensus on *any* moral belief would be problematic for non-skeptical moral realists anyway. At any rate, I argue later in this chapter that the existence of deep moral disagreement does not pose a skeptical challenge for moral knowledge.

comes at no cost to oneself, and given that this trait was selected for, we seem to have an evolutionary explanation for widespread acceptance of this moral belief.[5]

To complete the argument, we need an explanation of why the moral beliefs that would have been (indirectly) selected for are unlikely to converge upon realist moral truth, in a way that undermines knowledge. Here is one such explanation. We have no assurance that natural selection processes would lead us in the direction of moral truth, which is metaphysically unrelated to fitness-promotion – at least if moral realism is true. A guarantee of fit between evolutionary influence and moral truth would be best explained by supposing that moral truth is somehow *constituted* by the evaluative tendencies or psychological dispositions that natural selection would have furnished us with. But this sort of mind-dependent constitutive connection is precisely what moral realism rejects. So moral realists seemingly must deny that there is a guaranteed fit, in which case the convergence of evolutionary influence and moral truth would be a coincidence, and so incapable of supporting moral knowledge (because luckily true belief fails to qualify as knowledge).[6] My interest is in evaluating the *Connection* premise itself, so I shall here grant that some such plausible explanation is available to the debunker. That is, I grant:

> *Coincidence*: If evolutionary forces indirectly influenced moral belief contents, then it would be a coincidence if our moral beliefs converged on the moral truth (as understood by the realist).

The explanation just surveyed also supplies an anti-luck epistemic principle according to which *Coincidence* together with *Connection* yields skeptical results. Since I will not challenge

[5] See Sober and Wilson (1998, Chapter 2) for an influential and far more sophisticated presentation of this sort of argument.

[6] This sort of debunking argument is proposed in Street (2006).

the move from the debunker's evolutionary explanation to skeptical results for the realist in this paper, I grant that some such bridging epistemic principle is available to the debunker:

> *Epistemic Bridging Principle*: If the convergence of our moral beliefs on the moral truth (as understood by the realist) is a coincidence, then we lack sufficient justification for our moral beliefs to count as knowledge.

Together, *Connection*, *Coincidence*, and *Epistemic Bridging Principle* entail moral skepticism, on the assumption that moral realism is true.

Although the debunker's crucial *Connection* premise has been subject to less scrutiny in the literature than the *Coincidence* and *Epistemic Bridging Principle* premises, there have been some important criticisms of it in the literature. For instance, some have challenged, on empirical grounds, the idea that the presence and nature of a 'moral sense' is as it would have to be for it to be plausibly considered an inheritable trait at all – for instance, it has been challenged that the capacity for moral cognition is a *sui generis*, (near-)universal psychological kind apt to be explained by appeal to natural selection.[7]

Suppose that these criticisms are correct, and the evolutionary debunking argument fails as it stands. Nevertheless, it seems the skeptical spirit of the evolutionary debunking argument can be preserved – and, indeed, strengthened – if the argument is recast in terms of *cultural* selection. For cultural selection forces can explain the 'inheritance' of behaviors and even particular beliefs (obviating the need for the qualifier 'indirect' in the cultural selection version of the *Connection* premise) via processes of social learning, rather than by genetic inheritance of traits. A cultural-selection version of the debunking argument will be thus immune to the criticisms above, which challenge the claim that a capacity for moral cognition could be a

[7] See Levy and Levy (2020, §2.2.2).

genetically inheritable trait explainable by natural selection. The *cultural*-selection debunker will maintain that 'inheritance' of moral beliefs are strongly influenced by the effects of learning, imitation, and social pressures, etc., rather than by natural selection. And the debunking argument, when transposed, *mutatis mutandis,* into a cultural-selection key, appears to have just as much skeptical import as the natural selection version of the argument: for we seem to have no reason to think that the moral beliefs passed down by learning, imitation, and social pressures imposed by the need for cooperation would tend to correspond to objective moral truth.

I now turn to critique the debunker's evolutionary hypothesis. This argument does not rely upon questioning the debunker's empirical claims directly, but rather draws on Millikan's biosemantic theory again to identify conceptual problems in the claim that evolutionary forces shape moral belief contents.[8]

Millikan proposes a disjunctive account of reproduction or copying, according to which reproduction can occur either (1) when it is a family of devices' proper function to directly produce certain items, where those items are alike in certain respects for a reason, or (2) when it is a proper function of a device to make later items match earlier ones.[9] The former is relevant to natural (biological) selection, and the latter is relevant to cultural selection and learning. In the case of genetically inheritable traits that have proper functions, the genes responsible for the production of the trait in question have the proper function of producing copies of themselves (in offspring), thereby also copying the trait in question. For example, the genes responsible for the production of hearts in current populations were directly copied from the genes of the previous generation, and these genes have that copying function because it produces items that can

[8] See discussions in Levy and Levy (2020) and Mogensen (2016) for some empirical and theory-driven challenges to the connection premise.

[9] This is a simplifying gloss of Millikan's presentation (1984, p. 24). Some details – important, but not for present purposes – have been left out, including Millikan's understanding of 'Normal' conditions.

circulate blood (under Normal conditions), thereby performing a function essential for survival in creatures with hearts (and so also those creatures' genes).

In the case of cultural reproduction and learning, reproduction of a device by copying it from earlier devices serves a purpose (function) of coordinating and stabilizing behavior in a way useful to (serving the purposes of) the producers of such behavior. The social practice of greeting, for instance, can be served well enough if we all (within a society) use the same behavior that has previously been used to greet, rather than having to come up with novel forms of greeting on each occasion. It will be important later to see that distinctions between levels and types of selection processes (such as (1) and (2) just described) make it possible for the very same item to have different proper functions that may conflict with each other.[10] It is also important to note (for the purposes of my argument in Section 4) that these various sorts of selection processes, while differing in crucial respects, are all genuine sorts of selection; cultural selection, for instance, is not somehow less fundamental than biological natural selection.[11]

I will now argue that the biosemantic notion of proper function points the way to a rejection of debunking arguments against the moral realist. The *Connection* premise of the debunking argument has been vaguely stated, in particular with respect to the means by which selection processes are supposed to shape moral belief contents. We have a basic picture of how this is supposed to go – certain evaluative tendencies or dispositions are supposed to be inheritable, and these tendencies/dispositions are supposed to have a propensity to yield certain moral beliefs in the individuals who have them. For instance, an innate disposition to have pro-social emotions such as sympathy toward group members influences our moral beliefs about the

[10] See Millikan (2004, p. 15): "The purposes that emerge from these various levels of selection are not always compatible with one another but are sometimes at cross-purposes".

[11] Millikan argues for an analogous point concerning proper function in Chapter 1 of (2004).

value of helping others, the thought seems to be. But the details of this connection is important. There are two explananda here. On the one hand, there are tendencies/dispositions and their proper functions; on the other, there are particular token moral beliefs with specific contents and their proper functions. The proper function of a disposition towards sympathy differs from the proper function of the moral belief that it is good to help others when it comes at no cost to oneself, and in ways important to the success of the debunking argument, as I explain below.

Dispositions to experience emotions of a certain kind in response to environmental stimuli are arguably apt for explanation in terms of natural (biological) selection, at least on the assumption that such dispositions are near-universal psychological traits that can confer a reproductive advantage. By contrast, moral beliefs, like beliefs generally, will be treated in the biosemantic framework as subject to a kind of selection, but not strictly *biological* selection, for the very reasons debunkers find it important to insist that the relation between evolutionary forces and moral belief is indirect. Instead, selection pressures on moral belief are more likely to be a species of cultural selection and social learning; moral beliefs, like other ordinary beliefs, are copied via expression in language from speaker to hearer, and will tend to proliferate if they often enough serve the purposes of speakers and hearers.

Traits subject to natural selection, by contrast, need not have *stabilizing* proper functions; they have proper functions, but these do not depend on a stabilizing, standardizing symbiotic interaction between their producers and consumers. Thus, the function of moral beliefs and moral claims essentially involve their effects on their consumers (cognitive mechanisms in thought, or hearers in the case of communication) in a way that the proper function of basic evaluative tendencies do not. In short: Evaluative tendencies presumably just need to get individuals with those tendencies to act in certain ways; moral beliefs and claims need at minimum to effect a certain coordination in use between producer and consumer. The most proximate proper functions of evaluative tendences and dispositions are different from the most proximate proper functions of moral beliefs and claims. (Compare to 5.3.2, where I contrasted full-blown ethical affirmation, which is linguistically articulate and so belongs to representational systems supporting logical, semantic, and epistemological relationships with other propositions, with *proto*-ethical affirmation, which involve basic moral emotions but lack the articulation needed to enter into such relationships).

This distinction between levels of explanation in selection processes does not immediately pose a problem for evolutionary debunking arguments: After all, it seems like just one way to reiterate the importance of emphasizing that evolutionary forces can at best only indirectly influence moral belief. However, what the distinction reveals is that if the debunker admits that natural (biological) selection applies to basic evaluative tendencies/dispositions, while cultural selection or social learning applies to moral belief contents, she must further hold that these selection processes will match up – that is, will and push moral beliefs in the same (non-truth-oriented) direction. If this is so, an obvious explanation of this fact would be that our basic evaluative tendencies (subject to natural selection processes) themselves shape the cultural selection processes that in turn have a more proximate influence on moral belief contents. But this connection needs to be argued for. It is not trivial. The existence of selection processes operating at different 'levels' (e.g. natural vs. cultural) makes it possible for the functions of devices operating at these different levels to conflict with each other.[12] For instance, the eye-blink reflex functions properly when its activation is caused by and successfully deflects (in a Normal way) a bit of sand nearing one's eye (thereby preserving good vision); but the eye-blink

[12] See Millikan (2004, Chapters 1 and 2).

reflex may act against your ocular-relevant purpose of enabling your doctor to put drops in your eye.[13] In each case – properly functioning eye blink, successfully keeping your eye open for your optometrist – the same distal function is served, namely, that your vision remains preserved, but there is nevertheless a conflict here. Since natural selection pressures relevant to basic evaluative tendencies, and cultural selection and social learning pressures relevant to moral belief contents operate at different levels of selection explanation, it is not guaranteed that their influence on specific moral belief contents will match.

But surely there is some debunking-supporting connection between evaluative tendencies/dispositions and moral belief contents, one is inclined to insist. If not, we would hope to be told why the *Connection* premise has seemed so plausible that it has not been regarded important to spell out its details in discussions of debunking arguments.[14] To deny that there is such a connection would appear to allow for the possibility that individuals frequently exhibit, say, a disposition to feel sympathetic responses at least minimally sufficient to motivate helping behavior, yet fail to believe that it is good to help others when it comes at no cost to oneself. Here are three responses to this worry.

First: it would be good to have some empirical evidence for such a connection, rather than relying solely on intuition, especially if the tenability of moral realism is at issue. Given that biological purposes and socio-culturally-driven purposes can come apart, it is not unreasonable to expect an explanation of a match in the moral case; why should we suppose that moral thought and discourse have the same function as certain emotional dispositions towards (say) altruistic helping? The burden of proof is on the debunker to provide such an explanation. While I regard

[13] Millikan (2004, p. 3).

[14] Vavova, for instance, writes: "while it is important that this argument is empirical, the particular empirical claim is not important. It is replaceable and, anyway, not philosophically interesting" (2015, p. 104).

this first response as sufficient for present purposes to pose a challenge to debunking arguments, which have generally avoided providing a detailed defense of the *Connection* premise, in what follows I aim to provide further reasons to think that a non-debunking *mis*match is possible. If these considerations are plausible, this would raise the bar for an adequate defense of the *Connection* premise, that would then need to address these considerations.

Second, even if it is granted that there is a match in the influence of biological and cultural selection forces on the content of moral belief, it may be that this fact is explained by some third factor that *is* sensitive to moral truth. So, for instance, if it can be maintained, without begging the question against the debunker, that certain kinds of cooperative behavior is morally good, that fact might be employed in an explanation both for why we would have emotional dispositions that tended to result in such cooperative behavior, and why we would tend to believe (say) that helping others when it comes at no or minimal cost to oneself is good.[15] 'Third-factor' responses to the evolutionary debunking argument have received much attention in the literature, and I have no positive contribution to add here, except to note that the third response I describe just below provides an alternative, so that if 'third-factor' responses fail, all is not lost for the moral realist.

Finally, third, one might admit that: Yes, in principle the evaluative tendencies selected for in humans can come apart from the direction in which other more influential selection pressures have pushed our moral belief contents. However, this is not so surprising as it seems at first; in fact, it may be *expected*, if morality is to be about more than just "spreading our seed".[16] While it may remain that *some* moral attitudes and general moral beliefs have been selected for, what this response denies is that this is generally the case, and it denies that the influence of

[15] See Enoch (2010, 2011); Skarsaune (2011); and Wielenberg (2010, 2014) for some such 'third-factor' responses.
[16] Vavova (2021, p. 721).

selection processes is so strong as to outweigh the influence of other factors in play. I consider this third response in a bit more detail just below.

