

Higher-Order Expressivism: The Dyadic Nature of Moral Judgement



Graham Bex-Priestley
The University of Sheffield

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Submission date: September 2018

For Keanu Reeves

Declaration

This doctoral thesis is a presentation of my own research. It has not been submitted for another qualification to this or any other university. Wherever contributions of others are involved, every effort has been made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative discussions.

Preface

I've had a lot of fun writing this thesis. A PhD is a long project, and it's easy to get so bogged down in the details that we forget how privileged we are. For the last five years I have studied (and taught) my favourite subject. It has of course been a difficult endeavour, but a moment's reflection reveals how vital returning to philosophy has been for my happiness. For me, philosophy is not merely a subject but an outlet, a meditation. It is where I can focus on just *one* thing. I consider myself exceptionally lucky for all the opportunities I have had throughout my life to get me to this point, and I am humbled by the thought that so few of us are so fortunate.

The ideas that sparked this thesis began before I applied for university. At times I would feel guilty for not giving the money I'd earned at my Saturday job to those who needed it more, and I wondered whether I was in breach of a genuine moral requirement. This gave rise to more abstract questions about the foundations of morality. *Are there any objective moral demands? Is morality real or did we just make it up? Would it matter either way?* Other philosophical topics intrigued me, particularly free will and consciousness – like morality, notoriously tricky to mesh with a scientific worldview – but I found morality the most mysterious.

It was Helen Frowe who gave me my first taste of formal metaethics in my second year of university. Helen was openly partial to Jonathan Dancy's non-natural realism. While I didn't want to reduce ethics to "just doing science," I couldn't bring myself to believe in a special non-natural realm of moral facts. Quite apart from its queerness and the threat from widespread moral disagreement, realism (or my internal caricature of it) seemed to give no place to the human sentiment. So when I was introduced to the theory that moral sentences are not true or false but expressions of our emotional attitudes, I was sold.

Unsurprisingly my views have changed in the decade that has passed since then. Yet the core remains the same: I still defend the idea that moral

sentences express our moral sentiments. Perhaps the biggest change in that time is my sympathy towards realistic metaethics. Whereas I once would describe myself as ‘shamelessly anti-realist,’ I now believe expressivism leads fluidly to a form of *quasi*-realism according to which there are moral facts that we can discover through non-scientific means. This thesis is the result of a struggle to keep ethics what it is without recourse to anything non-natural.

All of this could not have been accomplished without a great number of people, and I would like to acknowledge some of them here. My three supervisors have been incredibly helpful in different and complimentary ways. Jimmy Lenman has given me an enormous amount of his time and he always struck the right balance of guidance and allowance for academic freedom. I am grateful to have had one of the world’s leading expressivists supervise my PhD. Jimmy’s help was not restricted to matters directly about my thesis. He has aided me with submissions to journals, conference talks, job applications and problems of a more personal nature. I could not have hoped for better support.

To my first secondary supervisor Holly Lawford-Smith, thanks are due for keeping an eye on the bigger picture, for detailed help with a paper that has now become chapter 4, and for letting me stay at her empty Edinburgh flat when I visited for a conference. To my second secondary supervisor Yonatan Shemmer, I am much obliged for his incisive comments on my written work and for the many in-depth conversations about things that puzzle us both about expressivism. The time we spent co-authoring our paper on disagreement has also been extremely useful for getting to grips with issues that appear in this thesis.

I heartily thank my friends Stephen Ingram and Lewis Brooks. It’s hard to overstate the magnitude of the influence these two philosophers have had on my thinking. Both have continually sharpened my understanding, challenged my sloppy arguments, shared with me their astoundingly vast knowledge of the literature, and have been exceedingly generous in giving up their time to read and comment on a draft of this thesis; without them it would have looked very different. In fact it was Stephen who, after hearing about my interest in metaethics, sent me a selection of key readings before I started my MA in 2013, and among them was Mike Ridge’s famous 2006 paper introducing ecumenical expressivism. I was hooked. Little did I know that I would be defending a version of it in my viva five years later.

This brings me to Mike Ridge himself. It is obvious throughout the thesis how much of an intellectual debt I owe him. There are references to his work in all five chapters, and while I often argue against him, these are usually small brush strokes in a big picture. I like to think we're on the same side. Further to his written work, I thank him for welcoming me so warmly to Edinburgh and for helping me with my own philosophical inquiry at various points. Even before I had met him in person I vividly remember chatting with him (nervously!) on the phone about expressivism and engaging in helpful lengthy email exchanges about my PhD proposal. It is wonderful that the chief ecumenical expressivist is equally kind to and enthusiastic about the expressivists in the early stages of their careers.

Similar things can be said about Teemu Toppinen. I was convinced by his 2013 paper in which he argued for a kind of ecumenical expressivism that he called the *higher-order state view* and that I call, for shallow aesthetic reasons, *higher-order expressivism*. My thanks go out to him not only for his influential research but for the comments he has given and the discussions we have had about several pieces of my work that now appear in this thesis. It's nice to have another metalhead to look up to.

I am grateful to the University of Sheffield and the White Rose College of the Arts and Humanities for funding this research and providing such a thriving environment for teaching and learning. This environment includes the large and splendid department of postgraduates who have listened to me pose most of the arguments you see in this thesis, and whose feedback I am thankful for. The other audiences I am indebted to are the audience of New Directions for Expressivism (a conference co-organised by Stephen Ingram and myself) and the audience of the 2018 annual conference for the British Society for Ethical Theory, both of whom heard me present what has now become chapter 5 and a forthcoming paper in *Ethical Theory and Moral Practice* (printed here with the publisher's kind permission).

A great deal of help and encouragement has come from outside of academia. Without the support of my parents Karen and Nigel, as well as my grandparents Betty, Danny, and the late Grace and Ken, I would not have achieved the marks I did at undergraduate and MA level that enabled me to undertake this PhD. My sister Tara has been a pillar of emotional support, being both a shoulder to lean on and a fountain of good humour.

Last and certainly not least, I must acknowledge the tremendous role that my spouse Novenka has played. I cannot describe everything they have

done without doubling the length of this preface. It has ranged from giving strategic advice about the thesis, to proof reading my work, to listening to me practise my presentations. They were a central part in the smooth running of *New Directions for Expressivism*. They are also responsible for my return to philosophy at the postgraduate level. Novenka could see I was unsatisfied with my career teaching maths, so when Jenny Saul advised I had a decent chance of getting funding for an MA in philosophy, they convinced me to go for it. Therefore to Novenka I express my deepest gratitude for being willing to sacrifice stability and financial security for my happiness... and for keeping us stocked with sci-fi.

Contents

Introduction	1
~1~ What is Expressivism?	5
1. Expressing States of Mind	6
2. Quasi-Realism.....	21
3. Semantics and Metasemantics	33
4. Halting Minimalism	50
~2~ The Problem with Purity	65
1. The Frege-Geach Problem	66
2. Minimalist Solutions.....	71
3. Higher Order Attitude Solutions.....	75
4. Inferentialist Solutions	79
5. Structural Solutions	87
6. Hybrid Solutions.....	95
~3~ Metaphysics for Expressivists	111
1. Robust, Relaxed, or Quasi?.....	112
2. Metaphysical Sidesteps	118
3. The Reducibility of Moral Properties.....	125
4. Metaphysical Options for Expressivists	128
5. For They Shall be Called the Children of God.....	135
6. Point Scoring.....	147

~4~ An Assertoric Theory of Disagreement.....	151
1. Disagreement in Attitude	152
2. Disagreement in Prescription.....	158
3. The Assertoric Theory.....	163
4. Why This Matters.....	170
~5~ Error and the Limits of Quasi-Realism	177
1. The Quasi-Realist Project.....	178
2. Expressivism and Error.....	180
3. Immunity to Fundamental Error	183
4. The Limits of Quasi-Realism.....	190
Summary	195
References.....	197

Introduction

Morality is peculiar. Suppose you're talking about an upcoming election with Gertrude. She says to you "UKIP is the best party. They have the right values and they won't betray this great country. I ought to vote for them. More than that, it is my *duty* to vote for them." After the election, you find out to your relief (being no UKIP supporter yourself) that Gertrude did not vote for UKIP but some other party. Still, you want some explanation. "Why didn't you vote for UKIP?" you ask Gertrude. "Did you change your mind? Did someone bribe you? Did your hand slip?" "No, nothing like that," she replies. "I still think UKIP is the best. I have never stopped thinking it. Their candidate here is unsurpassed too. I just voted for someone else, that's all."

There's something odd about Gertrude. It's natural to think that she's confused or else she's lying, to you or to herself. Normally we experience some pressure to act in line with our moral judgements, even if it's regularly overcome by fear, selfishness, laziness, or other forces. This is evidence that moral judgements are desire-like in nature. The case of Gertrude seems analogous to the so-called "friend" who keeps telling you they're planning to visit you in hospital but day after day never shows up: you doubt your friend ever had the intention at all, and you doubt whether Gertrude *really* believes she ought to vote for UKIP at heart.

On the other hand, moral judgements have the features of beliefs. They can be true or false, and Gertrude (you are convinced) said something false when she said she had a moral duty to vote for UKIP. This contrasts with desires. Desires aren't the kinds of things we readily call true or false, unlike beliefs. This is the dyadic nature of moral judgement. Our moral convictions seem to have the power to motivate us in the way our desires do, and at the same time they appear to be assessable as true or false.

This thesis is ultimately a defence of *higher-order expressivism* according to which moral judgements are hybrid mental states composed of beliefs and desires. However, only two of the chapters (2 and 3) contain arguments

in favour of this specific kind of expressivism. The first chapter argues for a characterisation of expressivism that will be useful for friends and foes alike, and the final two chapters address problems that are common to all expressivists.

There is a trend in metaethics to speak of “the normative” as a whole as opposed to just one aspect of it. Expressivists, error theorists, realists and so on often adopt the same theory towards *all* normative thought and discourse. There is something attractive about unification. Yet I do not have strong intuitions about the domain of, say, non-moral practical reasons. Presumably it’s because I do not feel the same pull of objectivity. In the moral realm, you generally shouldn’t kill people no matter how much you want to. But in the non-moral realm, if you want to dance, go ahead and dance! In this thesis I focus on morality and leave open whether expressivism extends into every other normative domain.

The five chapters are far from equal in length. They have ended up in order of size from largest to smallest. Chapter 1 has four sections, each of which is long enough to be its own chapter. I decided not to split the chapter into four separate ones because every section works towards answering the same question: *What is Expressivism?* Each of the five chapters is intended to be self-contained. In principle you could read them in any order.