Talk of ‘selection processes’ may encourage thinking in terms of how a given trait, capacity, or tendency would promote reproductive fitness for the *individual* or *groups* whose members possess it. Discussion of proper function, too, can easily be read in terms of individual/group survival – what has this trait been doing to promote the reproductive success of its *possessor*, one may be inclined to ask. Here it is important to clarify that strictly speaking, an explanation of a device’s proper functioning is an explanation of what that device has been doing in recent history to explain *its* – that is, the *device’s* – survival. The proliferation of individuals or groups of organisms that possess a trait with a proper function will undoubtedly frequently enter into such explanations; what the genes responsible for the production of hearts do to keep themselves around is, at some level, to help keep creatures with hearts around. So too with evaluative tendencies and moral beliefs; at the very least, these must not actively and routinely hinder the survival of the creatures who have them. This goes some way towards explaining the intuitive pull of debunking arguments, explaining, for instance, why we *don’t* generally observe widespread acceptance of norms that sanction needless killing. It also vindicates Foot’s observation that there are limits to what objects moral terms can be meaningfully applied to (1959). To borrow Foot’s memorable example, we cannot sensibly say that it is morally good that one clasp one’s hands together three times in an hour. The explanation I am offering for this oddity seems to be quite different from Foot’s, however: it is not an analytic, *a priori* point about moral concepts, but is grounded (in Millikanian spirit) in the actual historical facts about the conditions in which humans evolved. These conditions might have been different; but then our terms would have meant something other than they do, and would not qualify as ‘moral terms’.

In the same way, Millikan's Swampman's (a perfect replica of Davidson created by sheer cosmic accident) doings are not explained by his possessing beliefs and desires, because he has the wrong history to *have* beliefs and desires.

While considerations to do with the survival of individuals harboring moral belief place some restrictions on the application of moral terms, it is important to note that strictly speaking, what *moral beliefs* have to do to keep *themselves* around is at best indirectly related to what promotes the proliferation of human beings, and need not be always and unceasingly directed at the proliferation of the species. For creatures, like us, who have their own purposes – purposes that are potentially in conflict with or at least come apart from the natural or biological purposes of various of our traits – it may be that moral thought and discourse serves the purpose of tracking moral truth even if this does not particularly contribute to the proliferation of the species. Our widespread moral beliefs just have to serve some purpose of ours that is not essentially in conflict with our survival. Having achieved a position of relative affluence on the whole – i.e., not being in a position where all available resources and capacities must be constantly devoted our own survival – we humans can afford to employ certain of our capacities, themselves perhaps a result of natural selection pressures in our evolutionary history, for various other non-fitness-related ends.

To sketch this possibility out a bit further: here are some general-purpose capacities, apparently distinctive of *homo sapiens*, that may have contributed to our survival in the conditions in which we evolved, but which may now be co-opted towards the end of moral conduct: (i) a capacity for sophisticated reasoning; and (ii) a tendency to value truth as such. (i) and (ii) are seemingly fitness-enhancing, equipping us with a drive to acquire true beliefs and to use them to successfully navigate our environments. Since (i) and (ii) are domain-general – and being domain-general again seems advantageous as it enables flexibility in navigating a variety of kinds of environment (including physical and social environments of varying complexity) – we would expect such capacities to be available to be co-opted towards a variety of ends, allowing them to acquire new proper functions as they facilitate success in these new domains. Evolution plausibly has played a role in equipping us with general capacities and tendencies like (i) and (ii). But this does not entail that what counts as properly functioning use (thus, use which has been selected for) of such capacities is always and only promotion of reproductive fitness. If it serves *our* purposes – if not our genes' – to track objective moral truth, our moral beliefs will perform their proper functions under Normal conditions only if they are true. Capacities like (i) and (ii), I am suggesting, are supplied by natural selection and enable moral cognition; but they do not determine the contents of the outputs of such cognition in any sense strong enough to underwrite the debunker's conclusion.

6.2.4 The argument from moral skeptical hypotheses

The familiar problem of radical skepticism is generated from our apparent inability to know that some radical skeptical hypothesis does not obtain. As an example, let us consider the hypothesis that I am currently a brain in a vat being stimulated so that everything seems to me as it currently does (e.g. everything seems to me as though I am currently in the library writing my dissertation). Call this the BIV hypothesis. The intuitive thought is that we cannot possibly have any evidence that could make it rational to rule out the BIV hypothesis; any potential evidence one might have that the BIV hypothesis is false is evidence that one would still think one had even if the hypothesis were true. My belief that I am not a BIV is thus rationally *insensitive* to the truth of the matter. Yet if I cannot rationally claim to know that I am not a BIV, the skeptical thought continues, I could not rationally claim to know anything that is logically incompatible with the BIV hypothesis, such as that I am currently in the library, or even that I have hands. This appears to follow from an intuitive closure principle on rationally grounded knowledge, according to which I can come to gain new rationally grounded knowledge by inferring it from other things of which I have rationally grounded knowledge. If I really had rationally grounded knowledge that I am currently in the library, I would be able to infer, and so come to know, that I am not a BIV. But we have already seen that I apparently cannot rationally claim to know that I am not a BIV. So I can conclude that I lack rationally grounded knowledge that I am currently in the library.

A parallel skeptical argument can be generated for the moral domain. As Sinnott-Armstrong (2006, pp. 79-80) points out, *moral nihilism* – the view that there are no moral facts, properties, or truths – is like the BIV hypothesis in that it is an internally logically consistent hypothesis that is logically compatible with how things appear to us. Thus, there seems to be in principle no evidence we could have that would allow us to rationally rule out moral nihilism from the outset. Moral skepticism appears to follow by the same reasoning employed to generate the radical skeptical conclusion in the previous paragraph.

My goal in this chapter is to explain the possibility of moral knowledge, not to solve the radical skeptical problem. My particular concern is to examine arguments that *specifically* target moral knowledge, not all knowledge. Insofar as the skeptical argument of the previous paragraph does not turn on any features specific to the moral domain, but rather is just a particular application of the radical skeptical problem, it is not my concern to provide a response to it. But Sinnott-Armstrong has argued that the moral skeptical problem is importantly different from the radical skeptical problem. Moral nihilism, unlike the BIV hypothesis, is a theory that some people actually believe to be true, and which can be supported by various reasons, "including the pervasiveness of moral disagreement and our supposed ability (with the help of sociobiology and other sciences) to explain moral beliefs without reference to moral facts" (2019). Since the various reasons for thinking moral nihilism might be true are separately addressed in this chapter, I do not take them yet to provide reason to think the argument from skeptical hypotheses provides a special problem for moral knowledge. Still, in addressing what I take the most pressing version of the argument from disagreement below, I make use of an epistemological theory that dissolves the radical skeptical problem anyway. At that point, I shall be able to explain how the skeptical argument from moral nihilism can *also* be dissolved in the same way (see Appendix; see also Johnson, 2019).

6.3 The argument from disagreement

Moral realism supposedly faces a problem when it comes to accounting for moral disagreement.[17] The intuition generating this problem is that moral disagreements appear to be more widespread, or more intractable, than we would expect if moral realism were true. Now, 'the' argument from disagreement can take many different forms, some more compelling than others. I think the strongest version of the argument from disagreement, to be discussed below, poses a special problem for the hybrid account of this dissertation, and so I shall be concerned to leverage additional theoretical resources in order to address it.

6.3.1 The argument from disagreement (I): Inference to the best explanation

Some versions of the argument from disagreement target the *metaphysical* claim shared by moral realists that there are moral facts or properties that are in some sense objective and mind-independent, rather than the *epistemic* commitment of non-skeptical moral realism that we

[17] See Tersman (2006); Sinnott-Armstrong (2006); Enoch (2009, 2011); Shafer-Landau (2003); Wedgwood (2014); Ridge (2014, Ch. 2); Audi (2014); Fritz (2018), and Vavova (2014) for discussion of the problem of disagreement for moral realism.

can have knowledge of such facts or properties. One way for such an argument to take shape is as an inference to the best explanation (IBE). This sort of argument starts from the observation that actual moral disagreement is different from disagreement in paradigmatically realist domains in that it is more widespread and intractable. It is then proposed that the best explanation of these differences between moral disagreement and disagreement in realist domains is that there are no mind-independent moral facts or properties for us to really disagree about.

The argument might also be put as follows: if moral realism were true, we would expect there to be less moral disagreement than there in fact is, or that such disagreement would be rationally resolvable more often than they seem to be. The breadth and depth of moral disagreement then provides evidence that realism is false.[18] The skeptic proposes that the best explanation of moral disagreement is that our ethical judgments simply reflect our own (or our culture's own) moral outlook, rather than some objective moral reality. A version of this simple argument can be found in Mackie (1977, p. 36), which I quote at length (Mackie labels this the 'argument from relativity'):

> The argument from relativity has as its premiss the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community. Such variation is in itself merely a truth of descriptive morality, a fact of anthropology . . . . Disagreement on questions in history or biology or cosmology does not show that there are no objective issues in these fields for investigators to disagree about. But such scientific disagreement results from speculative inferences or explanatory hypotheses based on inadequate evidence, and it is hardly plausible to interpret moral disagreement in the same way. Disagreement about moral codes seems to reflect people's adherence to and participation in different ways of life. The causal connection seems to be mainly that way around

[18] See Enoch (2011, Ch. 8) for discussion of arguments along these lines.

In order to assess this argument, as an inference to the best explanation, we should consider alternative explanations of the 'data' about moral disagreement that are compatible with realism. For instance, one might argue that the best explanation of the breadth and depth of moral disagreement is not that moral realism is false, but rather simply that when it comes to moral matters our views are more likely to be shaped by our self-interests than in other domains such as scientific inquiry (Enoch 2011, pp. 191-195). Still, it should be noted that insofar as self-interest is epistemically distorting in the moral domain (and this does seem plausible), this explanation still yet may not go far enough to address moral skepticism.

One way to respond to the IBE argument is to argue that the premise that moral disagreement is significantly different from disagreement in other areas is, if not false, greatly exaggerated (Enoch 2011, pp. 190-191). For instance, one might hold that *apparent* moral disagreements are best understood as really disagreements concerning non-moral matter of fact. It is a matter for further consideration just how much apparent disagreement on moral matters can be explained in terms of disagreement in non-moral belief, and where this leaves the IBE argument. Nevertheless, it seems to me overly optimistic to think that *all* such 'apparent' moral disagreement can be explained away in this manner. That is, I think we should expect there to be at least some fundamental moral disagreement. Still, I concede that such disagreement may be less frequent than how it is portrayed by various arguments from disagreement.[19] The question is how much fundamental disagreement is needed to warrant substantial skeptical results.

6.3.2 The argument from disagreement (II): Semantic competence

Another version of the argument from disagreement takes on a *semantic* character, rather than a metaphysical or epistemic one. The basic idea behind this argument is that the pattern of

[19] See for instance Ridge (2014, pp. 64-76), who maintains that fundamental normative disagreement is especially ubiquitous.

moral disagreement we see both within and across cultures and times, as well as intuitions about merely possible disagreement, speak against the linguistic hypothesis that moral terms such as ‘wrong’ and ‘good’ have the primary semantic function of referring to naturalistic kinds, as certain forms of naturalist moral realism maintain. In particular, the intuition that we can have genuine moral disagreements with individuals and communities that apply the terms “wrong”, “good”, etc. to actions very different from those which *we* would apply them to suggests that those terms must be doing something other than (simply) referring to some property shared by the items to which they supposedly apply. If moral terms did have this simple referential function, why is it that we disagree so much in moral matters, and why is it that these disagreements do not appear to be simple instances of talking past each other? Versions of this family of argument from disagreement can be found in Hare’s case of the missionary and the cannibals (1952, p. 148), and more recently in Horgan and Timmons’ (1991) Moral Twin Earth Argument.

Now, the semantic competence version of the argument relies on certain assumptions about semantic competence that have not gone without challenge. For instance, Horgan and Timmons’ Moral Twin Earth argument applies specifically to Cornell-style moral realism (as exemplified by Boyd, 1988; Brink, 1986; Sturgeon, 1985; Railton, 1986), which is committed to an externalist causal theory of reference for moral terms (following Putnam, 1975, and Kripke, 1980). An attractive feature of the semantic commitments of Cornell-style moral realism is that it appears able to explain Moorean Open Question intuitions – something that analytic reductionist moral realism struggles to do. The idea is that moral predicates can refer to naturalistic properties even if they cannot be reductively analyzed in terms of those properties. Given their unanalyzability, it is intelligible for someone to ask “Yes, X promotes pleasure, but is X good?”

even if the predicate ‘good’ just refers to the naturalistic property of promoting pleasure. Horgan and Timmons’ idea is that this casual-historical externalist approach cannot account for the intuition that speakers from two communities, one from Earth, and one from Twin Earth, whose historical use of “wrong” was causally regulated by different sorts of moral properties, can successfully disagree on moral matters, rather than simply talking past each other as the causal-historical picture of reference for moral terms seems to predict.

Now, as discussed in Chapter 5 fn. 13, the biosemantic account I have proposed, while similar to Cornell-style moral realism in endorsing a semantic externalist account of the referents of moral terms, completely avoids the Moral Twin Earth problem. According to Millikan’s view, what matters for the content of a significant linguistic device are the effects tokens of that device have had on receivers that account for the device’s continued proliferation. Significantly, on this view, content is *not* determined by speaker conceptions. Accordingly, the intuitions of competent speakers about disagreement have very little ‘probative value’ (as Dowell, 2016, puts it) for discovering semantic content. Thus, the Moral Twin Earth argument, in relying on speaker intuitions about when there is genuine moral disagreement, begs the question against Millikan’s biosemantics as applied to the ethical domain (see Dowell, 2016; Millikan, 2010).

Ridge (2014) provides a version of the semantic competence argument that is more general, in that it does not presuppose a causal-historical picture of reference. Ridge takes it to be a plausible general principle about semantic competence for referring predicates that competence “requires mastery, however implicit, of some reliable method for deciding whether that predicate applies in a given case” (2014, p. 68). Thus, on the assumption that moral predicates refer, “any two competent speakers should, simply in virtue of their competence, both have at their disposal a method which would lead them to converge in their application of normative predicates, for the most part, at least under suitably ideal circumstances" (2014, p. 69). Assuming the principle about semantic competence above, the best explanation of the fact that speakers fully competent with moral claims persistently disagree with each other, Ridge maintains, is that moral terms do not have a primarily referential semantic function.

However, even the above assumption about semantic competence is rejected by the biosemantic proposal. Semantic competence, in Millikan's view, requires very little on the part of the speaker – indeed, Millikan maintains that speakers can competently use a term even after just one encounter with that term, and without any clear conception of what it might refer to (2010, 2017). This minimal conception of semantic competence is supported by, among other things, (i) the point that young children learn words too quickly to plausibly suppose that they learn something like the necessary and sufficient conditions for the applications of those words along with learning the words (Millikan, 2017, pp. 32-33; see also Chomsky, 1995). And (ii) the point that competent users of a term can agree in many of their judgments *despite* having very different conceptions associated with that term – for instance, the way that *I* tell weasels from other animals is very different from the way an expert on weasels can so tell (for one thing, the way I might tell the difference is by *consulting* the expert, where this is clearly not the way that the expert tells) (Millikan, 2010, p. 47). Since the biosemantic account endorsed in Chapter 5 rejects Ridge's principle for semantic competence, I set Ridge's argument aside as not applying to my account. A wide range of moral disagreement is compatible with moral terms having a referential function. However, as we saw in the last chapter, there still must be enough stability in application and resulting behavior for moral terms to aid in social coordination. Thus the present response to Ridge's argument may need to be supplemented with a 'debunking' strategy (discussed above) for showing that there is not as much moral disagreement as some (such as

Ridge) take there to be. Still, the relevant stability in use is not to be explained in terms of the conceptions employed by speakers, but rather by the stabilizing proper functions of the terms themselves.

6.4 Moral realism and the skeptical argument from disagreement

The most compelling versions of the argument from disagreement, to my mind, are epistemological arguments that target the possibility of moral knowledge. Although strictly speaking, moral realism carries no commitment to the denial of skepticism about moral knowledge, generally, those drawn to a realist stance take a non-skeptical view on moral knowledge. Should moral skepticism be unavoidable on the assumption that moral realism is true, I take it a main attraction of moral realism would be undercut.

6.4.1 Disagreement and skepticism[20]

Prima facie, disagreement is relevant to knowledge. *Conciliatory* stances in the epistemology of disagreement maintain that the discovery that one disagrees with an epistemic peer seems to make it rational to at least lower one's confidence in one's initial position.[21] A refusal to lower one's confidence seems to amount to a rejection of the possible relevance of any evidence one's interlocutor might have. Not only does it appear irrational to refuse to consider a salient, relevant possible source of evidence; it seems intellectually arrogant to do so. Initially, then, being conciliatory seems to be both the epistemically rational and the intellectually humble response to discovered disagreement with one's epistemic peers.[22]

[20] Parts of this section are drawn from Johnson (2019, 2021).

[21] Feldman (2005); Christensen (2007); Elga (2007).

[22] As Carter and Pritchard (2016) put it: "A widely shared insight in the disagreement literature is that, in the face of a disagreement with a recognised epistemic peer (such as between Hawking and Penrose), the epistemically virtuous agent should adopt a stance of *intellectual humility - that is, a stance where one exhibits some measure of epistemic deference by reducing one's initial confidence in the matter of contention*" (p. 3, emphasis added, footnote omitted).

In ordinary cases of peer disagreement, granting the conciliatory thesis does not have absurd skeptical implications, for two reasons. First, even if disagreement over some isolated proposition P makes it rational to lower one's confidence in P (or not-P) to such a degree that one's doxastic state would not count as knowledge even if it were true, this only would rationally require skepticism concerning *whether P* – wholesale skepticism about the domain of inquiry to which P belongs does not follow. And second, in instances of ordinary peer disagreement, generally, disputants can trust that a rational resolution to the disagreement is in principle possible. For instance, consider the standard 'restaurant bill' case (Christenson 2007, p. 193), wherein two people, call them Joe and Joan, equally competent in simple mathematical calculations (and in other relevant epistemic capacities) arrive at conflicting conclusions about what each owes for the restaurant bill. Because this is a disagreement between epistemic peers, conciliationists maintain that Joe and Joan should each lower their confidence in their respective judgments about how much they owe. But this skeptical result is surely temporary – Joe and Joan can easily resolve their disagreement, and come to a well-justified consensus about what they each owe, simply by re-checking their initial calculations, or consulting a calculator, etc. That is, ordinary peer disagreements are *shallow* disagreements.

Proponents of the argument from disagreement towards moral skepticism, however, will maintain that moral disagreement is unlike ordinary disagreement in ways relevant to moral knowledge. First, unlike ordinary cases of peer disagreement, moral disagreement is supposedly rampant, such that we find disagreement not only over isolated moral propositions, but also disagreement over entire moral outlooks. The 'restaurant bill' case described above represents an isolated instance of mathematical disagreement; lowering one's confidence in response to that disagreement, by itself, should not make it rational to also lower one's confidence in one's mathematical beliefs as a whole. But in some (at least possible, if not actual) moral disagreements, what is at issue is not a specific moral claim (such as "It was wrong for Joe to lie about how much he liked the dinner, even though he did it to avoid hurting our host's feelings"), but rather, claims that have bearing on one's entire moral outlook. Consider, for instance, some of the moral-political questions dividing conservatives and liberals in the current U.S. political context, including questions over the role of government, moral desert and social welfare, racial justice, and more. Haidt (2012) even argues that liberals and conservatives differ in the weight they place on different foundational moral principles, with liberals relying more on what he calls the 'care/harm' foundation, and less on the 'loyalty/betrayal', 'authority/subversion', and 'sanctity/degradation' foundations. Accordingly, if conciliationism is right, peer disagreement over moral propositions fundamental to one's moral outlook might make it rational for one to lower confidence in *all* of one's moral positions to such an extent that one can no longer rationally claim moral knowledge.[23] While I suggested above that there may be less moral disagreement than there seems (given that much moral disagreement may be based in differences in non-moral belief), still, it seems that at least on occasion, we are confronted with genuinely moral disagreements concerning moral propositions central to our respective moral outlooks.

A second way in which moral disagreements can differ from ordinary peer disagreement is in their apparent intractability. Unlike disagreements between experts in, say, physics or chemistry, it seems possible for rational moral disagreement to persist even when there is no

[23] One might argue that because moral propositions fundamental to one's moral outlook are intertwined with one's moral judgments as a whole, a one-off disagreement about such a proposition is insufficient to undermine one's moral knowledge in general. Indeed, I shall go on to endorse a proposal in this spirit. However, for the time being, note that it seems possible for one's fundamental moral commitments to be persistently challenged over many occasions, rather than just a 'one-off' instance. If enough of one's core moral commitments are challenged often enough in peer disagreement, it seems plausible to think, from a conciliatory perspective, that the accumulated epistemic significance of the disagreements would make it rational to lower confidence in one's moral judgments in general.

underlying disagreement in non-moral belief, no errors of reasoning on the part of any disputant, and no other obvious cognitive mistakes that could account for the disagreement. Well-informed, rational people can intelligibly disagree on moral questions fundamental to their moral outlooks. Indeed, this is just what we should expect: As Rawls (2005) tells us, it is part of living in a liberal democracy that there be a plurality of reasonable religious, moral, and philosophical doctrines. Inevitably, some of these doctrines will conflict. When deep disagreement occurs, the prospects for reaching a rational resolution where disputants converge on a single position appear dim.

In sum: insofar as we endorse the plausible-seeming conciliatory stance on the epistemic significance of peer disagreement, the breadth and depth of actual or possible moral disagreement among peers or presumed peers seems to rationally compel us towards moral skepticism. After all, if it is possible for me to disagree with an interlocutor radically and fundamentally in moral matters, where I should assume (both as a rational, but also a moral/political point) that my interlocutor is well-informed, rational, and making no obvious cognitive errors, how can I be sure that *my* moral outlook is correct, rather than my interlocutor's? And if I have no rational grounds for accepting my current moral outlook rather than my disagreeing interlocutor's, it seems irrational for me to regard myself as having moral knowledge. Insofar as moral realists are concerned to establish the legitimacy of our claims to moral knowledge, the possibility of deep moral disagreement appears to present a problem.

6.4.2 The hinge epistemic account of deep disagreement

In this section, I begin developing a realist-friendly anti-skeptical response to the argument from disagreement discussed in the previous section. While this response depends upon a controversial hinge epistemology, a dialectical advantage is that it can grant the skeptic's claims about the extent and rational irresolvability of moral disagreement. That is; unlike other prominent responses to the argument from disagreement, the response sketched here can grant that moral disagreements are indeed more common and more intractable than disagreements in other factual domains. What the present response denies is that the possibility of such deep moral disagreement has the skeptical import claimed by the argument from disagreement.

6.4.3 Hinge epistemology

The hinge epistemic account of deep disagreement holds that deep disagreements are not rationally resolvable when they concern hinge commitments. According to the version of hinge epistemology I prefer (based on the work of Pritchard, 2012, 2016), ‘hinges’ are commitments which are held with maximal subjective certainty, yet which are (for that very reason) not directly responsive to rational considerations and difficult to abandon. Because these commitments are not directly responsive to rational considerations and difficult to abandon, neither the mere fact of disagreement, nor the reasons put forward in the course of an argument can lead one to rationally doubt the hinge, and neither can one simply abandon the hinge and thereby come to be able to subject the relevant contested proposition to rational evaluation.

Hinge epistemology, which takes inspiration from Wittgenstein’s *On Certainty* (1969), is often presented as a response to radical epistemological skepticism. The core idea, shared by various accounts of hinge epistemology, is that certain of our commitments are exempt from doubt, because of the special role those commitments play in making the epistemic practice of giving and asking for reasons (including raising doubts) possible in the first place.

There are different views on how best to understand the notion of a hinge commitment, each with different implications for the analysis of disagreement. Here are four dimensions along which hinge epistemologies differ. First, there is a division among views that take hinge commitments to have propositions as their objects (including Pritchard, 2012, 2016; Coliva, 2010, 2015, 2016b; Wright, 1986, 2004), and those that take hinge commitments to concern non-propositional objects that are neither true nor false (Moyal-Sharrock, 2004, 2016). There is then a second division between views that take hinge commitments to have a positive epistemic status (Wright, 2004, 2014; Williams, 1991), on the one hand, and views that take the attitude of hinge commitment to be removed from the scope of rational evaluation altogether, and so not even enjoying epistemic entitlement (Pritchard, 2012, 2016; Coliva, 2010, 2015, 2016b), on the other hand. Relatedly, whereas on Wright's view our entitlement to accept a hinge proposition is defeasible, in that one may only rationally accept the hinge in the absence of any reason to think the hinge false (2004, p. 181), for Pritchard, our hinges are not responsive to evidence even in this way. On Pritchard's view, *because* hinges are held with maximal subjective certainty, any reason that might be offered to think the hinge false will seem less certain to one than the hinge itself, and so one would instead have reason to reject the purported counter-evidence rather than the hinge.

Third, some views seem to take hinge commitment to be a context-sensitive notion, in that an attitude's status as hinge commitment is specific either to a particular domain of inquiry or intellectual project (e.g., Wright, 2004; Williams, 1991). Other views take a commitment's status as a hinge commitment for a person to be largely independent of the context of inquiry (Pritchard, 2012, 2016). For instance, on Williams' view, the proposition that the world did not come into existence just five minutes ago will count as a hinge commitment relative to certain domains of inquiry (e.g. history), but not relative to others (e.g., philosophical reflection on skepticism) (Williams, 1991, pp. 121-125). By contrast, on Pritchard's view, if the proposition that the earth did not come into existence just five minutes ago is held as a hinge commitment by someone, then that person will hold this as a hinge commitment regardless of context of inquiry (2016, p. 106).

And finally, some views hold that there are principled limits on what kinds of propositions can play the hinge commitment role (Wright, 2004; Coliva, 2016b) – in particular, only 'Moorean certainties' such as the proposition that one has two hands can be hinge propositions. Whereas others hold that (almost) any proposition can be a hinge commitment for a particular person, so long as it plays the relevant role (Pritchard, 2012, 2016). Thus, on Pritchard's view it could be that, in the right circumstances (given one's upbringing, the beliefs of one's culture, etc.), one could be maximally subjectively certain of, and therefore have a hinge commitment to, just about any proposition. It is this feature of Pritchard's view that I think makes it the most natural framework for developing a hinge epistemic account of *disagreement*.