After we find out what expressivism is in chapter 1, I then turn to the problems with pure versions of expressivism according to which moral judgements are only desire-like states. Yes, chapter 2, *The Problem with Purity*, is about the fearsome Frege-Geach problem. I argue that pure expressivists are unlikely to be able to solve it but higher-order (and other hybrid) expressivists can.

In chapter 3, *Metaphysics for Expressivists*, I make a case for the inescapability of metaphysics even on the expressivist picture. I argue that higher-order expressivism has another advantage over pure expressivism in this arena before showing why we must think moral properties are reducible to bog-standard descriptive ones. I end by arguing that this expressivist take on moral naturalism is better than cognitivist versions of naturalism.

Chapter 4 is about how to understand moral disagreement. If moral judgements are desire-like, why does it make sense to say that two people morally disagree? A difference in desire doesn’t normally seem like a

disagreement. I propose *An Assertoric Theory of Disagreement* whereby people disagree if they assert inconsistent things (and the problem of inconsistency is addressed in chapter 2).

Finally, chapter 5, *Error and the Limits of Quasi-Realism*, is about how to understand the thought that our moral beliefs might be mistaken. I defend the standard expressivist construal of it being a case of thinking we might change our minds after we've gathered more evidence, cohered our beliefs and generally improved our epistemic situation. I then argue that this yields a transcendental argument against scepticism: it tells us that it is incoherent to believe we might be utterly *unable* to access the moral truth. The upshot is that expressivists cannot "mimic" realism completely, but if this simply means that our door is closed to radical scepticism whereas theirs is open, this is all the better for expressivism.

~1~

What is Expressivism?

In this chapter I explore previous characterisations of expressivism and I argue all of them are wrong. In its early days expressivism, then called *emotivism*, was thought of as the thesis that moral sentences are not truth-apt but instead express mental states other than belief. Modern expressivists disagree. Now we expressivists think there can be true moral sentences and that the mental states expressed by such sentences are beliefs. What is the contemporary expressivist view? According to Ralph Wedgwood (2007), Mark Schroeder (2008) and others, it is a semantic thesis about the meaning of moral terms. Contrastingly, Alex Silk (2013), Alejandro Pérez Carballo (2014), Michael Ridge (2014), Matthew Chrisman (2015) and Sebastian Köhler (2017) construe expressivism as a *metasemantic* thesis about *how* moral terms get their meaning. Often we also see expressivists characterised as rejecting any metaphysical foundation for morality. In this chapter I argue against all of this, along with a few other attempts to solve the problem of creeping minimalism. The end result, apart from implying all the encyclopaedia entries need updating, is that it is no longer clear what makes expressivism attractive in the first place.

In §1 I discuss the Humean distinction of beliefs and desires and how to understand the sense in which moral sentences might express desires. In §2 I go on to examine the problem of creeping minimalism, which has the consequence that expressivism cannot be characterised as the denial that moral sentences express beliefs. In §3 I argue against thinking of expressivism as a linguistic thesis in semantics or metasemantics. §4 is dedicated to some attempts to characterise expressivism in light of creeping minimalism: Simon Blackburn's idea that the philosophical journey matters more than the destination, James Dreier's (2004) "explanation" explanation,

and Jamin Asay's (2013) metaphysical attempt all fail. The answer closest to the truth is Neil Sinclair's (2006) suggestion that expressivism gives a different functional characterisation of the mental states involved in moral judgement, but it is still fundamentally incorrect because so-called 'besire' theorists can give precisely the same characterisation. To get around this, I propose expressivism should be understood as the view that moral beliefs are *necessarily* desire-like states.

1. Expressing States of Mind

Expressivism certainly involves a thesis about the kinds of mental states expressed by moral sentences. According to expressivism moral sentences express desire-like states, which might be emotional states of approval and disapproval, or commitments or goals of some sort, or plans to live, act, and feel in certain ways. In this section I will cash this out in more detail. Once I have done so, this will be far from the full story. While many metaethical theorists, called pure cognitivists, would deny that moral sentences express desire-like states, there are now several hybrid cognitivists who agree with the expressivist thesis just mentioned. Nevertheless, this will be a useful place to begin. First, let us examine a division of mental states that is very widely accepted, especially by expressivists, and yet one which I will ultimately argue pure expressivists should deny.

1.1 The Humean Theory of Motivation

A moment ago I took a sip of my water. This action of mine could be explained by the conjunction of two of my mental states at the time: my desire to quench my thirst and my belief that taking a sip of my water would quench my thirst.¹ One without the other would not be enough to motivate me to do what I did. According to the Humean Theory of Motivation, this is always the case.² Actions are driven by desires and guided by beliefs.

¹ Perhaps it also requires that I have "put the relevant desire and belief together" (Smith 1994: 92) in some sense. If it is true that motivation requires connecting the dots, as it were, then hybrid expressivists such as Toppinen (2013), Ridge (2014) and myself can afford a weaker form of motivational internalism than pure expressivists, which I draw attention to in §6 of chapter 3. I ignore this "putting together" aspect in this chapter.

² The theory is so called because its advocates are typically broadly in line with Humean traditions, but it is not clear that Hume himself would have subscribed to it. For instance,

Beliefs by themselves do not motivate – simply believing that water quenches thirst will not move me to do anything if none of my desires involve quenching existing thirst.

In the above paragraph I used the words ‘belief’ and ‘desire’ loosely, but within the Humean theory they are terms of art and should not be held hostage to ordinary language use. Indeed, modern expressivists think moral judgements are beliefs in the ordinary language sense but desires in the Humean sense. Let us therefore introduce terminology to keep things clear: from here on, the lower case ‘belief’ and ‘desire’ will refer to whatever ordinary language users refer to with the terms, and the upper case ‘BELIEF’ and ‘DESIRE’ will be used as specified by the Humean theory.

The most common way to elucidate these Humean notions is in terms of “directions of fit,” the method (but not the phrase) often attributed to Elizabeth Anscombe (1957). If there is a mismatch between the content of our propositional attitude and the truth of it, BELIEFS are to change to fit the way the world is, whereas DESIRES generally motivate us change the world to fit them. For example, if I BELIEVE it is raining and it becomes apparent that it isn’t, I normally change my mind. If I DESIRE coffee and I notice I don’t have any, I normally retain that mental state and change the world by getting coffee. Of course that is only what *normally* happens. Sometimes things go wrong; I see it but don’t believe it, or I want it but don’t go and get it. To capture what constitutively distinguishes the two kinds of state, descriptive theories (Smith 1987, 1994) speak of the *tendencies* or *dispositions* – not guarantees – of mental states to change or move us whereas normative theories (Gregory 2012) cite *reasons* to change or move us. Here is a (very rough) characterisation of Michael Smith’s descriptive view:

A mental state that p is a BELIEF if and only if the apprehension that not p tends to cause the extinction of that state.

A mental state that p is a DESIRE if and only if it yields a disposition to make or keep p the case.³

Here is Alex Gregory’s normative view:

Setiya (2004: 372) argues that “Hume does not accept the Humean theory of motivation.” I am not concerned with exegesis in this thesis.

³ See especially Smith (1994: 115).

For a mental state that P to be a [BELIEF] that P is for it to be the case that if not-P, the fact that not-P gives you an objective reason to abandon the state.

For a mental state that P to be a [DESIRE] that P is for it to be the case that if not-P, the mental state gives you a subjective reason to make P the case. (Gregory 2012: 18, italics in original removed.)

Is the normative view of mental states closed off to expressivists, who are trying to explain (at least one aspect of) normativity? One can be forgiven for the suspicion but there is no vicious circularity here, or if there is it is general: Gregory's view implies metaethics is part of normative inquiry, regardless of which metaethical theory is on the table. If Gregory's normative view is true and expressivism about the reasons he cites is also true, then it simply means that saying 'moral sentences express DESIRES' is itself an expression of some DESIRE. If realism is true, it means saying 'moral sentences express BELIEFS' expresses a BELIEF. Why should we worry about one more than the other? Perhaps you have the intuition that there is a fact of the matter about what mental state moral judgements are, or that the realist and the expressivist have a genuine disagreement rather than a mere clash of DESIRES. However, this is to levy general attacks on the expressivist theory. In §2 we will see why expressivists believe in moral facts and in chapter 4 of this thesis I present an expressivist-friendly theory of disagreement. If these succeed, we can continue thinking there are (normatively laced) facts about the mental states we express with moral sentences and that there are genuine metaethical disagreements.⁴

I will not be concerned with adjudicating between the descriptive and normative theories of mental states or any further sub-divisions in those camps. The main point is that however BELIEFS and DESIRES are to be understood, on the Humean picture motivation requires two states, one of each. Importantly, BELIEFS by themselves are motivationally inert. If you BELIEVE a mouse is in front of you, this will have no bearing on your actions unless you have a relevant DESIRE. The latter category is broad: it includes a fear of mice, dislike of them, wishing that they be killed, the intention to squish them, and so on.

⁴ The normative theory of mental states would be more worrying to some error theorists. They have had to grapple with the argument that the error theory implies there is no reason to believe it (Streumer (2017: ch.9) even argues he *cannot* believe his error theory!), but the conjunction of the normative theory of mental states plus the error theory entails there are no BELIEFS at all, undermining a core component of its own thesis (Streumer 2013).

Anti-Humeans (Nagel 1970, McDowell 1979, Scanlon 1998, Shafer-Landau 2003) believe motivation can result from *single* simple mental states. The most common example is a moral belief. It is plausible that Yusuf's belief that he ought to help Sandy across the road might motivate him directly to do so. According to these philosophers, at least some moral beliefs are correctly described as both BELIEFS and DESIRES as categorised above, and are called 'besires'. The moral belief that abortions are bad, for instance, could be simultaneously a BELIEF that abortions are bad and a DESIRE that, say, fewer abortions occur. (Note that the content of the DESIRE is different from the content of the BELIEF, despite them both being descriptions of the same mental state.)