Although Pritchard's view allows for variability in personal hinge commitments, it also captures what all hinges have in common, in virtue of which they are hinges: namely, the functional role the hinge commitment plays in the cognitive economy of the person holding it. All hinges are alike in that they are immune from direct rational evaluation and difficult to abandon, because they are held with maximal subjective certainty. And despite the possibility for variation in hinge commitment between people, we should also expect that generally there will be agreement in our hinge commitments; indeed, Moorean certainties (such as commitment the proposition that one has two hands) and anti-skeptical commitments (such as commitment to the proposition that the earth did not pop into existence just 5 minutes ago) are hinge commitments that nearly everyone will share.[24] As I think of hinge epistemology, in the game of giving and

[24] As Pritchard puts it, all the various personal hinge commitments 'codify' the more basic 'uber hinge commitment' we all share to the proposition that one is not fundamentally and radically mistaken in one's beliefs. This uber hinge commitment then entails the denials of skeptical hypotheses, thus generating anti-skeptical hinge commitments which will also be nearly universally shared (2016, pp. 94-103).

asking for reasons, hinge commitments are like the board on which this game is played, in that they are not themselves subject to requests for reasons, nor do they directly provide reasons (they are not moves on the board), but they are commitments we must hold if we are to give and ask for reasons at all (they *are* the board).[25]

Thus, drawing from Pritchard's non-epistemic propositional view, we can define the notion of 'hinge commitment' as follows:

> *Hinge Commitment*: S has a hinge commitment to the proposition that p just in case S is subjectively maximally certain that p is true, where (for that reason) S's commitment is *a*rationally held, in that it is not based on any particular reasons for thinking p true, and is generally resistant to purported reasons for thinking p false.

The arationality and the maximal subjective certainty are crucially related. Because hinges are maximally subjectively certain, no evidence can lead one to rationally reject one's hinge (again, by one's own lights), as the hinge will be more certain than any reasons speaking against it. But by the same token, no evidence can speak in favor of a hinge, either, as any such evidence will also be less certain than the hinge itself. Thus, we should understand Pritchard's view as adopting the following principle, which I dub 'the rational grounding principle' (see Pritchard, 2012, pp. 256-257; 2016, pp. 63-66):

> *The Rational Grounding Principle*: Rational grounds for S to doubt (or believe) proposition p must themselves be more subjectively certain to S than the proposition p which is to be doubted or believed.

[25] This feature of hinge commitments (their being removed from the scope of rational evaluation) distinguishes them from fundamental epistemic principles (as discussed by Lynch, 2010, 2012, 2016). Fundamental epistemic principles cannot themselves be epistemically justified in a non-circular manner, yet they do still provide a source of epistemic justification for beliefs. Hinge commitments do not themselves directly justify other beliefs; rather, holding the hinge commitment is a prerequisite for being able to have justification for one's beliefs.

When we combine this principle with the observation that, for any epistemic agent, there will be some propositions the agent holds with a maximum subjective degree of certainty, the core commitments of Pritchard's hinge epistemology fall out:[26] First, whatever is held to a maximum degree of subjective certainty must be rationally groundless – by definition, there is nothing more certain for the agent that could stand as its rational ground. Second, whatever is held to a maximum degree of subjective certainty cannot directly provide rational grounds for knowledge of other initially less certain propositions. For, the hinge, being itself rationally groundless, cannot serve as the rational ground for accepting other propositions. If it could, then reason to doubt the proposition so grounded could provide reason to doubt the hinge commitment grounding that proposition. But hinges, being maximally subjectively certain, are immune to rational doubt. As a result, hinge commitments are themselves arational, standing outside of our ordinary epistemic practices. Yet we must have *some* hinge commitments in place, because it is only relative to (though not by appeal to) these maximally subjectively certain commitments that we are able to provide rational grounds for doubting and believing other propositions. As Pritchard concludes, the fact that hinge commitments stand outside the scope of rational evaluation shows that our practices of rational evaluation are necessarily local – it is thus simply not possible to subject all of one's cognitive commitments to doubt at once, as radical skepticism would have us do.

[26] One might wonder why it would be true that epistemic agents must hold some commitments that are maximally subjectively certainty and yet arational. A full defense of hinge epistemology goes beyond the scope of the present chapter, but the basic thought is that it is part of the 'logic' of rational evaluation that if some grounds are to be 'more' or 'less' certain for S, they are so only relative to some maximally certain proposition(s) that S holds. Given the rational grounding principle, however, such a proposition could not be directly rationally supported by other propositions -- hence, it is rationally groundless for S. And since such a proposition is rationally groundless for S, it also could not directly rationally support other propositions for S. Thus, holding some propositions with maximal certainty is a prerequisite for rational evaluation altogether, yet such propositions would be immune from rational evaluation themselves.

Because a hinge commitment is an attitude towards a proposition that involves commitment to the *truth* of that proposition, and given that there can be *divergence* between the hinge commitments of individuals it makes sense to say we can *disagree* about hinges. Now, as noted earlier, the attitude of hinge commitment towards a proposition is not directly responsive to epistemic reasons for thinking that proposition true or false, and not easily abandoned. This is why disagreements that directly concern hinge commitments will be persistent and rationally unresolvable. Nevertheless, hinge commitments can change over time, so there remains the possibility that hinge disagreement can eventually be resolved.

Though hinge disagreement will be rare, I do think that there are some real life instances. It is difficult to say to any degree of precision whether a given disagreement is a hinge disagreement, because this will largely depend on the role that the proposition under dispute plays in the cognitive economy of each disputant. Pritchard offers the following as examples: “someone raised in a religious community where God’s existence is taken as an obvious fact of life is likely to have religious hinge commitments that would be alien to someone raised in a largely secular environment. Or consider someone raised in a deeply politically conservative social milieu, as opposed to someone brought up in a commune exclusively populated by people of a left-wing political persuasion. Clearly, one would expect this to lead to individuals with very different hinge commitments regarding core political matters” (2018, p. 3).

I would add, as a possible historical example, the attitudes held by white slave-owners towards slavery in the pre-civil war American South, compared with the attitudes of abolitionists. It would not be surprising, I think, for a white person raised in a plantation setting, confronted with slavery as an everyday fact of life, to hold that the institution of slavery is morally justified as a hinge commitment. Racist attitudes held even today may, for some people, play the hinge commitment role.[27] As a concrete instance, I offer the case of Derek Black, a former white nationalist. From a young age, Black was raised to accept white nationalist ideals - “his father, Don Black, had created Stormfront, the Internet’s first and largest white nationalist site, with 300,000 users and counting. His mother, Chloe, had once been married to David Duke, one of the country’s most infamous racial zealots, and Duke had become Derek’s godfather,” and “Derek had been taught that America was intended as a place for white Europeans and that everyone else would eventually have to leave. He was told to be suspicious of other races, of the U.S. government, of tap water and of pop culture” (Saslow, 2016).[28] Given this sort of upbringing, I suggest that Black’s white nationalist views are plausible candidates for hinge commitments with which we should disagree.

6.4.4 Hinge disagreement

The argument from moral disagreement that I wish to address was that, granting the initially plausible seeming conciliationist idea, together with the point that some moral disagreements between epistemic peers are not directly rationally resolvable, yields the result that we cannot rationally claim knowledge when it comes to contested moral matters. Since the possible targets of intractable moral disagreement include matters that are fundamental to whole moral outlooks, the possibility of intractable moral disagreement threatens a skeptical result for a fairly wide range of the moral domain.

In response to the skeptical argument from disagreement, I argue that intractable moral disagreements are plausibly explained as disagreements in hinge commitment, and that a peculiar feature of disagreements in hinge disagreement is that they *do not* make it rational for disputants to lower their initial confidence in their views, thus avoiding the semi-skeptical conclusion of the

[27] These examples illustrate, of course, that hinge commitments are not guaranteed to be true.
[28] Thanks to Tracy Llanera for introducing me to this story and influencing my thinking on it.

argument. Conciliationism may be the correct view on what it is rational to do in cases of *shallow* disagreement. But I shall hold that the conciliatory response is not correct when it comes to deep disagreement.

Because a hinge commitment is an attitude towards a proposition that involves commitment to the *truth* of that proposition, and given that there can be *divergence* between the hinge commitments of individuals, it makes sense to say we can have disagreement about hinges. Hinge disagreement will be unlike shallow disagreements such as the disagreement in the restaurant bill case. In the restaurant bill case, each of the disputants is plausibly committed to there being, in principle, some epistemic method that could be appealed to settle the dispute – say, checking a calculator. Disputants could not rationally defer to an independent epistemic method for adjudicating the dispute unless they were prepared to hold the results of that method as more certain than their initial confidence in their own judgments about how much is owed for the bill. It is the nature of hinge commitment, however, that one could *not* hold the result of some epistemic method as more certain than one's hinge commitment itself. Rather, it is the other way around: if the result of some epistemic method or principle conflicted with one's hinge commitment, this could only give one reason to reject that method or principle, or to double-check that it was executed properly. (Example: when I am outside on a hot summer's day, I do not lower my confidence in my belief that it is over 85 degrees Fahrenheit when I look at a thermometer that reports the temperature at -18 degrees Fahrenheit: rather, I assume that there is a fault with the thermometer). As discussed earlier, the attitude of hinge commitment towards a proposition is not directly responsive to epistemic reasons for thinking that proposition true or false. This is why disagreements that directly concern hinge commitments will be persistent and rationally unresolvable, in just the way that the deep moral disagreements driving the argument from disagreement are supposed to be. For this reason, I think that a plausible way to explain the persistence and irresolvability of some deep moral disagreement is by treating such disagreement as disagreement in moral hinge commitments.

Conciliationists consider the fact of disagreement to provide reason to lower one's initial confidence. But hinge commitments are such that they can be neither supported nor called into question by rational considerations. Thus, when it comes to hinge disagreement, the fact of disagreement cannot provide reason for disputants to lower their initial confidence. The significance of this for the argument from disagreement is: if (as I think is plausible) it is right to explain some deep moral disagreement as hinge disagreement, then such deep moral disagreement is indeed rationally irresolvable, but it also does not merit skeptical conclusions. We can continue to claim moral knowledge, since we are not rationally compelled to lower our confidence in our fundamental moral outlooks in light of disagreement. Yet having such commitments enables one to have rationally grounded knowledge of other propositions. Deep moral disagreement was supposed to be problematic insofar as the disagreement merits skepticism about moral propositions central to one's moral outlook. The anti-skeptical point here is that one's moral hinge commitments are (i) likely central to one's overall moral outlook, and (ii) cannot be rationally abandoned simply in light of disagreement, since they are not directly responsive to rational considerations).

This concludes my offered response to the argument from disagreement against moral knowledge. It is more complex than I would like it to be, and it relies upon a radical epistemological stance. Nevertheless, it seems to me that this is the correct response to the problem. In the next section, I conclude by considering an apparent and unwelcome implication of this solution, namely, that it warrants *dogmatism* concerning one's own moral views. I argue that there is, on the contrary, a sort kind of intellectual humility that is appropriate for our fundamental moral commitments.

6.4.5 Moral hinge commitments and intellectual humility

We can now return to a bit of unfinished business from Chapter 2. In the context of discussing the problem of fundamental moral error for quasi-realists (2.5,4), we saw that Ridge treats certainty that P as a matter of having a much higher credence in P than not P. I proposed instead that we think of certainty in terms of 'super-stable' belief, which we can now see essentially amounts to *hinge commitment*. I argued that Ridge's response to the problem of fundamental moral error requires the 'much-higher-credence' account of certainty: Thus, if the hinge epistemology I have discussed is correct, Ridge's solution to the problem of fundamental moral error fails. I noted, however, that Ridge is on *prima facie* good grounds in rejecting the hinge account in favor of his much-higher-credence account. This was because it is difficult to see how adhering to one's hinge commitments in the face of peer disagreement could amount to anything other than dogmatism or intellectual arrogance. The problem of fundamental moral error is supposed to show that the quasi-realist must be unpardonably smug about her moral commitments; now the hinge epistemic account seems to do no better.

What is needed to address this concern is an account of intellectual humility that is appropriate to our hinge commitments. This is what I now provide. It seems to me that a way to exhibit intellectual humility about one's hinge commitments is to recognize their rational groundlessness. As Wittgenstein remarks, "[t]he difficulty is to realize the groundlessness of our believing" (1969, §166). Intuitively, it seems a mark of intellectual arrogance to think that one's beliefs are always capable of being given compelling rational support. We exhibit intellectual humility by having the self-awareness to know when our reasons give out.

The account I propose for intellectual humility concerning hinge commitments is the following:

> *Intellectual Humility for Hinge Commitment:* (i) an awareness of one's hinge commitments as rationally groundless, and (ii) a willingness to stand by one's hinge commitments, in the sense of taking proper responsibility for the hinge commitments one has.

Regarding (i): I do not see this as requiring that ordinary folks be aware of the details of hinge epistemology. Rather, realizing the rational groundlessness of certain of one's commitments is a way of owning the fact that one has not 'earned' those commitments through intellectual effort (it is thus a way of owning one's limitations).[29] This recognizes the Wittgensteinian point that it is part of becoming a member of a community of epistemic agents that we must 'swallow down' some propositions as being beyond doubt without argument.[30] A failure to recognize the rational groundlessness of certain of one's commitments amounts to intellectual hubris. This may provide a possible interpretation for why Wittgenstein admonishes Moore for asserting that he *knows* he has hands (1969, §151); in so doing, Moore presents himself as being in a position to demonstrate that this is so – that is, Moore might be charged with intellectual hubris. In this sense, arguments for radical skepticism can be seen as an important corrective, and are properly regarded as humbling.[31]

[29] In this way, the view fits with Whitcomb *et al.*'s (2017) view of intellectual humility as appropriate awareness and taking ownership of one's cognitive limitations. But the limitation here is not merely a contingent limit; it is not as if we could persuade everyone if only we were excellent debaters. Rather, this is a principled limit.

[30] Wittgenstein (1969, §143): "I am told, for example, that someone climbed this mountain many years ago. Do I always enquire into the reliability of the teller of this story, and whether the mountain did exist years ago? A child learns there are reliable and unreliable informants much later than it learns facts which are told it. It doesn't learn at all that that mountain has existed for a long time: that is, the question whether it is so doesn't arise at all. It swallows this consequence down, so to speak, together with what it learns".

[31] In this respect, the attitude of intellectual humility I identify is similar to Hazlett's (2012), since on his view, one can exhibit intellectual humility regarding one's own knowledge by taking up a higher-order attitude of agnosticism about *whether* one knows - one can have knowledge and yet suspend judgment about whether one does have that

But in considering arguments for radical skepticism, we should not become overly modest in our estimation of our epistemic positions. An acceptance of radical skepticism in this regard would amount to an intellectual *meekness*. The radical skeptic underestimates her epistemic abilities to such an extent that she claims no knowledge whatsoever; in attempting to doubt even that she has hands, she attempts to disavow any intellectual commitments at all. This is why the proper attitude towards our hinge commitments will involve (ii) – taking responsibility for the hinge commitments one in fact has. As I understand it, 'standing by' one's hinge commitments means recognizing their subjective certainty and continuing to endorse them. One can fail to stand by and take responsibility for one's hinges by (misleadingly) presenting them as open to rational revision. In more mundane contexts, where radical skepticism is not under discussion, taking proper responsibility for one's hinge commitments might include not concealing those commitments (including in self-deception).[32] In the context of deep disagreements, one way to *fail* to take responsibility for one's hinge commitment is to (misleadingly) present that commitment as though it were an ordinary (if firmly held) belief, in principle open to rational revision. This can occur when one engages in a dialogue concerning whether p while concealing the fact that the question of whether p is not genuinely open for one – this would be a bad faith effort at dialogue.

The self-awareness involved in the intellectually humble attitude to take towards one's hinge commitments is (following Whitcomb *et al.*, 2017) a mean between obliviousness and

knowledge. Similarly, I am maintaining, we can recognize the rational groundlessness of certain of our beliefs, without thereby being rationally compelled to abandon those beliefs.

[32] This relates to Pritchard's discussion of 'dialectical posturing' (2017 27-29). A dialectical poseur engages in a debate inauthentically. "By this I mean that there are parties to the dispute who, far from expressing their genuine convictions about the subject matter at hand, are instead merely playing a certain role, wearing a dialectical hat, if you will" (27) – whether they consciously mean to or not, I would add. Someone who asserts (and so presents themselves as knowing) that there is no such thing as knowledge is a dialectical poseur, in this sense.

obsessiveness over one’s limitations. Obliviousness to one’s hinges would amount to a failure to recognize where one's reasons give out. The oblivious person will continue to offer reasons that do not really have any bearing on their own commitment; the provided reasons would be a post-hoc rationalization of the commitment. Obsessiveness over one's limitations may lead one to think reasons have given out before they really have; the obsessive would be so uncertain of his own ability to present his authentic reasons for belief, that he will likely avoid argument too often.

Now, this position may still seem to amount to an endorsement of *dogmatism*, rather than an articulation of a peculiar kind of intellectual humility. After all, my account maintains that even after we recognize that some of our commitments are rationally groundless, we may nevertheless continue to adhere to those commitments, and indeed hold them to a maximum degree of certainty. However, I think the attitude I have recommended we take towards our hinge commitments is properly described as one of intellectual humility, rather than dogmatism. ‘Dogmatism’ carries with it the implication that the dogmatic are *culpably* unresponsive to reasons: the dogmatic *improperly* refuse to believe in accordance with their evidence. It strikes me that one can only refuse to do what one (thinks one) can do. We cannot willingly lower our confidence in our hinge commitments, even in the face of purported counter-evidence – thus, we cannot improperly refuse to do so. We are epistemically innocent with respect to the rational non-responsiveness of our hinge commitments. By contrast, the attitude of humility I identify - that of realizing the groundlessness of the hinge commitments – *is* under the voluntary control of agents. The attitude involves a recognition of a particular sort of cognitive limit. When it comes to hinge disagreement, I suggest that one can exhibit intellectual humility by doing what it takes to recognize a limit on one's dialectical position. Eventually, reasons give out, and when they do, we should see that we are all equals at least in the groundlessness of our believing.

6.5 Conclusion

In this Chapter, I defended the possibility of moral knowledge against a variety of skeptical arguments. As we saw, the biosemantic framework adopted in Chapter 5 contributes to the response to several skeptical arguments, including the evolutionary debunking argument, and the semantic competence version of the argument from disagreement. I then identified what I take to be the strongest argument from disagreement, namely that deep moral disagreement rationally compels a strong but limited skepticism about morality. In order to respond to this problem, I defended a version of hinge epistemology according to which deep disagreement does not rationally compel one to lower one's initial confidence in one's core moral commitments. The efforts of this chapter do not constitute a definitive proof of the possibility of moral knowledge, as they rely on a controversial hinge epistemology, and because there may yet be other skeptical arguments that need to be taken seriously. Still, I think we can end on an optimistic note.

In the broader context of the dissertation, the contribution of this chapter is twofold; it explains the possibility of moral knowledge, and accounts for the extent and depth of moral disagreement. This finally accomplishes my proposed task at the start of the dissertation to provide an account of ethical thought and discourse that at once accounts for the cognitivist appearances of ethical thought and discourse and the distinctive features of the ethical domain. The offered hybrid theory relies on nonstandard ideas in the philosophy of mind, language, and epistemology relevant to metaethics; (i) a neo-expressivist account at odds with standard ways of laying out the logical space of metaethical views; (ii) a biosemantic framework that rejects certain standard assumptions in the philosophy of language as applied to metaethics; (iii) a radical hinge epistemology that rejects standard ways of thinking about the relations of rational support between beliefs. That we have turned away from the standard options is no bad thing in itself. I would like to think that to make any real progress on an intellectual problem, one needs to continually reassess the lens through which one examines that problem. This is what I have attempted. If the answers I have offered are ultimately unsatisfactory, I hope they nevertheless cast the questions in a new light.

Appendix: Radical skepticism and moral skeptical hypotheses

I am now in a position to return to the argument from skeptical hypotheses raised in section 6.2.4 and show how hinge epistemology dissolves both the radical skeptical problem and its moral analogue. Hinge epistemology would be an insufficient response to the radical skeptical problem if it only rejected a central component of the skeptical problem without explaining why that component fails. The skeptic's urgings initially seem just as compelling as before; how could we be *justified* in holding any proposition to be maximally certain? Even if we are justified in believing ourselves to have knowledge of everyday matters of fact (and so radical skepticism is false), the radical skeptical problem seems to teach us something about the limits of our knowledge. Does hinge epistemology ignore those lessons?

The radical skeptical problem gains traction from its reliance on an intuitively plausible principle of epistemic closure. Since the skeptical challenge concerns rationally-grounded knowledge, we must construe the relevant closure principle in terms of rationally grounded knowledge as well, as follows (Pritchard, 2016, p. 23):

> *The Rationally Grounded Knowledge Closure Principle* (The Closure$_{\text{RK}}$ Principle): If *S* has rationally grounded knowledge that *p*, and *S* competently deduces from *p* that *q*, thereby forming a belief that *q* on this basis while retaining her rationally grounded knowledge that *p*, then *S* has rationally grounded knowledge that *q*.

I agree with Pritchard (and others) that this closure principle is hard to deny – for one thing, it's denial carries with it a commitment to what DeRose (1995) described as 'abominable conjunctions', such as: "I know that I have two hands, and I do not know that I am not a handless brain in a vat". But if we accept the closure principle above, skeptical conclusions appear difficult to avoid. For it is natural to think that one cannot have rationally grounded knowledge that one is not a BIV – after all, everything would appear to one exactly as it does if the BIV hypothesis were true, so there seems to be no possible evidence one could cite in support of one's conviction that it is not. And if one cannot have rationally grounded knowledge that the BIV hypothesis is false, then, through an application of the Closure$_{\mathrm{RK}}$ Principle, one can come to deduce that one cannot have rationally grounded knowledge that one has hands – or any other ordinary matter of fact.

Hinge epistemology provides a distinctive solution to the skeptical problem just discussed. Hinge commitments can neither directly rationally support belief in other propositions, nor are they directly rationally supported by other beliefs. Thus, in Pritchard's terminology, hinge commitments are removed from the scope of rational evaluation. Concerning radical skepticism, this means that hinge commitments do not appropriately figure in applications of the Closure$_{\mathrm{RK}}$ Principle. Hinge commitments cannot be directly supported by other beliefs one might have, *including through competent deduction* from those beliefs – *ipso facto,* one cannot have rationally grounded knowledge of a hinge commitment through a competent deduction from one's other rationally grounded knowledge. Likewise, hinge commitments cannot be called into question through applications of the Closure$_{\mathrm{RK}}$ Principle. Such commitments are thus immune from these sorts of skeptical worries. Hinge epistemology dissolves the radical skeptical problem by showing that the problem is generated by the mistaken but initially alluring idea that rational evaluation is in principle unlimited in scope. The skeptic does not simply extend our ordinary epistemic practices, but distorts them.[33]

[33] An important qualification is in order considering the discussion of inference and the Wishful Thinking problem in Chapter 4. In claiming that hinge commitments cannot enter closure-style inferences, we must understand 'inference' as 'inferr*ing'* -- a causal transition between doxastic states normatively constrained by laws of logic, where one takes the premises to provide rational support for the conclusion. When we consider 'inference' instead as just a logical relation that holds between propositions, independently of any *act* of inferring, it is clear that the

As discussed above, hinge epistemology dissolves the radical skeptical argument by showing that there are principled limits on the scope of rational evaluation. However, this still leaves room for *local*, or *domain-specific*, skeptical problems to gain a foothold. (If hinge epistemology dispelled all forms of skepticism at once, it would prove too much). Domain-specific skepticisms call into question our knowledge in one specific area of inquiry, such as mathematics, religion, witchcraft, etc. So, the topic of this chapter – moral skepticism – is an instance of domain-specific skepticism. And domain-specific skepticisms, unlike the radical skeptical problem, can be motivated by noting certain contrasts between the domain in question and other domains about which we uncontroversially have knowledge. The argument from disagreement is just one instance of a domain-specific motivation for skepticism, as I discussed above in 6.2.4.

Let us consider again the analogue of the radical skeptical problem in the moral case, and whether the dissolution to the radical skeptical problem given above carries over to the moral skeptical problem. Sinnott-Armstrong (2006, pp. 79-80) provides an argument for moral skepticism along the following lines:

propositions towards which we have hinge commitments do legitimately enter inference in this latter sense, since they are semantically and logically just like any other proposition and exhibit logical relations with other propositions. This qualification I think helps clarify and strengthen hinge epistemology, since it wards off possible misunderstanding. Again: it is *not* being maintained that the steps from

1. I do not have rationally grounded knowledge that I am not a BIV, and
2. If I do not have rationally grounded knowledge that I am not a BIV then I do not have rationally grounded knowledge that I have hands, to
3. I do not have rationally grounded knowledge that I have hands.

is an *illogical* inference, at least when we focus on inference as a logical property rather than mental act. Clearly, the argument is valid. Instead, what is being maintained is that given that an agent S has a hinge commitment to the proposition that S has hands, S cannot rationally form a belief in 3, even as a result of accepting the premises and drawing the conclusion, since accepting 3 would call into doubt S's hinge commitment, and it is the nature of hinge commitments that they cannot be doubted in this way. By the same taken, S cannot begin with the premise that she has rationally grounded knowledge that she has hands to conclude that she can have knowledge that she is not a BIV. *Neither* the denial of the BIV hypothesis, nor the claim that one has hands are appropriate items for rationally grounded knowledge.

1. S cannot have rationally grounded knowledge that moral nihilism is false.
2. If S cannot have rationally grounded knowledge that moral nihilism is false, then S cannot have rationally grounded knowledge that it is wrong to torture infants just for fun.
3. But S can have rationally grounded knowledge that it is wrong to torture infants just for fun.

Moral nihilism, similarly to the brain-in-a-vat hypothesis, is a logically consistent view, is compatible with all the non-moral facts, and cannot be refuted without begging the question, so there does not appear to be evidence available that could rule out moral nihilism as false. (2) above is just a particular application of the Closure$_{RK}$ Principle. And (3) represents a purported instance of moral knowledge; if we can have any moral knowledge at all, surely, we could know that it is wrong to torture infants just for fun. Since the skeptical problem here turns on the Closure$_{RK}$ Principle, a piece of uncontroversial moral knowledge, and a moral skeptical hypothesis, it is structurally identical to the closure-based radical skeptical problem.