With the advent of hybrid theories in metaethics a clarification must be made. According to hybrid theories, each moral judgement involves *two* states, a BELIEF and a DESIRE.⁵ David Copp (2014: 51) argues the moral judgement that torture is wrong is the "belief that torture is wrong as well as ... commitment to a policy of opposing and avoiding wrongful acts." Hybrid theorists need not believe in besires despite holding moral judgements to have both directions of fit because a besire is meant to be *one* simple state. As Smith (1994: 119) writes, "[b]esires are not supposed to be gerrymandered mereological sums of two states – a belief plus a desire – each of which could occur without the other. Besires are conceived of as unitary states that have both the representational characteristics of beliefs and the motivational characteristics of desires." The difference between Copp and the anti-Humeans, then, is that Copp thinks moral judgements are two states that come apart: someone could BELIEVE torture is wrong without having any DESIRE and thus, despite BELIEVING torture is wrong, does not make a fully-fledged moral judgement (which requires the commitment he describes).⁶ Contrastingly, anti-Humeans think a moral judgement *can* be a single state irreducible to BELIEFS that aren't DESIRES or DESIRES that aren't BELIEFS. Given that many anti-Humeans maintain that moral judgements aren't *always* motivational – that is, they are sometimes merely BELIEFS and not besires – it's an incredibly fine distinction! All it seems to boil down to is, in the instances where a person

⁵ They either *are* the two states taken together (Boisvert 2008), or they are single higher-order states composed out of the two of them (Toppinen 2013), or they are relations between the two of them (Schroeder 2013).

⁶ Copp (2018: 1343-1347) thinks there is a *basic* way of thinking about moral properties that does not involve the motivational aspect and an *internal* way of thinking, which does. To use moral language felicitously, one must be using the internal way of thinking (*ibid*: 1354).

both BELIEVES a moral proposition and has the appropriate DESIRE, whether or not these are two separate states or simply two accurate descriptions of the same single indivisible state; and in the instances where a person does *not* have the relevant DESIRE, whether or not to call the mere BELIEF in a moral proposition a fully-fledged moral judgement.

The Humean Theory of Motivation is a common assumption to make in many metaethical debates. Connie Rosati (2016: §3.1) writes it is “fair to say that Humeanism continues to be the dominant view.” I believe it to be universally accepted among expressivists. Expressivists explain the reliable connection of moral judgements with motivation via the thesis that moral judgements are DESIRES. However, as we shall see in §4.4 of this chapter, pure expressivists such as Blackburn should *reject* Humeanism and accept that moral judgements are *besires*: simple states with both directions of fit. This is controversial. Blackburn and others (e.g. Sinclair 2006) would not classify moral judgements as BELIEFS and Blackburn (1998: 101-4) argues that *besires* do not exist. Nevertheless I hope to make a good case by appealing to the responsiveness of moral judgements to the moral facts. Now, with the mental divisions in hand, we are ready to evaluate early characterisations of the expressivist view.

1.2 The Good Old Days

The early expressivists were called *emotivists* (Ogden and Richards 1923, Ayer 1936, Stevenson 1937, 1944) and *prescriptivists* (Carnap 1935, Hare 1952). Dreier (2004: 23) beautifully, almost regretfully, describes this as a time when it “used to be easy to tell a moral realist from a moral irrealist.” To find out, you’d simply ask whether there were moral facts. A realist would say ‘yes’ whereas an early expressivist would say ‘no’, and that would be that. Of course there were other views that denied the existence of moral facts too. Subjectivists and relativists might allow one sense of moral fact (in the minimal sense to be discussed in §2), but they deny there are any *objective* moral facts. Error theorists agree with the early expressivists that a plain ‘no’ is the correct answer. What was the difference, then, between error theorists and expressivists? According to error theories, moral discourse is assertoric – we assert moral propositions and attempt to describe a moral reality that isn’t there. ‘Stealing is wrong’ is *apt* for truth but isn’t true, and the same goes for any substantive moral claim. Moral *language* is the same on the error theory as it is according to realistic theories.

Their quarrel lies in the metaphysics – unlike realists, error theorists think moral properties either do not exist or cannot be instantiated.

By contrast, the early expressivists denied that moral language is used to make claims at all but is used to express DESIRES instead. The declarative grammatical form, on their view, is deceptive. Prescriptivists thought the underlying meaning of sentences like ‘stealing is wrong’ was to be cashed out in terms of imperatives such as ‘do not steal.’⁷ Thus moral sentences, like any command, do not have truth conditions but satisfaction conditions. Alternatively, emotivists thought that moral sentences express emotional attitudes. For instance, an utterance of ‘stealing is wrong’ might express a negative emotional state of disapproval towards the act of stealing, akin to saying ‘stealing? Boo, hiss!’⁸ One attraction of these views is that, like the error theory, they do not commit us to the existence of supposedly “queer” objective morality, but unlike the error theory they do not take ordinary moral practice to be infected with a huge mistake. In other words, people aren’t making untrue moral assertions all the time because moral utterances aren’t assertions of anything. It’s not as though the person saying ‘stealing is wrong’ is missing the mark, they’re simply using the sentence to express an attitude or to tell people not to steal. Furthermore, these theories fit well with what seems to be the sociological function of morality: to regulate each other’s behaviour.

The Good Old Days are gone. While the early expressivists denied the existence of moral assertions, moral propositions and so on, modern expressivists do not. The main reason modern expressivists accept moral language to be assertoric seems to be to do with *creeping minimalism*, to be discussed later. However there are independent reasons to believe moral language is assertoric. First of all, it simply fits better with common sense intuition that we do indeed make genuine moral claims. When I declare ‘eating meat is wrong,’ it *feels* like I have asserted something which may or may not be true. It seems as though what I have said is answerable to the world in a way that commands (‘do not eat meat!’) and ejaculations (‘boo to

⁷ “But actually a value statement is nothing else than a command in a misleading grammatical form. It may have effects upon the actions of men, and these effects may either be in accordance with our wishes or not; but it is neither true nor false. It does not assert anything and can neither be proved nor disproved.” (Carnap 1935: 24.)

⁸ “But in every case in which one would commonly be said to be making an ethical judgement, the function of the relevant ethical word is purely ‘emotive’. It is used to express feeling about certain objects, but not to make any assertion about them.” (Ayer 1936: 107.)

eating meat!') are not. A second reason involves thinking about what moral claims are and are not coherent with. Suppose a drug baron is offering a bribe to a new police officer.⁹ He says to her "You really ought to refuse this bribe, but take it and be on your way." This seems coherent. Conversely, 'refuse this bribe, but take it and be on your way' does not. A third reason involves thinking about the function of the speech acts and when they do or don't succeed. When a sincere demand is made, the purpose is to get someone to do something.¹⁰ When we make assertions in conversation, the point is normally to transfer information and get our audience to accept something. Ejaculations of 'boo!' or 'ouch!' seem to be made merely to evince some feeling – football supporters booing or cheering succeed in doing what the language is there for, even if they haven't "transferred" their feelings to anyone else. In the bribery case I described earlier, the normal function of what the drug baron said will have been achieved if the police officer takes the bribe while believing she ought to refuse, and will have failed were it the other way around. There is good reason, then, to think standard utterances of declarative moral sentences are assertions.¹¹

Nevertheless, the legacy of the early expressivists that modern expressivists subscribe to is the idea that moral sentences express a DESIRE. This is the core of the expressivist picture. (Since other theories hold this to be the case it cannot be the whole story. If it were, this chapter would not need to be written.) It was introduced as a way of sidestepping tricky metaphysical questions: rather than jumping straight into examining the nature of slippery moral properties, look instead at the way moral language is used. When Wade said abortion was wrong he was expressing his disapproval and that's all there is to it, so there's no need to worry about what moral reality is like – or so it has been contended. (I argue against this in chapter 3.) Yet expressivists have not always been clear about what they mean by the term 'express' in sentences such as 'moral sentences express

⁹ Thanks to Jimmy Lenman for this illustration of the point.

¹⁰ I am not conflating pragmatics and semantics. I do not argue that the purpose, or the intended use, constitutes or otherwise determines meaning, or anything like that. I am simply using our intuitions about whether the normal linguistic function is satisfied as evidence for the kind of utterance it is.

¹¹ Exactly what assertions are is controversial, but I contend that expressivists can help themselves to any theory of assertion on the table in chapter 4. Since there are normative theories of assertion we might worry the issue of circularity rears its head again as it did with normative theories of mental states, but if it does I would believe it to be similarly benign. Circularity arises again in the next subsection.

DESIREs.’ What exactly is this relation between sentences and mental states?

1.3 Expression for Expressivists

Expressivism has been described by proponents and adversaries alike as a theory of normative language, a core component of which holds that we ought to “explain the meaning of a term” by “what states of mind the term can be used to express”¹² and the states of mind expressed by normative terms are DESIREs. This already looks like a strong commitment to make. Is this how meanings of terms are normally explained? And if not, is it how they ought to be? These kinds of questions will be answered in §3. For now, I simply note that *expression* forms what is often thought to be integral to expressivist theory. I will also use the terms ‘approval’ and ‘disapproval’ as placeholders for whatever DESIRE expressivists think is expressed by sentences like ‘veganism is right’ and ‘tax evasion is wrong.’

There are many ways in which I can express something. I could frown and express displeasure, or I could dance and express joy. I could repeatedly tap ‘down’ on my arcade stick to make my Super Street Fighter II Turbo character rapidly duck and stand¹³ in order to express disrespect for my opponent. I could write a musical piece to express many complex things at once. More philosophically, a sentence might express a proposition. These are not what modern expressivists have in mind. The *expression relation* is a relation between sentences and mental states, and whatever it turns out to be, it must be the case that metaethical realism implies that moral sentences express BELIEFS and expressivism implies that moral sentences express states like approval and disapproval.

Schroeder believes expressivism has unpalatable implications for what the expression relation must be. In chapter 2 of *Being For* (2008) and in chapter 1 of *Expressing Our Attitudes* (2015), Schroeder quickly dispatches three expression relations that expressivists cannot use and presents this as a disadvantage, and then puts forward an account on our behalf which he

¹² Gibbard (2003: 7).

¹³ A motion known as ‘teabagging’ in the fighting game community.

deems problematic.¹⁴ This subsection will be split into two parts. In the first part I argue it is no great loss to be unable to use the first three potential expression relations, and in the second I discuss the merits of Schroeder's own *assertability* relation and argue, against Schroeder himself, that it is not in any way a cost to expressivism.

1.3.1 Prior Expression Relations

Schroeder begins with the *same-content* theory of the expression relation. On this account, a sentence expresses a mental state if and only if they share the same content. The sentence 'grass is green' expresses the belief that grass is green because they are about the same thing – they share propositional content with each other. However, since expressivists begin with moral thought and do not help themselves to moral propositions (at least at the beginning of their story, but not at the end as we'll see in §3.3), this option is unavailable. Furthermore, if 'murder is wrong' expresses, for instance, a desire that nobody murders, then as highlighted by Frank Jackson and Philip Pettit (1998: 245) the sentence appears to have a very different propositional content from the mental state it expresses: it is true that murder is wrong, but is not true that nobody murders.

At this point, Schroeder (2008: 25) begins to worry. "This is bad! Expressivism started with the idea that it could take advantage of whatever everyone else had to say about the [expression relation]. But it turns out that they cannot!" Throughout the chapter, Schroeder describes the same-content account as "the most natural account" (*ibid*: 24) and "the standard conception of semantics" (*ibid*: 32), perhaps in an attempt to stress that expressivism has bolder commitments in the philosophy of language than its proponents had previously realised. In later work (2015: 41) he continues to think this is "a considerable blow" to one of expressivism's core ideas. I believe his worries are unwarranted. Schroeder is overstating the power of the same-content account. It may very well be a "natural" account in the sense that it is the first that comes to mind, but trouble is found when extending it beyond ordinary descriptive sentences.

¹⁴ Schroeder includes a small subsection in his (2015: 43-44) book addressing expression as *implicature*. I do not discuss it here because only hybrid cognitivists like Barker (2000), Copp (2001, 2009) and Finlay (2005) take the expression of DESIRES to be merely implicated.

If I say ‘shut the door,’ the sentence does not express the belief that the door is shut. If any belief is expressed at all, it will be a belief the door *isn't* shut. In uttering it, it seems to me that I primarily express my desire for the door to be shut (or for my audience to shut it), perhaps alongside a belief that the door is currently open and perhaps also that it is possible for my audience to shut the door. The content of my desire – that *the door is shut* – has the same content as the sentence ‘the door is shut,’ but imperatives and declaratives are distinct sentences with different meanings. Indeed, for any proposition *p*, we have the declarative ‘*p*,’ the imperative ‘make it the case that *p*,’ and the interrogative ‘is it the case that *p*?’ The propositional content of all three sentences is *p*, but only the declarative has equivalent truth conditions (or any truth conditions at all), and only the declarative expresses the belief that *p*. While the propositional content of each sentence is the same, the propositional attitude expressed is not. It may be perfectly fine to say that ordinary descriptive sentences share their propositional content with beliefs, but not every sentence expresses a belief. The same-content account absurdly implies that ‘the door is shut’ expresses the desire that the door is shut, and that ‘shut the door’ expresses the belief that the door is shut, since they share their content. We should not mourn the loss of the same-content account.

Schroeder then discusses and dismisses the *causal* account, according to which a sentence *S* expresses a given belief because it “is typically *caused*, in normal, sincere, cases, by the belief” (2008: 25). Ayer, for instance, thought of moral utterances as ejaculations or exclamatives – emotional responses to things that morally excite us similar to the way wincing or saying ‘ouch!’ can be caused by a sudden pain or the way Frodo (my cat, not the hobbit) might purr as a result of feeling mollified. Expressivists have come a long way from this account, but we do learn a lesson from seeing what is wrong with it. If I lie to you by telling you that I like your new haircut when really I don’t, my utterance of ‘I like your new haircut’ is clearly not caused by the relevant belief, because I don’t *have* that belief. Since a liar does not change the meaning of the sentences they utter, we reach our lesson: it is possible to express a mental state one does not have.

Finally, Schroeder discusses Allan Gibbard’s (1990: 85) view of the expression relation. On this view, clearly influenced by Paul Grice, one expresses a mental state by *intending* to convey to the audience that one is in it. This evades the problems with the same-content and causal accounts. When asking a question, the speaker might intend to convey their desire for

their audience to tell them something, and liars intend to convey the false information that they are in a particular mental state. Gibbard (2003: 78-9) abandoned the account in light of Mark van Roojen's (1996: 325) objection. Essentially the *intentional* account does not distinguish between different sentences if they are used to convey the same information. To use Gibbard's example, Holmes may intend to convey to his audience that he has the belief that Moriarty has arrived by uttering 'Moriarty has arrived.' However, he may also intend to convey that information by saying 'I believe that Moriarty has arrived,' or 'I am hereby intending to convey to you that I have the belief that Moriarty has arrived,' both of which are sentences with different truth conditions and different meanings.

Gibbard correctly acknowledges that this is a problem with the *intentional* account and not expressivism. He then draws the conclusion that cognitivists will be in the same boat, but Schroeder disagrees. "It is true that everyone owes an account of expression that does not result in this conflation. But the same-content account is sufficient to avoid that conflation. The same-content account is all the cognitivists need."¹⁵ However, as I have said earlier, Schroeder does not consider how problematic the same-content account is when applied to sentences that aren't descriptive. This is not to blame Schroeder. It is easy to overlook imperatives and interrogatives when the literature does so as well. I hope I have shown that, while none of the three expression relations discussed are adequate for expressivists, this is simply because they are inadequate. It is not a special problem for expressivism.

1.3.2 The Assertability Expression Relation

Schroeder's own suggestion is that sentences express what they do in virtue of their assertability conditions. I think this is the most promising account, and it can even be extended in certain ways if we so desire. Assertability conditions of sentences are the mental states one must have in order to be linguistically permitted to assert those sentences. For instance, one must (at least) believe grass is green in order to assert 'grass is green,' therefore the sentence 'grass is green' expresses the belief that grass is green. If somebody does *not* believe that grass is green and asserts 'grass is green' they are making a linguistic error in the sense that they are breaking the rules of the

¹⁵ Schroeder (2008: 28).

language, either intentionally (e.g. when the speaker is lying) or accidentally (e.g. when the speaker is learning a language and doesn't yet grasp it, or when the speaker is mistaken about their own mental states).¹⁶ The precise "norm of assertion" differs from theory to theory. Some think that merely believing p is sufficient for linguistic permission to assert it, but others think that more is required – perhaps justified belief, or a particular level of credence, or even knowledge (Williamson 2000). For our purposes, it is enough to know that there *are* linguistic rules governing when it is or isn't linguistically permitted to use a given sentence, and those rules relate to the mental state of the speaker.

Importantly, assertability conditions are not truth conditions. Suppose Honest Jane and Liar Larry both think less than half of the world's CO₂ is stored in its oceans and are asked to consider the sentence 'about 93% of CO₂ is stored in the oceans.' Honest Jane is speaking linguistically correctly but falsely when she denies the sentence and Liar Larry, who intends but fails to live up to his name, is speaking linguistically incorrectly but truly when he asserts it. Applying this to moral language on the expressivist view, we must admit that Wade, who disapproves of abortion, speaks linguistically correctly when he asserts 'abortion is wrong,' but we need not think his disapproval of abortion has anything to do with the truth conditions of that sentence. Since I believe abortion is permissible I do think Wade has made a mistake, but it is a *moral* mistake and not a linguistic one.

The *assertability* expression relation is this: a sentence S in language L expresses mental state M if and only if the rules governing language L permit one to assert S only if one is in M. Do we need an account of where these rules come from? I do not think so, so long as it is plausible that there are such rules. The rules may even arise differently for different languages, especially when we compare natural languages to stipulated or formal languages. But whether they have come about as a sort of tacit convention in our community or whether by expert decree, we clearly have a sense of whether a speaker has or has not used language correctly.

I mentioned earlier that the account could be extended. As it stands the account only applies to assertions, implying only declarative sentences in

¹⁶ Relatedly, Ridge (2006b: 487) argues that "a speech-act is sincere just in case the speaker *believes* that she has the state of mind *she believes* it expresses" which is essentially the point that sincerity is a matter of *trying* to follow the rules of the language, whether or not you manage it.

natural languages express mental states at all. Now, we could leave it there and stipulate that sentences only express mental states when they are assertable. I see no theoretical problem with this and so I shall not develop an extension in detail. However I will mention that, if we find it natural to say that other kinds of sentences express mental states, we could say something similar about them too. Perhaps interrogatives and imperatives come with similar rules: ask Q only if you are in M*, or command C only if you are in M**. The details would take more spelling out than they're worth, but I hope this gesture will satisfy those who think 'shut the door' expresses the desire for the door to be shut, or that anyone who makes this demand while secretly wanting it to remain open is in some sense misusing the language.

Schroeder (2015: 53-54) is not totally pleased with his proposed expression relation and raises some worries for it. I think he worries too much. First off, he notes that to say a moral sentence expresses a DESIRE is "to say that it is *semantically correct* to assert that sentence only if you are in that state of mind. So this raises the obvious question of whether *semantic correctness* is a normative concept."¹⁷ (Schroeder here is assuming expressivism is a thesis about all normativity.) If it is, though, all this means is that, given the truth of expressivism, claims about which mental states are expressed by sentences are themselves expressions of DESIRES. For example, the sentence 'the sentence "stealing is wrong" expresses disapproval of stealing' may express linguistic disapproval (however this is characterised) of *asserting 'stealing is wrong' without morally disapproving of stealing*. Therefore, if expressivism is true about morality *and* the expression relation, coming to accept the truth of expressivism is to have a certain DESIRE. But without an argument that DESIRES are problematic this is simply a consequence, not an objection.

In a comparable fashion to appealing to a normative theory of mental states in a theory of normative concepts, there is a certain circularity in this that might give some people pause, perhaps for a similar reason a cousin of mine yelled "Stop!" as I moved to point a video camera at the screen it was live feeding to, as if it would have split the universe in two. Once again though, realism about all these concepts would contain the same circularity (realists think they are expressing BELIEFS when they say 'moral sentences

¹⁷ Schroeder (2015: 53). Schroeder uses the term '*semantic correctness*' but, so as not to conflate it with views in semantics we'll discuss in §3, I have opted for the term 'linguistic correctness' throughout this subsection.

express BELIEFS') and there is no clear reason to worry any more about circularity in either case. What might be worrying (but isn't) is if our metaethical theories had strong and diverging commitments for how we ought to carry out our normative inquiry. Suppose, for instance, that normative inquiry as directed by expressivism led to the normative conclusion that moral sentences express DESIRES, whereas realistic normative inquiry led to the denial of this conclusion.¹⁸ Then we might be in trouble, as the theories would be self-contained and their proponents could not hope to persuade one other by rational means. Yet this is no reason to shun *expressivism* in particular. Besides, most metaethical theories do not differ much in their recommended methods of normative inquiry anyway.

Schroeder's first worry about his assertability expression relation was based on the hypothesis that linguistic rules are genuinely normative, but this is not necessarily the case; the fact an assertion would be linguistically prohibited need not be a reason against doing so. Situations seem possible where you *only* have reasons to lie, which involves intentionally breaking one of these linguistic rules, such as telling a Nazi officer 'these children are not Jewish' when you believe they in fact are. (Conversely, the fact that you believe Gary has an awfully large nose makes it linguistically permissible to assert it, but morally it may well be impermissible.) This is similar to the rules of chess. Moving the rook diagonally is prohibited, but there may be no reason to follow the rules and all the reason to break them. Yet *if* you have reason to play chess correctly, you have reason to avoid moving the rook diagonally. Likewise, if you have reason to use language correctly as we usually do, you have reason to obey its rules. This seems to me more plausible than the rules of a language giving us categorical reasons to obey them, but as I have discussed, I see no sufficient reason for expressivists *in particular* to be worried about normativity creeping all the way down.