However, the structural similarity of the problems alone is insufficient to guarantee that hinge epistemology yields an anti-skeptical result in the moral case. What blocks the *radical* skeptical problem from taking hold is the point that necessarily, for any given epistemic agent, she will have some commitments that (i) are preconditions for rational evaluation in general, and (ii) cannot figure in instances of the Closure$_{RK}$ Principle and so are immune to skeptical doubt for her. But when it comes to a specific domain, such as morality, it seems a possibility that some particular agent A may lack any *moral* hinge commitments (without yet lacking in hinge commitments altogether). If so, then it is open to A to subject any moral proposition, including moral skeptical hypotheses, to rational evaluation. Accordingly, it would be open for A to competently deduce that she cannot have rationally grounded moral knowledge from her recognition that she cannot have rationally grounded knowledge that moral nihilism is false. In short; it is seemingly not a necessary condition for being an epistemic agent that one have, specifically, *moral* hinge commitments. So there is nothing about hinge epistemology in general that blocks the *moral* skeptical argument above.

There is, then, at least the following skeptical result: any person who lacks moral hinge commitments cannot rationally avoid moral skepticism, unless some alternative response to the moral skeptical argument above can be given. This result is plausible, since an individual who does not hold any moral propositions to a maximum degree of certainty is naturally described as more skeptical about morality as a whole than one who did hold some moral propositions as maximally certain. At any rate, I think this limited skeptical result should not be too worrying, since I suspect that most if not all epistemic agents do have moral hinge commitments. In Johnson (2019), I suggested that certain basic moral commitments, such as commitment to the proposition that it would be wrong to pour gasoline on a cat and ignite it just for fun, are as likely to stand as hinge commitments for many agents as anything is. Given that we do have moral hinge commitments, the moral skeptical argument above is dissolved in the same way that the radical skeptical problem was. Still, as I shall discuss in the next section, reflection on the moral skeptical problem does not leave everything just as it was – an important lesson we can learn from thinking about skeptical problems is that when it comes to our believing, in morality and otherwise, we ultimately lack rational grounds for our most firmly held views. Rather than plunging us into ‘epistemic vertigo’, as Pritchard (2016) describes it, I think the recognition of this groundlessness should encourage a distinctive sort of intellectual humility, as described in 6.4.4.

Bibliography

Anscombe, G.E.M. (1963). *Intention* (2nd ed.). Ithaca: Cornell University Press.

Artiga, M. (2014). “Teleosemantics and Pushmi-Pullyu Representations”. *Erkenntnis, 79*(S3): 1-22. https://doi.org/10.1007/s10670-013-9517-5

Audi, R. (1998). *Moral Knowledge and Ethical Character*. New York: Oxford University Press.

Audi, R. (2014). “Normative Disagreement as a Challenge to Moral Philosophy and Philosophical Theology.” In M. Bergmann and P. Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution* (pp. 61-79). Oxford: Oxford University Press.

Austin, J. (1962). *How To Do Things with Words*. London: Clarendon Press.

Ayer, A.J. (1936). *Language, Truth, and Logic*. London: Gollancz.

Bach, K. (1999). “The Myth of Conventional Implicature”. *Linguistics and Philosophy, 22*(4): 327-366. https://doi.org/10.1023/a:1005466020243

Balaguer, M. (1998). *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

Barker, S. (2000). “Is Value Content a Component of Conventional Implicature?” *Analysis, 60*(3): 268-279. https://doi.org/10.1093/analys/60.3.268

Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Oxford University Press.

Bar-On, D. (2013). “Origins of Meaning: Must We ‘Go Gricean’?” *Mind and Language, 28*(3): 342-375. https://doi.org/10.1111/mila.2013.28.issue-3

Bar-On, D. (2019). “Crude Meaning, Brute Thought (or: What Are They Thinking?!).” *Journal* *for the History of Analytic Philosophy* 7(2): 29-46. https://doi.org10.15173/jhap.v7i2.3483

Bar-On, D. (2022). "How to Do Things with Nonwords: Pragmatics, Biosemantics, and Origins of Language in Animal Communication". *Biology and Philosophy, 36*(50): 1-25. https://doi.org/10.1007/s10539-021-09824-z

Bar-On, D., and Chrisman, M. (2009). "Ethical Neo-Expressivism". In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 4* (pp. 132-165). Oxford: Oxford University Press.

Bar-On, D., Chrisman, M., and Sias, J. (2014). "(How) Is Ethical Neo-Expressivism a Hybrid View?" In G. Fletcher & M. Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics* (pp. 223-247). Oxford: Oxford University Press.

Bar-On, D., and Sias, J. (2013). "Varieties of Expressivism." *Philosophy Compass, 8*(8): 699-713. https://doi.org/10.1111/phc3.12051

Bedke, M.S. (2009). "Moral Judgment Purposivism: Saving Internalism from Amoralism". *Philosophical Studies, 144*: 189-209. https://doi.org/10.1007/s11098-008-9205-5

Bergman, K. (2019). *Communities of Judgment: Towards a Teleosemantic Theory of Moral Thought and Discourse*. Dissertation, Uppsala University.

Bergman, K. (2021). "Bargaining and Descriptive Content: Prospects for a Teleosemantic Ethics". *Biology and Philosophy, 36*(5): 1-23. https://doi.org/10.1007/s10539-021-09817-y

Bjorklund, F., Bjornsson, G., Eirksson, J., Olindar, R.F., and Strandberg, C. (2012). "Recent Work on Motivational Internalism." *Analysis, 72*(1): 124-137. https://doi.org/10.1093/analys/anr118

Bjornsson, G. (2002). "How Emotivism Survives Immoralists, Irrationality, and Depression".

Southern Journal of Philosophy, 40: 327-344. https://doi.org/southernjphil200240345

Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.

Blackburn, S. (1993a). "How to Be an Ethical Anti-Realist". In *Essays in Quasi-Realism* (pp. 166-181). New York: Oxford University Press. Originally published in *Midwest Studies in Philosophy, 12*(1): 361-375 (1988).

Blackburn, S. (1993b). "Attitudes and Contents". In *Essays in Quasi-Realism* (pp. 182-197). Oxford: Oxford University Press. Originally published in *Ethics, 98*(3): 501-517 (1988).

Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.

Blackburn, S. (2009). "Truth and *A Priori* Possibility: Egan's Charge Against Quasi-Realism." *Australasian Journal of Philosophy, 87*(2): 201-213. https://doi.org/10.1080/00048400802362182

Bloomfield, P. (2001). *Moral Reality*. Oxford: Oxford University Press.

Bloomfield, (2018). "Tracking Eudaimonia." *Philosophy, Theory, and Practice in Biology* 10(2): 1-24. https://doi.org/10.3998/ptpbio.16039257.0010.002

Boghossian, P. (2014). "What is Inference?" *Philosophical Studies, 169*(1): 1-18. https://doi.org/10.1007/s11098-012-9903-x

Boisvert, D. (2008). "Expressive-Assertivism". *Pacific Philosophical Quarterly, 89*: 169-203. https://doi.org/10.1111/papq.2008.89.issue-2

Boisvert, D. (2014). "Expressivism, Nondeclaratives, and Success-Conditional Semantics". In G. Fletcher & M. Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics* (pp. 22-50). Oxford: Oxford University Press.

Boyd, R. (1988). "How to be a Moral Realist." In G. Sayre-McCord (ed.), *Essays on Moral Realism* (pp. 181-228). Ithaca: Cornell University Press.

Brink, D. (1986). "Externalist Moral Realism." *Southern Journal of Philosophy, 24*: 23-40. https://doi.org/10.1111/j.2041-6962.1986.tb01594.x

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Burge, T. (1979). "Individualism and the Mental." *Midwest Studies in Philosophy, 4*(1): 73-122. https://doi.org/10.1111/j.1475-4975.1979.tb00374.x

Bykvist, K., and J. Olson. (2009). "Expressivism and Moral Certitude". *The Philosophical Quarterly, 59*(235): 202-215. https://doi.org/10.1111/j.1467-9213.2008.580.x

Camp, E. (2019). "Metaethical Expressivism" In T. McPherson and D. Plunkett (eds.), *The Routledge Handbook of Metaethics* (pp. 87-101). New York: Routledge.

Carruthers, P. (2013). "Mindreading the Self". In S. Baron-Cohen, M. Lombardo, & H. Tager-Flusberg (eds.), *Understanding Other Minds: Perspectives from Developmental Social Neuroscience* (pp. 467-486). Oxford: Oxford University Press.

Carter, J., and Pritchard, D. (2016). "Intellectual Humility, Knowledge-How, and Disagreement". In C. Mi, M. Slote and E. Sosa (eds.), *Moral and Intellectual Virtues in Western and Chinese Philosophy: The Turn Towards Virtue* (pp. 49-63). New York: Routledge.

Cholbi, M. (2009). "Moore's Paradox and Moral Motivation." *Ethical Theory and Moral Practice, 12*(5): 495-510. https://doi.org/10.1007/s10677-009-9158-6

Chomsky, N. (1995). "Language and Nature". *Mind, 104*(413): 1-61. https://doi.org/10.1093/mind/104.413.1

Chrisman, M. (2013). “Emotivism”. In Hugh LaFollete (ed.), *the International Encyclopedia of Ethics*. https://doi.org/10.1002/9781444367072.wbiee052

Chrisman, M. (2016). *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics.* Oxford: Oxford University Press.

Chrisman, M. (2017). *What is This Thing Called Metaethics?* New York: Routledge.

Christensen, D. (2007). “Epistemology of Disagreement: The Good News”. *Philosophical Review, 116*: 187-218. https://doi.org/10.1215/00318108-2006-035

Christensen, D. (2011). “Disagreement, Question-Begging, and Epistemic Self-Criticism”. *Philosophers' Imprint* 11. http://hdl.handle.net/2027/spo.3521354.0011.006

Camp, E. (2019). “Metaethical Expressivism”. In T. McPherson and D. Plunkett (eds.), *The Routledge Handbook of Metaethics* (pp. 87-101). New York: Routledge.

Coliva, A. (2010). *Moore and Wittgenstein: Scepticism, Certainty, and Common Sense*. London: Palgrave Macmillan.

Coliva, A. (2015). *Extended Rationality: A Hinge Epistemology*. New York: Palgrave Macmillan.

Coliva, A. (2016a). *The Varieties of Self-Knowledge*. London: Palgrave Macmillan.

Copp, D. (1995). *Morality, Normativity, and Society*. Oxford: Oxford University Press.

Copp, D. (2001). “Realist-Expressivism: A Neglected Option for Moral Realism”. *Social Philosophy and Policy, 18*(2): 1-43. https://doi.org/10.1017/s0265052500002880

Copp, D. (2009). “Realist-Expressivism and Conventional Implicature”. In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 4* (pp. 167-202). Oxford: Oxford University Press.

Copp, D. (2014). “Can a Hybrid Theory Have it Both Ways? Moral Thought, Open Questions, and Moral Motivation”. In G. Fletcher & M. Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics* (pp. 51-73). Oxford: Oxford University Press.

Cuneo, T. (2014). *Speech and Morality: On the Metaethical Implications of Speaking*. Oxford: Oxford University Press.

Darwall, S. (1983). *Impartial Reasons*. Ithaca: Cornell University Press.

Darwall, S. (1995). *The British Moralists and the Internal ‘Ought’: 1640-1740*. Cambridge: Cambridge University Press.

Davidson, D. (1979). “Moods and Performances.” In A. Margalit (ed.), *Meaning and Use* (pp. 9-20). Reidel.

Davis, W. (2002). *Meaning, Expression, and Thought*. Cambridge: Cambridge University Press.

DeRose, K. (1995). “Solving the Skeptical Problem”. *Philosophical Review, 104*(1): 1-52. https://doi.org/10.2307/2186011

De Waal, F.B.M. (2012). “The Antiquity of Empathy.” *Science, 336*(6083): 874–876. https://doi.org/10.1126/science.1220999

Dezecache, G., Eskenazi, T., & Grèzes, J. (2016). “Emotional Convergence: A Case of Contagion?” In S.S. Obhi and E.S. Cross (eds.), *Shared Representations: Sensorimotor Foundations of Social Life* (pp. 417-438). Cambridge: Cambridge University Press.

Doherty, R.W. (1997). “The Emotional Contagion Scale: A Measure of Individual Differences.” *Journal Of Nonverbal Behavior, 21*(2): 131–154. https://doi.org/10.1023/A:1024956003661

Dorr, C. (2002). “Non-cognitivism and Wishful Thinking.” *Nous, 36*(1): 97-103. 10.1111/1468-0068.00362

Dowell, J. (2016). "The Metaethical Insignificance of Moral Twin Earth." In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 11*. Oxford: Oxford University Press.

Dreier, J. (1990). "Internalism and Speaker Relativism." *Ethics, 101*(1): 6-26. https://doi.org/10.1086/293257

Dreier, J. (1996). "Expressivist Embeddings and Minimalist Truth." *Philosophical Studies, 83*(1): 29-51. https://www.jstor.org/stable/4320685

Dreier, J. (2004). "Meta-ethics and the Problem of Creeping Minimalism." *Philosophical Perspectives, 18*(1): 23-44. https://doi.org/10.1111/j.1520-8583.2004.00019.x

Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth.