The other worry Schroeder (2015: 53-54) ends with is a fear that expressivism would collapse into a form of relativism, or at least have some relativistic sounding consequences.¹⁹ His thought is that the assertability expression relation implies "you speak truly when you say, 'if I didn't

¹⁸ Or worse, what if every metaethical theory ultimately refutes itself? Would that mean no metaethical theory is correct?

¹⁹ Arguments of this sort have been surprisingly persistent (Jackson and Pettit 1998, Peacocke 2004, Suikkanen 2009). I do not discuss them in this thesis because I believe Schroeder's (2014) general objection to them is decisive.

disapprove of murder, then it would be semantically incorrect for me to say that murder is wrong'."²⁰ He thinks this is bad because he has the intuition that it is *correct* to say that murder is wrong (given that murder is indeed wrong), and he goes on to air a suggestion that we build *knowledge* into the assertability conditions of sentences. "The idea would be that you would only have to admit that your counterfactual self spoke in a way that was semantically correct, if you were willing to ascribe *truth* to what your counterfactual self said."²¹ I think this is a mistake. What was especially handy about the assertability expression relation was its divorcing of assertability conditions from truth conditions. This allowed us to make sense of Honest Jane sticking to the rules and Liar Larry flouting them. In short, Schroeder's intuition of "correctness" is tracking the truth of the statement instead of whether the rules of language use were correctly followed. We should not pay heed to this intuition.

I hope I have shown that expressivists are not losing out on anything when it comes to the expression relation that has so often been taken for granted. The accounts Schroeder argues do not work for the expressivist turn out not to work for anyone, and the assertability expression relation Schroeder posits on behalf of the expressivist does not have the downsides he fears.

This brings me to the end of §1. We have looked at the division of mental states into BELIEFS and DESIRES, we have seen a first pass at defining expressivism as the view that moral sentences express DESIRES instead of being truth-apt assertions that express BELIEFS, and we have examined what this *expression* claim might amount to. In the next section we will see why this first pass is incorrect.

²⁰ Schroeder (2015: 53).

²¹ *Ibid.*

2. Quasi-Realism

In this section I discuss why the lines between realism and expressivism are far blurrier than they used to be. The culprit is the popularity among expressivists of deflationary theories of truth, or “truth minimalism,” which allows them to accommodate much of what seems attractive about realist theories. The trouble is that minimalism *creeps* – if truth can be deflated, why not all the claims that seem to distinguish realism from expressivism? I begin by outlining these core realist claims before showing why, on the assumption that truth and related concepts can be deflated, they do not help us characterise the expressivist position.

2.1 Accommodating Realist Intuitions

‘Realism’ is a term of art and has been used as lightly as to mean any theory according to which moral judgements are beliefs that are sometimes true, and as heavily as to refer to only robust forms of realism according to which moral facts exist as non-natural, irreducible entities. I shall land somewhere in between. I define realism (for now) as any metaethical theory subscribing to the following:

- (1) There are moral facts.
- (2) Moral facts are mind-independent and non-relative.
- (3) Moral judgements are beliefs about these things.
- (4) Declarative moral sentences express these beliefs.

Realists can differ among themselves in many ways but particularly when it comes to the nature of moral facts and how we know about them. Are they natural facts or non-natural? Are moral truths analytic or synthetic? We will look at these kinds of issues in chapter 3, but the point here is that realists, in my sense, agree with all four of the above claims. Error theorists deny (1) and (2), although they typically agree moral facts *would* be mind-independent and non-relative if they existed, and subjectivists, relativists and constructivists deny (2).

The realist claims appear attractive to many. It is strongly intuitive that, say, it is a fact that torturing for fun is wrong, and that this fact is the case regardless of what anyone thinks or how they feel about it, and that it isn’t merely true relative to some framework or context but that it is true objectively. However, many have felt that realism cannot be backed up

metaethically due to various worries about persistent moral disagreement, the difficulties of reducing moral properties to natural ones, the queerness of irreducibly moral properties and how they could supervene on natural ones, the connection to motivation of and authority over our actions, and other strong objections.

The earliest expressivists were shamelessly anti-realist and would have denied all four of the realist claims. They were happy to accept, for instance, that a moral sentence such as ‘stealing is wrong’ “expresses no proposition which can be either true or false.”¹ This clearly contradicts folk practice – ordinary speakers will attribute truth and falsehood to moral claims. While this is no decisive objection by itself since the folk may simply be wrong, the problem goes deeper. Moral thought and discourse sometimes involve hypothesising the truth of moral propositions, disjoining moral sentences with prosaically descriptive ones, attributing moral knowledge to others, making valid arguments with moral premises and conclusions, and so on. If we throw all of this out there won’t be much of ethics left to work with.

Contemporary expressivists tend to be uncomfortable about attributing widespread error to common sense practice. Indeed, one of expressivism’s supposed advantages over error theory was that it doesn’t postulate a huge error infecting morality, but if everyone who says ‘that is true’ or ‘that is false’ of moral sentences is making a mistake, we have once again attributed widespread and systematic error to the folk. Blackburn (1984) began *the quasi-realist project* with the Wittgensteinian conviction that ordinary language is, by and large, just fine as it is. It *works*. We should aim to keep moral thought and discourse intact, and if the philosopher cannot make sense of it, it is the philosopher who is more likely at fault. Here is Blackburn:

But perhaps there is no mistake [permeating ordinary moral thinking]. I call the enterprise of showing that there is none – that even on antirealist grounds there is nothing improper, nothing “diseased” in projected predicates – the enterprise of *quasi-realism*. The point is that it tries to earn, on the slender basis, the features of moral language (or of other commitments to which a projective theory might apply) which tempt people to realism. (Blackburn 1984: 171.)

¹ Ayer (1936: 107).

One may worry we give the folk too much credit. Couldn't ordinary speakers be mistaken? Scientists are not deterred by the prospect of counterintuitive discoveries, so perhaps philosophers should not be either. After all, aren't philosophers meant to be questioning the things we all take for granted? Well yes, but consider the analogy with science more carefully. It was a great thing when we discovered we are mostly empty space, or that simultaneity is relative to a frame of reference, but scientific discoveries leave everyday practice intact. We still talk about tables and chairs, we just know more about their underlying structures than we once did. When Granny says 'the Sun moved behind the tree,' the heliocentric model of the universe does not refute her statement. Granny makes no such astronomical commitments. Instead, science can give us a more precise specification of the truth conditions for what she said – perhaps that the Earth has spun on its axis such that the tree now stands between Granny and the Sun.²

Discoveries can change how we conceive of ourselves and our place in the universe. They won't stop us going to the shop to buy food, but they may tell us interesting things about what we're up to when we do. Similarly, nothing in the philosophy of mathematics is going to stop us counting our money in the usual ways. The philosopher might tell us interesting things about what numbers are and what we're up to when we count, but if they tell us two plus two isn't really four – or worse, that it's actually five – we'll be sceptical.³ Likewise for ethics. If the expressivist has to admit through gritted teeth that it isn't *really* true that stabbing kittens for fun is wrong, we'll be sceptical. The worry the quasi-realist wants to discharge is that "once we see the disappearance of some favoured conception of objectivity and truth, or once we see what these amount to, we can no longer exercise judgement as before."⁴ The goal, then, is to develop a form of expressivism that allows us to keep intact our pre-theoretical intuitions about torture being objectively, mind-independently wrong, and the project begins with minimalism about truth.

² Echoes of Davidson (1973) and his principle of charity abound.

³ A healthy scepticism should not rule these theories out of court. There may be strong reasons to deny the existence of numbers (Field 1980) that outweigh the implausibility of mathematics being false. But they had better be *very* strong reasons. Lewis (1991: 58-9) perhaps put it best: "Mathematics is an established, ongoing concern. Philosophy is as shaky as can be. To reject mathematics for philosophical reasons would be absurd ... That's not an argument, I know. Rather, I'm moved to laughter at the thought of how presumptuous it would be to reject mathematics for philosophical reasons."

⁴ Blackburn (1993: 3).

2.2 Minimalism

Moral practice raises troubling metaphysical and epistemological worries. It looks as though we ascribe moral properties to things, so in metaethics we wonder whether there are such properties and how we can know about them. The same can be said of our talk about truth. People say of some sentences or propositions that they are true, and of others that they are false. In doing so, are we ascribing some property or are we doing something else? If truth is a property, what kind of property is it and how do we know about it? The theories of truth most popular among contemporary expressivists are deflationary. I am lumping them all together, and I use ‘truth minimalism’ and ‘deflationary theories of truth’ interchangeably. Defenders of these theories make the same sidestep expressivists do: instead of jumping right in and looking at the truth property itself, look at the act of ascribing it and leave the metaphysics out of the story. Part of the story is that to ascribe truth to a sentence or proposition is to do nothing over and above asserting that very same sentence or proposition. (The other part explains the concept’s utility for generalisation.) Here, rather ironically given his qualms about ascribing truth to moral sentences, is Ayer:

If I say that it is true that Shakespeare wrote *Hamlet*, or that the proposition “Shakespeare wrote *Hamlet*” is true, I am saying no more than that Shakespeare wrote *Hamlet*. Similarly, if I say that it is false that Shakespeare wrote the *Iliad*, I am saying no more than that Shakespeare did not write the *Iliad*. And this shows that the words ‘true’ and ‘false’ are not used to stand for anything, but function in the sentence merely as assertion and negation signs. (Ayer 1935: 28.)

There are different ways to fully spell this idea out but the basic scheme is the same – we explain our willingness to apply the truth predicate to a sentence in terms of our willingness to assert the sentence itself.⁵ If we apply this to moral sentences we get an idea of how the quasi-realist can vindicate moral truth. If I say that it is true that stealing is wrong, or that the

⁵ See Field (1994) and Horwich (1998). There are complications when it comes to ascribing truth to sentences and propositions we do not know, since I might proclaim everything the oracle told you is true without knowing exactly what the oracle told you. This might be understood as a commitment to affirm any proposition I discover the oracle said or else to withdraw my assent to ‘everything the oracle told you is true.’ There is also a distinction between minimal understandings of truth-aptness of sentences and minimal conceptions of *truth*, which I shall ignore as they tend to travel in tandem.

proposition ‘stealing is wrong’ is true, I am saying no more than that stealing is wrong. In conjunction with the expressivist understanding of moral assertions, this equivalence implies that ‘it is true that stealing is wrong’ expresses a DESIRE, which we are calling disapproval of stealing. Similarly, if I say that it is false that premarital sex is wrong, I am saying no more than that premarital sex is not wrong. Once again, due to the equivalence, the sentence ‘it is false that premarital sex is wrong’ also expresses a DESIRE. It is unclear exactly what kind of DESIRE it would be. In chapter 2 I will argue that it is yet to be satisfactorily cashed out by pure expressivists and that hybrid expressivists can provide a good account of this DESIRE. Still, *if* the expressivist can present us with an adequate account of the state of mind expressed by negated sentences, then we will have an account of the state of mind expressed by attributions of falsehood. In short, expressivists need to solve the negation problem (and the Frege-Geach problem more generally) before being able to speak of moral truth and falsity. We will proceed as if they have successfully done so.

To summarise thus far, expressivists have a story of what is going on when someone asserts (and, we are supposing, when someone negates) a moral sentence. Deflationary theories of truth don’t need much more, and they certainly don’t demand metaphysical explanations. Other features often desired are *syntax* (“the syntactic surface features of a paradigmatic truth-apt” discourse) and *discipline* (“a sufficient degree of normative discipline establishing clear standards of appropriate and inappropriate use”).⁶ These features of moral discourse are covered in the Frege-Geach problem to be discussed in chapter 2 and in our previous discussion of the expression relation in §1.3 of this chapter. If all goes well, *and* if a deflationary theory of truth is correct, then the quasi-realist will have vindicated the use of the truth predicate and will be able to continue attributing truth and falsehood to moral sentences in good faith. The project of quasi-realism would be up and running.

2.2.1 Is Minimalism Actually True?

I am attracted to deflationary theories of truth but I do not have space to defend them here. I can only reiterate what others have said in the past: “deflationary theories of truth are in the same spirit as expressivist theories

⁶ Both references to Lenman (2003: 34).

of normative predicates,"⁷ "the Minimalist conception of truth and the expressivist conception of ethics fit together so well"⁸ and the line of argument for deflationary theories of truth "is *precisely* the pattern of argument used to motivate expressivism."⁹ Since the arguments in favour of minimalism are similar to those in favour of expressivism it is not surprising to see the same people defending both of these theories together. As for myself, insofar as I am an expressivist I am a minimalist about truth. In a way this is backed up by a principle of charity about folk discourse – if expressivism is the correct metaethical theory, the most charitable way to interpret the folk when they ascribe truth to moral sentences is along deflationary lines.

All that being said, I acknowledge the conceptual space available for the truth of expressivism and the falsity of minimalism. What then? There are two options. First of all, Ridge (2014: ch.7) argues that his version of expressivism is compatible with more robust theories of truth, such as the correspondence theory. If Ridge is right and there are other theories of truth available to the expressivist, the project of quasi-realism can go ahead. If it cannot, then we have an easy answer to the title question of this chapter: expressivism is the view that moral sentences express DESIRES and are not truth-apt. There is another option too. It might be the case that a plurality of theories about truth are correct, perhaps in different domains, and that some of them make sense on a realist picture of morality but not an expressivist one. If that were shown to be the case then perhaps we could distinguish realism from expressivism via something like the following idea: expressivism allows for minimal truth of moral sentences but not for any stronger truth, whereas realism allows for moral sentences to have more or all of the different truth properties.

The easy answers to the title question would be nice in a way but it seems odd that differing theories of moral *thought* are best distinguished by theories of *truth*, and furthermore, the easy answers wouldn't make for an interesting opening chapter. What's difficult is characterising expressivism as a distinct metaethical theory on the hypothesis that minimalism is the only correct theory of truth. Since it has seemed to many that there *is* a difference between expressivism and realism *regardless* of what theory of

⁷ Dreier (2015a: 451).

⁸ Dreier (2004: 24).

⁹ Ridge (2014: 201).

truth turns out to be correct, it is on this hypothesis that Dreier raises the problem of creeping minimalism.

2.3 Minimalism Creeps

Given that moral sentences can be true, there is a cheap way to vindicate the thought that there are moral facts. Take any moral sentence you think is true, such as ‘torturing innocent people merely to pass the time is wrong.’ Since this sentence is true, it is also a *fact* that torturing innocent people merely to pass the time is wrong. Saying something is true and saying something is a fact amounts to roughly the same thing. Some people may feel a little uneasy with this idea. Fair enough, you might think, there is a sense of ‘fact’ according to which any truth is also a fact. It’s a fact that grass is green and it’s also a fact that I am in Sheffield. But there is a different, more metaphysically robust sense of fact whereby facts are what *make* truths true, and they are not always best described “disquotationally” in the way we might say ‘grass is green’ is made true by the fact that grass is green – the sentence ‘I am in Sheffield’ when uttered by me is made true by the fact that *Graham Bex-Priestley* is in Sheffield, and plausibly there are no ‘I-mentioning’ facts to make such sentences true. I am sympathetic, although not wedded, to this line of thought and will discuss facts as truthmakers in §4.3. For now, it is enough to note that there *is* a sense in which it is a fact that I am in Sheffield, and we quasi-realists can help ourselves to that sense in order to vindicate talk of moral facts: someone asserting it is a fact that torture is wrong is expressing disapproval of torture.¹⁰ This is a moral claim, and if you share this negative attitude towards torture then *ipso facto* you think it is a fact that torture is wrong. Quasi-realists, then, can continue asserting there are moral facts without incoherence.

Just as we “deflated” talk of facts to truth and talk of truth to the sentences themselves, we can also deflate properties. Saying, for instance, that torture has the property of wrongness is just a long-winded way of saying torture is wrong. In this way, the quasi-realist can make sense of an utterance of ‘torture has the property of wrongness’ or ‘the property of

¹⁰ Perhaps ‘it is a fact that’ adds more force or indicates more certainty. If this is more than mere pragmatic implicature, then perhaps we can amend the mental state: ‘X is wrong’ expresses disapproval, and ‘it is a fact that X is wrong’ expresses disapproval with some (presumably vague) minimum level of certainty.

wrongness is instantiated by torture' as an expression of disapproval.¹¹ If you disapprove of torture, *this just is* the ascription of the property of wrongness to torture, and you are linguistically permitted to assert that torture has a moral property – wrongness.

We have seen how a quasi-realist can speak of moral facts and moral properties. What about the intuition that they are mind-independent? Can the quasi-realist assert this in good faith, if they are so inclined? Yes they can. Just as the assent to moral facts and properties are thought of as moral claims, mind-independence can also be interpreted morally. Essentially it is a question of *what matters*. To figure out whether to accept mind-independence, we consider whether the moral facts would change if we felt differently. Since I believe that torturing innocent people to pass the time would be wrong even in worlds where everyone thought it was okay, or where everyone approved of it, or where any other relevant mental state was changed, then I believe that this moral claim is true mind-independently. The debate gets complicated when we consider people like Christine Korsgaard (1996) who would agree that there are moral claims that remain true across all possible worlds, but who think they do *because* of some universal mind-dependent fact (about, say, rational agency). The relevant issue is about what we feel *determines* the rightness or wrongness of actions. I save deeper discussion of these metaphysical issues to chapter 3. The point for now is that the issue of what rightness and wrongness depend on – whether it depends on what we think or plan to do, or what we *would* think or plan to do after some process of reflection, or instead on what would maximise pleasure or some other fact that doesn't depend on our thinking – is understood by the quasi-realist as a moral issue. Talk of mind-independence can thus be vindicated as expression of, say, disapproval of torture *whether or not* it is accepted by the community. Here is Blackburn:

To say that an ethical view is true is just to reaffirm it, and so it is if we add the weighty words 'really', 'true', 'fact', and so on. To say that it is objectively true is to affirm that its truth does not vary with what we happen to think about it, and once more this is an internal, first-order ethical position. (Blackburn 1998: 296.)

¹¹ I think the story is a little more complicated than this and will argue the pure expressivist has trouble talking *about* moral properties beyond ascribing them. Chapter 3 will discuss in more detail how to make sense of the language and metaphysics of moral properties.

The quasi-realist can now interpret talk of mind-independent moral facts and properties as legitimate expressions of approval and disapproval. Minimalism “sucks the substance out of heavy-duty metaphysical concepts.”¹² An expressivist, by engaging in the project of quasi-realism, captures the first and second of my listed characteristics of realism.

What about numbers (3) and (4)? One may think expressivists would be giving up on their view if they admitted moral judgements were beliefs. But minimalism doesn't stop creeping. When someone says ‘Jenny believes torture is wrong’ the most charitable assignment of truth conditions to their claim (on the assumption expressivism is true) is that Jenny disapproves of torture. To believe torture is wrong *just is* to disapprove of torture, and to disapprove of torture is to believe torture is wrong. Quasi-realists, then, think that moral judgements *are* beliefs after all.¹³ Perhaps, though, these are merely beliefs and not BELIEFS, and what distinguishes expressivism is the denial that moral beliefs are BELIEFS properly understood. This is the line Sinclair (2006) takes, and I will argue against it in §4.4. It will turn out that the quasi-realist must think of moral beliefs as BELIEFS as well. But let us continue creeping for the last stretch. If moral sentences express beliefs, then we have good reason to think we *assert* them, because “assertion has the following analytical tie to belief: if someone makes an assertion ... she has a belief whose content can be captured by means of the sentence used.”¹⁴ Likewise, we can deflate description and representation: Jenny represents torture as being wrong, which just means she disapproves of it, and she describes it as being wrong with the sentence ‘torture is wrong,’ which just means she asserts it. In short, it looks like the quasi-realist agrees (or *can* agree, if she is so inclined) with the four listed characteristics of moral realism.

Minimalism has crept. Quasi-realists can seemingly agree with everything realists say, and I count myself among them. There are moral facts, moral truths are mind-independent, moral judgements are beliefs, and moral sentences express these beliefs. But this leaves us with a problem. I want to *distance* myself from realists, and realists want to distance

¹² Dreier (2004: 26).

¹³ “I genuinely believe that pain is bad, and my expressivistic theory, filled out, explains what this believing consists in.” (Gibbard 2003: 183.)