Dunaway, B. (2019). "Realism and Objectivity". In T. McPherson & D. Plunkett (eds.), *The Routledge Handbook of Metaethics* (pp. 135-150). New York: Routledge.

Egan, A. (2007). "Quasi-Realism and Fundamental Moral Error." *Australasian Journal of Philosophy, 85*(2): 205-219. https://doi.org/10.1080/00048400701342988

Ekman, P., and Friesen, W.V. (1971). "Constants Across Cultures in the Face and Emotion". *Journal of Personality and Social Psychology, 17*(2): 124-129. https://doi.org/10.1037/h0030377

Elga, A. (2007). "Reflection and Disagreement". *Noûs, 41*(3): 478-502. https://doi.org/10.1111/j.1468-0068.2007.00656.x

Enoch, D. (2009). "How is Moral Disagreement a Problem for Realism?" *Journal of Ethics, 13*(1): 15-50. https://doi.org/10.1007/s10892-008-9041-z

Enoch, D. (2010). "The Epistemological Challenge to Normative Realism: How Best to Understand It, and How to Cope with It". *Philosophical Studies, 148*(3): 413-438. https://doi.org/10.1007/s11098-009-9333-6

Enoch, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

Eriksson, J. (2006). *Moved by Morality: An Essay on the Practicality of Moral Thought and Talk*. Dissertation. Uppsala: Department of Philosophy.

Eriksson, J. (2009). "Homage to Hare: Ecumenism and the Frege-Geach Problem". *Ethics, 120*: 8-35. https://doi.org/10.1086/606161

Eriksson, J. (2014). "Hybrid Expressivism: How to Think About Meaning". In G. Fletcher & M. Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics* (pp. 149-170). Oxford: Oxford University Press.

Feldman, R. (2005). "Deep Disagreement, Rational Resolutions, and Critical Thinking". *Informal Logic, 25*(1): 13-23. https://doi.org/10.22329/il.v25i1.1041

Feldman, R. (2006). "Reasonable Religious Disagreements". In L. Antony (ed.), *Philosophers Without Gods: Meditations on Atheism and the Secular Life* (pp. 194-215). New York: Oxford University Press.

Finlay, S. (2004). "The Conversational Practicality of Value Judgment". *Journal of Ethics, 8*(3): 205-223. https://doi.org/10.1023/B:JOET.0000031064.73238.8f

Finlay, S. (2005). "Value and Implicature". *Philosophers' Imprint* 5(4): 1-20. http://hdl.handle.net/2027/spo.3521354.0005.004

Finlay, S. (2010). "Recent Work on Normativity." *Analysis, 70*(2): 331-346. https://doi.org/10.1093/analys/anq002

Fletcher, G. (2014). "Moral Utterances, Attitude Expression, and Implicature." In G. Fletcher & M. Ridge (eds.), *Having It Both Ways: Hybrid Theories and Modern Metaethics* (pp. 173-198). Oxford University Press.

Foot, P. (1959). “Moral Beliefs”. *Proceedings of the Aristotelian Society, 59*: 83-104. https://www.jstor.org/stable/4544606

Foot, P. (2001). *Natural Goodness*. Oxford: Oxford University Press.

Fritz, J. (2018). “Conciliationism and Moral Spinelessness”. *Episteme, 15*(1): 101-118. https://doi.org/10.1017/epi.2016.44

Garner, R. (2007). “Abolishing Morality.” *Ethical Theory and Moral Practice, 10*(5): 499-513. https://doi.org/10.1007/s10677-007-9085-3

Geach, P. (1960). “Ascriptivism.” *Philosophical Review, 69*(2): 221-225. https://doi.org/10.2307/2183506

Geach, P. (1965). “Assertion.” *Philosophical Review, 74*(4): 449-465. https://doi.org/10.2307/2183123

Gertler, B. (2011). *Self-Knowledge*. New York: Routledge.

Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

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

Gibbard, A. (2006). “Normative Properties”. In T. Horgan and M. Timmons (eds.), *Metaethics After Moore* (pp. 319-338). Oxford: Oxford University Press.

Gibbard, A. (2012). *Meaning and Normativity*. Oxford: Oxford University Press.

Gibbard, A. (2015). “Global Expressivism and the Truth in Representation”. In S. Gross, N. Tebbens, and M. Williams (eds.), *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism* (pp. 210-223). Oxford: Oxford University Press.

Green, M. (2017). "Assertion". In *Oxford Handbooks Online*. https://doi.org/10.1093/oxfordhb/9780199935314.013.8

Greene, J.D. (2008). "The Secret Joke of Kant's Soul". In W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3* (pp. 35-80). Cambridge: MIT Press.

Grice, H.P. (1957). "Meaning". *The Philosophical Review, 66*(3): 377-388. https://doi.org/10.2307/2182440

Grice, H.P. (1975). "Logic and Conversation". In P. Cole and J. Morgan (eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 41-58). New York: Academic Press.

Grice, H.P. (1989). *Studies in the Ways of Words*. Cambridge: Harvard University Press.

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

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

Harman, G. (1975). "Moral Relativism Defended". *Philosophical Review, 84*(1): 3-22. https://doi.org/10.2307/2184078

Harms, W. (2000). "Adaptation and Moral Realism." *Biology and Philosophy, 15*(5): 699-712. https://doi.org/10.1023/A:1006661726993

Hazlett, A. (2012). "Higher-Order Epistemic Attitudes and Intellectual Humility". *Episteme, 9*(3): 205-223. https://doi.org/10.1017/epi.2012.11

Hopster, J. (2017). "Two Accounts of Moral Objectivity: from Attitude-Independence to Standpoint-Invariance". *Ethical Theory and Moral Practice, 20*: 763-780. https://doi.org/10.1007/s10677-017-9796-z

Horgan, T., and Timmons, M. (1991). "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research, 16*: 447-465. https://doi.org/10.5840/jpr_1991_19

Horwich, P. (1990). *Truth.* Oxford: Oxford University Press.

Horwich, P. (1993). "Gibbard's Theory of Norms." *Philosophy and Public Affairs, 22*(1): 67-78. https://www.jstor.org/stable/2265326

Horwich, P. (1994). "The Essence of Expressivism." *Analysis, 54*(1): 19-20. https://doi.org/10.2307/3328098

Hume, D. (1739/1888). *A Treatise of Human Nature*. L.A. Selby-Bigge (ed.). Oxford: Clarendon Press.

Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F., and Pettit, P. (1997). "A Problem for Expressivism". *Analysis, 58*(4): 239-251. https://doi.org/10.1111/1467-8284.00128

Johnson, D. (2019). "Hinge Epistemology, Radical Skepticism, and Domain Specific Skepticism." *International Journal for the Study of Skepticism, 9*(2): 116-133. https://doi.org/10.1163/22105700-20191302

Johnson, D. (2021). "Deep Disagreement, Hinge Commitments, and Intellectual Humility". *Episteme* 1-20. https://doi.org/10.1017/epi.2020.31

Johnson, D. (forthcoming). "Proper Function and Ethical Judgment: Towards a Biosemantic Theory of Ethical Thought and Discourse". *Erkenntnis*: 1-25. https://doi.org/10.1007/s10670-021-00481-y

Joyce, R. (2001a). "Moral Realism and Teleosemantics." *Biology and Philosophy, 16*(5): 723-731. https://doi.org/10.1023/A:1012280429613

Joyce, R. (2001b). *The Myth of Morality*. Cambridge: Cambridge University Press.

Joyce, R. (2006). *The Evolution of Morality*. Cambridge: MIT Press.

Kalderon, M. (2005). *Moral Fictionalism*. Oxford: Clarendon Press.

Kaplan, D. (1989). “Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals”. In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes From Kaplan* (pp. 481-563). Oxford: Oxford University Press.

Kitcher, P. (2005). “Biology and Ethics”. In D. Copp (ed.), *The Oxford Handbook of Ethical Theory* (pp. 163-185). Oxford: Oxford University Press.

Kohler, S. (2015). “What Is the Problem with Fundamental Moral Error?” *Australasian Journal of Philosophy*, *93*(1): 161-165. https://doi.org/10.1080/00048402.2014.928736

Korsgaard, C. (1996). “Skepticism about Practical Reason”. In *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Lenman, J. (2003). “Disciplined Syntacticism and Moral Expressivism.” *Philosophy and Phenomenological Research*, *66*(1): 32-57. https://doi-org/10.1111/j.1933-1592.2003.tb00242.x

Levy, A., and Levy, Y. (2020). “Evolutionary Debunking Arguments Meet Evolutionary Science”. *Philosophy and Phenomenological Research*, *100*(3): 491-509. https://doi.org/10.1111/phpr.12554

Little, M.O. (1997). “Virtue as Knowledge: Objections from the Philosophy of Mind”. *Noûs*, *31*: 59-79. https://doi.org/10.1111/0029-4624.00035

Locke, J. (1689). *An Essay Concerning Human Understanding*.

Lutz, M. (2014). “The ‘Now What’ Problem for Error Theory”. *Philosophical Studies*, *171*(2): 351-371. https://doi.org/10.1007/s11098-013-0275-7

Lycan, W. (1988). *Judgment and Justification*. Cambridge: Cambridge University Press.

Lynch, M. (2009). *Truth as One and Many*. Oxford: Oxford University Press.

Lynch, M. (2010). "Epistemic Circularity and Epistemic Incommensurability". In A. Haddock, A. Millar, and D. Pritchard (eds.), *Social Epistemology* (pp. 262-278). Oxford: Oxford University Press.

Lynch, M. 2012. *In Praise of Reason: Why Rationality Matters for Democracy*. Cambridge: MIT Press.

Lynch, M. (2013). "Expressivism and Plural Truth." *Philosophical Studies, 163*: 385-401. https://doi.org/10.1007/s11098-011-9821-3

Lynch, M. (2016). "After the Spade Turns: Disagreement, First Principles, and Epistemic Contractarianism". *The International Journal for the Study of Skepticism*, 6: 248-259. https://doi.org/10.1163/22105700-00603010

Lyons, V. and Fitzgerald, M. (2004). "Humor in Autism and Asperger Syndrome". *Journal of Autism and Developmental Disorders, 34*(5): 521-531. https://doi-org/10.1007/s10803-004-2547-8

Mackie, J.L. (1977). "The Subjectivity of Values." In J.L. Mackie, *Ethics: Inventing right and Wrong* (pp. 15-49). New York: Penguin.

McDowell, J. (1978). "Are Moral Requirements Hypothetical Imperatives?". *The Aristotelian Society, Supplementary Volume, 52*: 13-29. https://doi.org/10.1093/aristoteliansupp/52.1.13

McDowell, J. (1979). "Virtue and Reason". *Monist, 62*: 331-350. https://doi.org/10.5840/monist197962319

McNaughton, D. (1988). *Moral Vision: An Introduction to Ethics*. Oxford: Blackwell.

Mele, A. (1996). "Internalist Moral Cognitivism and Listlessness". *Ethics*, *106*: 727-753. https://doi.org/10.1086/233670

Miller, C.B. (2008). "Motivational Internalism". *Philosophical Studies*, *139*: 233-255. https://doi.org/10.1007/s11098-007-9115-y

Millikan, R. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge MA: MIT Press.

Millikan, R. (2004). *Varieties of Meaning*. Cambridge: MIT Press.

Millikan, R. (2005a). "Pushmi-Pullyu Representations." In *Language: A Biological Model*, p. 166-186. Oxford: Oxford University Press. Originally published in *Philosophical Perspectives* 9: 185-200. (1995).

Millikan, R. (2005b). "On Meaning, Meaning, and Meaning". In *Language: A Biological Model* (pp. 53-76). Oxford: Oxford University Press.

Millikan, R. (2010). "On Knowing the Meaning: With a Coda on Swampman." *Mind*, *119*(473): 43-81. https://doi.org/10.1093/mind/fzp157

Millikan, R. (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford: Oxford University Press.

Millikan, R. (2018). "Biosemantics and Words that Don't Represent". *Theoria*, *84*(3): 229-241. https://doi.org/10.1111/theo.12146

Millikan, R. (forthcoming). "Comment on Artiga's 'Teleosemantics and Pushmi-Pullyu Representations'". *Erkenntnis*. https://doi.org/10.1007/s10670-020-00354-w

Mogensen, A. (2016). "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?" *Philosophical Studies*, *173*(7): 1799-1817. https://doi.org/10.1007/s11098-015-0579-x

Morgan, A. (2016). “Hybrid Illocutionary Acts: A Theory of Normative Thought and Language that ‘Has it Both Ways’.” *European Journal of Philosophy*, *25*(3), 785-807. https://doi.org/10.1111/ejop.12161

Moore, G.E. (1903). *Principia Ethica*. Dover Publications.

Moyal-Sharrock, D. (2004). *Understanding Wittgenstein’s On Certainty*. London: Palgrave Macmillan.

Moyal-Sharrock, D. (2016). “The Animal in Epistemology”. *International Journal for the Study of Skepticism*, *6*(2-3), 97-119. https://doi.org/10.1163/22105700-00603003

Olson, J. (2011). “In Defense of a Moral Error Theory.” In M. Brady (ed.), *New Waves in Metaethics* (pp. 62-84). London: Palgrave-MacMillan.

Price, H. (1994). “Semantic Minimalism and the Frege Point.” In S. L. Tsohatzidis (ed.), *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*. New York: Routledge.

Price, H. (2013). “Part I: The Descartes Lectures.” In H. Price (ed.), *Expressivism, Pragmatism, and Representationalism* (pp. 1-64). Cambridge: Cambridge University Press.

Pritchard, D. (2012). “Wittgenstein and the Groundlessness of Our Believing”. *Synthese, 189*(2): 255-272. https://doi.org/10.1007/s11229-011-0057-8

Pritchard, D. (2016). *Epistemic Angst: Radical Skepticism and the Groundlessness of Our Believing*. Princeton: Princeton University Press.

Pritchard, D. (2017). “Disagreements, of Beliefs and Otherwise”. In C.R. Johnson (ed.), *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public* (pp. 22-39). New York: Routledge.

(2018). “Wittgensteinian Hinge Epistemology and Deep Disagreement.” *Topoi*: 1-9. https://doi.org/10.1007/s11245-018-9612-y

Putnam, H. (1975). “The Meaning of ‘Meaning’”. *Minnesota Studies in the Philosophy of Science, 7*: 131-193. https://hdl.handle.net/11299/185225

Parfit, D. (1998). “Reasons and Motivation”. *The Aristotelian Society, Supplementary Volume, 71*: 99-130. https://doi.org/10.1111/1467-8349.00021

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Rawls, J. (2005). *Political Liberalism* (expanded edition). New York, NY: Columbia University Press.

Railton, P. (1986). “Moral Realism”. *The Philosophical Review, 95*(2): 163-207. https://doi.org/10.2307/2185589

Ridge, M. (2006). “Ecumenical Expressivism: Finessing Frege.” *Ethics, 116*(2): 302-336. https://doi.org/10.1086/498462

Ridge, M. (2007). “Expressivism and Epistemology: Epistemology for ecumenical expressivists.” *Aristotelian Society Supplementary Volume, 81*(1): 83-108. https://doi.org/10.1111/j.1467-8349.2007.00152.x

Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.

Ridge, M. (2015). “I Might be Fundamentally Mistaken.” *Journal of Ethics and Social Philosophy, 9*(3): 1-21. https://doi.org/10.26556/jesp.v9i3.92

Roskies, A. (2003). “Are Ethical Judgments Intrinsically Motivational? Lessons from ‘Acquired Sociopathy’”. *Philosophical Psychology, 16*: 51-66. https://doi.org/10.1080/0951508032000067743

Rowland, R. (2016). “The Epistemology of Moral Disagreement”. *Philosophy Compass, 12*(2):

1-16. https://doi.org/10.1111/phc3.12398

Ruse, M. & E.O. Wilson (1986). “Moral Philosophy as Applied Science”. *Philosophy, 61*: 173-192. https://doi.org/10.1017/S0031819100021057

Saslow, E. (2016, October 15). “The white flight of Derek Black.” *The Washington Post*. Retrieved from: https://www.washingtonpost.com/national/the-white-flight-of-derek-black/2016/10/15/ed5f906a-8f3b-11e6-a6a3-d50061aa9fae_story.html?noredirect=on&utm_term=.3baf1bc9c30d

Schroeder, M. (2008a). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

Schroeder, M. (2008b). “Expression for Expressivists.” *Philosophy and Phenomenological Research, 76*(1): 86-112. https://doi.org/10.1111/j.1933-1592.2007.00116.x

Schroeder, M. (2009). “Hybrid Expressivism: Virtues and Vices”. *Ethics, 119*: 257-309. https://doi.org/10.1086/597019

Schroeder, M. (2010). *Non-Cognitivism in Ethics*. New York: Routledge.

Schroeder, M. (2014). “The Truth in Hybrid Semantics”. In G. Fletcher & M. Ridge (eds.), *Having it Both Ways: Hybrid Theories and Modern Metaethics* (pp. 273-293). Oxford: Oxford University Press.

Schroeter, L., and Schroeter, F. (2019). “Metasemantics and Metaethics”. In T. McPherson and D. Plunkett (eds.), *The Routledge Handbook of Metaethics* (pp. 519-535). New York: Routledge.

Schwitzgebel, E. (2006). “The Unreliability of Naive Introspection”. *Philosophical Review, 117*(2): 245-273. https://doi.org/10.1215/00318108-2007-037

Searle, J. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Sellars, W. (1969). “Language as Thought and as Communication”. *Philosophy and Phenomenological Research, 29*(4): 506-527. https://doi.org/10.2307/2105537

Shafer-Landau, R. (2003). *Moral Realism: A Defense*. Oxford: Oxford University Press.

Shields, M. (2019). “On the Pragmatics of Deep Disagreement”. *Topoi* 1-17. https://doi.org/10.1007/s11245-018-9602-0

Sinclair, N. (2006). “The Moral Belief Problem.” *Ratio, 19*(2): 249-260. https://doi-org/10.1111/j.1467-9329.2006.00323.x

Sinclair, N. (2007). “Expressivism and the Practicality of Moral Convictions”. *The Journal of Value Inquiry, 41*: 201-220. https://doi.org/10.1007/s10790-007-9080-x

Sinclair, N. (2012). “Metaethics, Teleosemantics, and the Function of Moral Judgment.” *Biology and Philosophy, 27*(5): 639-662. https://doi.org/10.1007/s10539-012-9316-4

Sinnott-Armstrong, W. (2006). *Moral Skepticisms*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (2019). “Moral Skepticism”. In N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/skepticism-moral/

Skarsaune, K.O. (2011). “Darwin and Moral Realism: Survival of the Iffiest”. *Philosophical Studies, 152*(2): 229-243. https://doi.org/10.1007/s11098-009-9473-8

Smart, J.J.C. (1984). *Ethics, Persuasion, and Truth*. Boston: Routledge & Kegan Paul.

Smith, M., Jackson, F., and Oppy, G. (1994). “Minimalism and Truth Aptness.” *Mind, 103*(411): 287-302. https://doi.org/10.1093/mind/103.411.287

Smith, M. (1994). *The Moral Problem*. Malden: Blackwell Publishing.

Snowdon, P. (2012). “How to Think About Phenomenal Self-Knowledge”. In A. Coliva (ed.), *The Self and Self-Knowledge* (pp. 243-262). Oxford: Oxford University Press.

Sober, E., & D.S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.

Stainton, R. (2016). “Full-on Stating.” *Mind and Language, 31(4),* 395-413. https://doi.org/10.1111/mila.12112

Stevenson, C.L. (1937). “The Emotive Meaning of Ethical Terms”. *Mind, 46*(181): 14-31.

Stevenson, C.L. (1944). *Ethics and Language*. New Haven, CT: Yale University Press.

Stocker, M. (1979). “Desiring the Bad: An Essay in Moral Psychology”. *The Journal of Philosophy, 76*: 738-753. https://doi.org/10.2307/2025856

Stokke, A. (2013). “Lying and Assertion.” *Journal of Philosophy, 110*(1): 33-60. https://doi.org/10.5840/jphil2013110144

Stoljar, D. (1993). “Emotivism and Truth Conditions.” *Philosophical Studies, 70*(1): 81-101. https://doi.org/10.1007/BF00989663

Strandberg, C. (2011). “The Pragmatics of Moral Motivation”. *The Journal of Ethics, 15*(4): 341-369. https://doi.org/10.1007/s10892-011-9106-2

Strandberg, C. (2012). “A Dual-Aspect Account of Moral Language”. *Philosophy and Phenomenological Research, 84*(1): 87-122. https://doi.org/10.1111/j.1933-1592.2010.00447.x

Street, S. (2006). “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies, 127*(1): 109-166. https://doi.org/10.1007/s11098-005-1726-6

Sturgeon, N. (1985). “Moral Explanations”. In D. Copp and D. Zimmerman (eds.), *Morality, Reason, and Truth* (pp. 49-78). Totowa: Rowan and Allanheld.

Svavarsdottir, S. (1999). “Moral Cognitivism and Motivation”. *The Philosophical Review, 108*:

161-219. https://doi.org/10.2307/2998300

Tersman, F. (2006). *Moral Disagreement*. Cambridge: Cambridge University Press.

Timmons, M. (1999). *Morality Without Foundations*. Oxford: Oxford University Press.

Tollhurst, W. (1995). "Moral Experience and the Internalist Argument Against Moral Realism". *American Philosophical Quarterly, 32*: 187-194.

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge: Harvard University Press.

Toppinen, T. (2017). "Hybrid Accounts of Ethical Thought and Talk." In T. McPherson & D. Plunkett (eds.), *The Routledge Handbook of Metaethics* (pp. 243-259). New York: Routledge.

Tresan, J. (2006). "De Dicto Internalist Cognitivism". *Noûs, 40*: 143-165. https://doi.org/10.1111/j.0029-4624.2006.00604.x

Tresan, J. (2009). "Metaethical Internalism: Another Neglected Distinction." *Journal of Ethics, 13*(1): 51-72. https://doi.org/10.1007/s10892-008-9042-y

Unwin, N. (1999). "Quasi-realism, Negation, and the Frege-Geach Problem." *Philosophical Quarterly, 49*(196): 337-352. https://doi.org/10.1111/1467-9213.00146

Unwin, N. (2001). "Norms and Negation: A Problem for Gibbard's Logic." *Philosophical Quarterly, 51*(202): 60-75. https//doi.org/10.1111/1467-9213.00214

van Roojen, M. (1996). "Expressivism and Irrationality." *The Philosophical Review, 105*(3): 311-335. https://doi.org/10.2307/2185703

van Roojen, M. (2015). *Metaethics: A Contemporary Introduction*. New York: Routledge.

Vavova, K. (2014). "Moral Disagreement and Moral Skepticism". *Philosophical Perspectives, 28*(1): 302-333. https://doi.org/10.1111/phpe.12049

Vavova, K. (2015). “Evolutionary Debunking of Moral Realism.” *Philosophy Compass, 10*(2): 104-116. https://doi.org/10.1111/phc3.12194

Vavova, K. (2021). “The Limits of Rational Belief Revision: A Dilemma for the Darwinian Debunker”. *Nous, 55*(3): 717-734. https://doi.org/10.1111/nous.12327

Wallace, R.J. (2006). “Moral Motivation”. In J. Dreier (ed.), *Contemporary Debates in Moral Theory* (pp. 182-195). Oxford: Blackwell.

Wedgewood, R. (2014). “Moral Disagreement Among Philosophers.” In M. Bermgann and P. Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution* (pp. 23-39). Oxford: Oxford University Press.

Whitcomb, D., Battaly H., Baehr J., and Howard-Snyder D. (2017). “Intellectual Humility: Owning Our Limitations”. *Philosophy and Phenomenological Research, 94*(3): 509-539. https://doi.org/10.1111/phpr.12228

Wielenberg, E.J. (2010). “On the Evolutionary Debunking of Morality”. *Ethics, 120*(3): 441-464. https://doi.org/10.1086/652292

Wielenberg, E.J. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford: Oxford University Press.

Wiggins, D. (1991). “Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs”. *Proceedings of the Aristotelian Society*, *91*: 61-85. http://www.jstor.org/stable/4545127

Williams, M. (1991). *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*. Cambridge: Blackwell.

Williamson, T. (2001). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Wisdom, J. (2017). “Proper Function Moral Realism.” *European Journal of Philosophy, 25*(4): 1660-1674. https://doi.org/10.1111/ejop.12252

Wittgenstein, L. (1969). *On Certainty*. G.E.M. Anscombe and G.H. von Wright (eds.); D. Paul and G.E.M. Anscombe (trans.). New York, NY: Harper and Rowe.

Wittgenstein, L. (1973). *Philosophical Investigations*, 3rd edition. G.E.M. Anscombe (trans.). New York: Prentice Hall.

Woods, J. (2014). "Expressivism and Moore's Paradox". *Philosopher's Imprint, 14*: 1-12. http://hdl.handle.net/2027/spo.3521354.0014.005

Wright, C. (1986). "Facts and Certainty". *Proceedings of the British Academy, 71*: 429-472.

Wright, C. (1992). *Truth and objectivity*. Cambridge: Harvard University Press.

Wright, C. (2004). "Warrant for Nothing (and Foundations for Free)?" *Aristotelian Society Supplementary Volume, 78*(1): 167-212. https://doi.org/10.1111/j.0309-7013.2004.00121.x

Wright, C. (2014). "On Epistemic Entitlement (II): Welfare State Epistemology". In D. Dodd and E. Zardini (eds.), *Scepticism and Perceptual Justification* (pp. 213-247). Oxford: Oxford University Press.

Wright, C. (2015). "The Reality of Privileged Access". In S. Goldberg (ed.), *Externalism, Self-Knowledge, and Skepticism: New Essays* (pp. 49-74). Cambridge: Cambridge University Press.

Yalcin, S. (2021). "Modeling With Hyperplans". In B. Dunaway and D. Plunkett (eds.), *Meaning, Decision, and Norms: Themes on the Work of Allan Gibbard* (pp. 307-341). Ann Arbor: Maize Books.