¹⁴ Wright (1992: 14). While the implication in the quoted sentence goes only one way, we have reason to treat it as a biconditional, since it seems equally analytically true that sentences primarily expressing beliefs are assertions. The nature of assertions is discussed in chapter 4.

themselves from me. And it is not simply for my own comfort that I don't want to be lumped with the realists – it is because I think realists face rather damning objections and I don't want to have to answer them. This is a lesson Sharon Street brings to the fore in her (2011) paper entitled 'Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have it Both Ways.' If we quasi-realists admit to the existence of mind-independent moral facts, we're going to have to explain what they are like, how we know about them, why there is such persistent moral disagreement, and so on. The rest of this chapter is about finding a solution to the problem of creeping minimalism. Before we proceed, though, I wish to comment upon how minimalism about truth creeps back onto itself.

2.3.1 Quasi-Realism about Truth

If truth minimalism is correct, it blurs the boundaries between metaethical realism and expressivism because it interprets the core realist claims as first-order moral claims to be argued about internally. But what has not been as much discussed is that minimalism equally applies to theories of truth themselves. Theories thought to be in competition with deflationary theories of truth include correspondence theories, where the truth of a sentence is a matter of it standing in some relation to reality; coherence theories, which deem a belief true if it is coherent with, say, a set of fully coherent beliefs; and pragmatist theories of truth, which define true propositions as, for instance, those we would believe at the end of ideal inquiry. However, if we take minimalism as our starting point, what is to stop it creeping into these very debates? One might interpret 'what Haifa said corresponds to reality' as another way of saying 'what Haifa said is true,' and so truth minimalists may agree with a core claim of correspondence theories. Sometimes people define deflationary theories as the view that truth is not a genuine property, but it's hard to see what stops us from thinking Haifa's statements have the property of truth if we indeed think they are true. And if there are truth properties, we can argue about their nature and what they depend upon, in the same way expressivists can morally argue about whether right and wrong depend on features of our minds. This point is not entirely original. Here is what Ayer wrote after revealing his deflationary view:

Must we then regard all 'theories of truth' as attempts to answer a fictitious question, and so as nonsense? No, we may regard them as

attempts to answer the genuine empirical question How are propositions actually validated? So that when Professor Schlick says that the truth of a synthetic proposition consists in its agreement with reality, and not merely in its logical compatibility with other propositions, we may assume that he is attempting to show what is the criterion that people actually apply to their beliefs. In this view, what he is saying is that, as a matter of fact, we are not prepared to accept a system of synthetic propositions as valid merely on the ground of its logical self consistency, but require also that it should agree with reality. Can we accept this as a correct account of our procedure? I think we can, if we give a suitable interpretation to the objectionably vague expression 'agreement with reality'. (Ayer 1935: 29.)¹⁵

It seems to me Ayer is saying that even if truth is deflated, the usual truth debates – whether truth is a matter of internal coherence, or correspondence with reality, or reaching stable belief, or something else – can continue as *first-order* inquiry. (Given Ayer's logical positivism he construed this as empirical enquiry about how we actually proceed, but we needn't do so.) The parallels between expressivism and deflationary theories of truth continue to be striking. Expressivist quasi-realism was intended to leave first-order ethics intact, such that ethicists can go on arguing about what matters in much the same ways as before, and many of the old metaethical debates can be construed at the first-order level of ethical inquiry. Similarly, deflating truth still leaves room for argument about when the truth predicate is to be applied, and above we see Ayer putting forward the view that these are procedural, epistemological arguments about when to accept (or in his terms 'validate') a proposition. In other words, the argument between defenders of coherence and correspondence theories is better understood not as about the metaphysical nature of truth but as about the

¹⁵ James (1907: 141) had a similar idea sprouting from a correspondence, not minimalist, starting point: "Truth, as any dictionary will tell you, is a property of certain of our ideas. It means their 'agreement,' as falsity means their disagreement, with 'reality.' Pragmatists and intellectualists both accept this definition as a matter of course. They begin to quarrel only after the question is raised as to what may precisely be meant by the term 'agreement,' and what by the term 'reality,' when reality is taken as something for our ideas to agree with." Thus, despite the correspondence-sounding nature of the claim that truth is agreement with reality, James thinks the debate can continue as one of explicating what agreement with reality amounts to. He of course argues for a pragmatist conception: "*True ideas are those that we can assimilate, validate, corroborate and verify. False ideas are those that we can not.*" (*Ibid*: 142.) "The 'absolutely' true, meaning what no further experience will ever alter, is that ideal vanishing-point towards which we imagine that all our temporary truths will some day converge." (*Ibid*: 150.)

correct way to form beliefs: is coherence enough, or are we to use some other procedure? My interpretation of Ayer in this passage may be incorrect, but the point still stands and I think it is an interesting one. The fact that truth minimalism creeps back onto itself to open up the possibility of quasi-realism about truth has generally gone unremarked.

In this section we have examined the project of quasi-realism and seen how it threatens to blur the distinctions between realism and expressivism, making it tricky to answer the title question of this chapter. In the next section we will look at characterisations of expressivism as a distinctive semantic thesis and as a *metasemantic* thesis, and I will argue expressivism is neither.

3. Semantics and Metasemantics

Expressivism has most often been characterised as a semantic theory about what moral words and sentences mean. Charles Stevenson's (1937) paper is entitled 'The Emotive Meaning of Ethical Terms,' expressivist theory is said by both Gideon Rosen (1998: 387) and Wedgwood (2007: 35) to be in the business of providing a "*psychologistic semantics*" and Sinclair (2011: 386) thinks that the "attitude expressed, and the co-operative action-guiding nature of this expression, gives the meaning of the sentence." The following passage from Schroeder illuminates how he understands expressivism:

I think we should understand expressivism as committed to an underlying semantic program ... (1) that the role of semantics is to assign an assertability condition to each sentence of the language, understood as the condition under which it is semantically permissible for a speaker to assert it. (2) These assertability conditions typically or always say that the speaker needs to be in a certain state. (3) Descriptive sentences inherit their propositional contents (truth-conditions) from the belief that it is their assertability condition to be in. And (4) the assertability condition of complex sentences are a function of the assertability conditions of their parts, where this function is given by the meaning of the sentential connectives that are used to form the complex sentence. (Schroeder 2008: 31-32.)

This looks like an incredible commitment for expressivists to make in the philosophy of language. While I am happy to adopt Schroeder's account of the expression relation, I resist the move made without argument from the viability of assertability *expression* to assertability *semantics*. I agree partly with (1) and (2) that a core component of expressivist *theory* is to ensure a DESIRE is assigned to each sentence of moral language, such that to accept a moral sentence is to have that DESIRE, and with (4) that expressivists must show that these DESIRES are related in the right way when it comes to complex sentences. This is the difficult Frege-Geach problem to be discussed in chapter 2. However, nothing in Schroeder's expression relation necessitates point (3). The fact that I must believe grass is green in order to linguistically correctly assert 'grass is green' needn't ground the truth conditions of 'grass is green.'

Recently, several philosophers (Silk 2013, Pérez Carballo 2014, Ridge 2014, Chrisman 2015, Köhler 2017) have gone against the semantic grain

and argued that expressivism is best understood as a *metasemantic* thesis. That is, expressivism is not engaged in the project of giving *meanings* to ethical terms but instead is a theory of *how* terms get their meanings. Let us try to get a clearer grip on the distinction. To create two non-overlapping fields which are naturally called ‘semantics’ and ‘metasemantics’ is a difficult task. Like with ethics and metaethics – an analogy Ridge (2014: 8) draws on – the division will be unclear at times and sometimes theories in one field will have implications in the other. I will have to make a few stipulations in the following paragraphs, and in doing so I will mostly be following suggestions found in Alexis Burgess and Brett Sherman (2014).

Semantics is the business of assigning meanings, or “semantic contents,” to sentences or subsentential expressions. It is a semantic fact (if it is a fact at all) that the French sentence ‘la neige est blanche’ means snow is white, and – somewhat controversially since Quine – that ‘bachelor’ means unmarried man. Pairing linguistic expressions with meanings is a semantic project; one that, incidentally, used to be thought of as a philosophical, introspective project but is now seen as a much more empirical one. The two languages in the pairing are the “object language” and the “metalanguage,” the former being the one we are trying to understand and the latter being one in which competence is assumed. Sometimes sentences have propositions as their meaning – for instance, the sentence ‘grass is green’ means (the proposition) that grass is green, but the sentence ‘I am in Sheffield’ perhaps has a *character* (Kaplan 1979) as its meaning, in this case a function from contexts to propositions. In the context of being uttered by me, it would output the proposition that Graham Bex-Priestley is in Sheffield, and as uttered by Liv Tyler it outputs the proposition that Liv Tyler is in Sheffield.

What Burgess and Sherman (2014: 3) call “basic metasemantics” is the project of discovering the non-semantic facts (if there are any) *in virtue of which* the semantic facts hold. Michael Johnson and Jennifer Nado (2017: 717) define a metasemantic theory to be “a theory of what makes it the case that an expression means what it does, rather [than] something else, or nothing at all.” Why is it true that ‘la neige est blanche’ means snow is white but ‘sfhd dgshgu’ means nothing? Did the community of language users form a tacit convention where we paired words with ideas (Davis 2003)? Do sentences mean what they do partly in virtue of being the objectively most natural assignments (Lewis 1983)? Is it to do with actual or counterfactual

dispositions of competent speakers (Johnson and Nado 2014)? Or perhaps there are simply no semantic facts to ground (Kripke 1982).

We might also contrast *theories of meaning* from basic metasemantics. A theory of meaning is an account of the meaning relation itself. Basic metasemantics might be thought of as looking for its *grounding*, whereas theories of meaning are interested in its nature or essence. For instance, perhaps for S to mean M, M has to be the truth condition(s) of S, or the representational content of S (clearly this is only plausible for declarative sentences), or maybe M has to satisfy the role S plays in our cognitive economy. Finally, we can consider the metaphysics of semantic values. What are meanings in the first place? Are they truth conditions, conceptual roles, mental states, propositions? If they are propositions, what are propositions: thoughts, sets of possible worlds, or something else?

The boundaries of all four of these areas are difficult to draw, and many answers to some of these questions will affect the answers to others. It is my contention that expressivism is compatible with at least the most popular answers to these questions, and hence does not make as large a commitment in the philosophy of language as many have supposed. I cannot show that expressivism is compatible with every possible answer. Instead I shall select a small smorgasbord of each and show how an expressivist could adopt them. This clearly won't be an exhaustive list but I hope to demonstrate that the burden of proof should be shifted to our opponents to argue why certain semantic or metasemantic positions exclude metaethical expressivism. I begin with semantics.

3.1 Semantics

Natural languages are untidy things. For those of us interested in meaning, formal semantics can be a helpful tool to clean up messy languages and model them. The models help us grasp various aspects of natural languages, such as context-sensitivity and compositionality, and how we as finite beings come to learn them in a matter of years. In more concrete instances, semantics can simply help us learn the meanings of the words in a given language.

It is fair to say that expressivists have historically been assumed to be undertaking a semantic project to uncover the meanings of moral terms. A semantic theory of moral terms might, for example, pair them with natural,

descriptive meanings: in the spirit of Jeremy Bentham, perhaps ‘good’ means pleasurable. Stevenson (1944: 22) and R. M. Hare (1952: 118) both believed that moral sentences *do* have such natural descriptive meanings, but these meanings are “secondary” and not necessarily relevant to the audience’s agreement. On their view the primary meanings of moral terms are emotive or prescriptive. James Lenman (2003: 35) writes that one way to understand quasi-realist expressivism is the project of finding “*patently* not truth-apt” sentences such as “Hurrah for our not killing innocent people!” that have equivalent *meanings* to seemingly truth-apt moral sentences like ‘it is wrong to kill innocent people.’ But when it comes to the meaning of moral terms it seems clear there are many answers an expressivist could give, and in fact there are many different answers expressivists *have* given.

Let us begin with the word ‘ought’. Gibbard (2003: 185) takes the Moorean stance that the word is not reducible to other terms. If you ask Gibbard what ‘ought’ means he’ll say he cannot tell you directly. He can tell you an explanatory story about the use of the word via the planning states he thinks it is used to express, but the meaning itself cannot be reduced. We might instead have a *reasons-first* view (Parfit 2011, Scanlon 2014) where, for instance, ‘you ought to dance’ means you have most reason to dance. But it doesn’t seem to me that an expressivist is committed to rejecting reasons-first views in virtue of their expressivist theory. Instead, they could tell their story about our use of ‘reasons’ language in virtue of some desire-like states of mind we express with the language, and the word ‘ought’ would be explained in terms of when we think we have most reason to do something (cf. Gibbard 2003: 191). Ridge, another expressivist, has a *standards* based semantics of terms like ‘good’ and ‘ought’, which he compellingly argues are context sensitive due to their wide varieties of use over both normative and non-normative matters. In the normative uses of ‘ought’ we’re interested in, ‘you ought to give to charity’ means “[a]ny acceptable standard of practical reasoning would, given contextually specified circumstances, recommend that you give to charity.”¹ (Note that ‘acceptable’ is the fundamental notion here, and so we are still left with a normative concept at bottom, which Ridge goes on to explain in terms of non-cognitive “normative perspectives.”)

The context sensitivity of normative terms is also discussed in great detail by the cognitivist Stephen Finlay (2014), who argues ‘you ought to

¹ Ridge (2014: 40).

dance' means, roughly, that dancing is the relevant option most likely to lead to a contextually specified end.² Alice giving some advice to her friend to help cheer her up might be a context where the goal is to make her friend happy, and her utterance 'you ought to dance' would be true if dancing was indeed the best (probabilistically speaking) way to achieve that goal and false otherwise. Finlay thinks all normative terms can be reduced to non-normative ones, but there is a way for expressivists (and realists) to adopt the semantics for 'ought' and the other end-relational terms he develops: allow, *contra* Finlay, that some contexts specify distinctively *moral* ends. 'You ought to turn yourself in' could then be interpreted in some contexts as making a moral recommendation and might mean that, say, turning yourself in would be the most likely way to obey the correct moral principles. In this way an expressivist could steal the impressive end-relational semantics Finlay develops while holding onto the idea that there are irreducibly moral concepts. The judgement about which moral principles are the correct ones could be explained in terms of desire-like attitudes such as Ridge's "normative perspectives."

When philosophers in metaethics write about the meanings of moral terms what we generally find is that some term is fundamental: 'good', 'ought', 'reason', 'acceptable', etc. There doesn't seem to be any reason why any of these options is off the table for an expressivist. As long as the judgement itself is construed as desire-like, expressivists can agree with certain realists (Boyd 1988, Enoch 2011, Parfit 2011) that the *meaning* of that term is not reducible to other terms. The only meanings of fundamental moral terms that initially look to be off the table are ones that reduce them to non-moral terms in the way that analytic naturalists do (Lewis 1989, Jackson 1998). But perhaps there is a way to accommodate even this.

3.1.1 Analytic Reduction for Expressivists

Non-natural moral realists believe there are moral properties which are not reducible to anything else. Some *natural* realists, often called synthetic reductionists, think moral properties are reducible to natural properties but moral concepts are not reducible to natural ones. Just like 'water' and 'H₂O' have different meanings but the same referent (an identity to be discovered

² See Finlay (2014: ch.3) for discussion of the meaning of 'ought' which is set out in far more technical and mathematical terms.

a posteriori), the words ‘good’ and ‘pleasurable’ might have different meanings yet refer to the same property, the property of being pleasurable. Analytic reductionists, on the other hand, believe moral terms have the same meanings as some other, ostensibly non-moral, terms. ‘Good’ might *mean* pleasurable, in the way that ‘bachelor’ means unmarried man and ‘circle’ means a two-dimensional shape with a boundary consisting of all the points equidistant from a given point. Is expressivism incompatible with the semantic claims of analytic reductionists?

Some theories of meaning are *normative* theories of meaning (Kripke 1982, Brandom 1994, Gibbard 2012). According to these theories, claims about what terms mean are not mere descriptions (of, say, actual usage in a community) but “normative claims seeking to legislate similarity of usage between terms.”³ They are best understood not as ‘is’ claims but as ‘ought’ claims: as claims about how words *ought* to be used, about what inferences are licensed from sentences using them, and so on. If these theories are correct, it is open to the expressivist to claim that ‘good’ means pleasurable. All they must do to convince us is give us good reasons to believe we ought to use ‘good’ and ‘pleasurable’ in similar ways, that we ought to infer one from the other, and so on. This does not commit them to thinking the mental state of judging something to be good is the exact same mental state as believing something to be pleasurable. It is simply the commitment to it being the case that we ought to use the terms in similar circumstances: if you’re willing to apply the term ‘good’ to the ice cream, the charitable donation and the sex last night then you ought to be willing to apply the term ‘pleasurable’ too, and vice versa. While I do not personally believe this to be the case, it is a matter for first-order normative inquiry and thus I do not rule it out *qua* expressivist. If meaning is normative, expressivists *can*, yet from my normative perspective shouldn’t, be analytic reductionists.

3.1.2 Semantic Programmes

We’ve looked at semantic claims that are fairly specific, such as hypotheses about what the word ‘ought’ means. More broadly speaking there are a variety of semantic *programmes* which look at systematic ways to model language. Examples include truth-conditional semantics, possible worlds semantics, conceptual role semantics, and the psychological semantics we

³ Lance and O’Leary-Hawthorne (1997: 63).

saw Rosen and Wedgwood attribute to expressivists. They are all in the business of assigning meanings to terms, but they differ in what they attribute as meanings more generally: when looking for the meaning of a term, truth-conditional semantic programmes will make reference to its potential contribution to a declarative sentence's truth-conditions, whereas psychologicistic semantics (also known as *mentalism*) will assign a particular mental state, such as a thought. Semantics itself can be carried out using any of these frameworks.

The most popular, “almost orthodox, way to pursue semantics”⁴ is truth-conditional. It would be a shame if expressivists were committed to psychologicistic semantics, thereby closing the door to important work in truth-conditional semantics we might like to use on, say, the meanings of indexicals if we favour truth-conditional approaches there. In general, truth-conditional semantics “is a powerful, highly fruitful, and widely accepted research program in philosophy *and* linguistics”⁵ and it would be odd if it were to fail because of a theory in metaethics.

In light of this, people like Ridge and Köhler aim to show that expressivism is compatible with truth-conditional semantics. The trouble is that truth-conditional semantics is carried out in terms of realistic sounding terms such as reference, the extension of a property, propositional content, and of course truth. However, those engaging in truth-conditional semantics are not doing metaethics, or even metasemantics – they are modelling meanings. They use these realistic sounding terms without further analysis. It is an open question whether these terms are to be interpreted in a robustly realistic sense or whether they can be interpreted quasi-realistically. As Pérez Carballo writes, “semanticists are not in the business of pronouncing on metaphysical issues. It would be incredible if the viability of current semantic theory as we know it depended on the outcome of a controversial metaphysical dispute.”⁶ Contemporary expressivists after all think that there are moral properties and truths. Whether expressivists can capture the notion of a moral *proposition* will be discussed in §3.3, but if we can, then on our working hypothesis that truth

⁴ Köhler (2017: 3).

⁵ Köhler (2017: 4). See also Schroeder (2010: 210): “the departure from truth-conditional semantics also means a departure from all the great successes of truth-conditional semantics at accommodating the compositional constraint and at explaining the semantic properties of complex sentences.”

⁶ Pérez Carballo (2014: 136).

and its correlates can be deflated we can use all the terminology without giving up our position. Once again the burden of proof is on our opponents to show why we cannot make use of the work done in truth-conditional semantics.

3.2 Metasemantics

In the previous subsection we saw that expressivism is compatible, not in competition, with a number of different semantic claims about the meanings of moral terms. In light of this, philosophers have recently begun characterising expressivism as a *metasemantic* theory instead. I am unconvinced. Recall that metasemantic theories are theories of *how* terms get their meaning or *why* they mean what they do, rather than theories of *what* terms mean. While I agree that expressivism is compatible with different semantic programmes and various assigned meanings to moral terms, this doesn't automatically mean expressivism is a metasemantic theory. Indeed it seems compatible, not in competition, with most basic metasemantic theories too.

Let us begin with ideationalist metasemantics, traceable to ideas found in John Locke's work. These theories all begin with thought and explain the assignment of linguistic meanings in terms of the mental states of the language users. Grice's (1989) influential theory is an example of this, and we came across it earlier in §1.3.1 in our discussion of Gibbard's Grice-inspired expression relation. Roughly, according to Grice, speakers may use a declarative sentence *S* to mean *p* if and only if the speaker intends their audience to believe *p* (or to believe that the speaker believes *p*) and for them to recognise this intention and form their beliefs accordingly. This is speaker-meaning, and what a speaker means with a sentence can vary drastically to what the *sentence* actually means. Grice, again roughly, thought sentence-meaning was determined by a group of language users having standard "procedures" to generally use sentence *S* to speaker-mean *p*. While I may use the sentence 'I have a scarf' in response to 'are you sure you want to go outside?' to speaker-mean that I want to go outside and I am prepared for the weather, that is a very specific and circumstantial use. The sentence 'I have a scarf' is only associated with the procedure of speaker-meaning that the speaker has a scarf, and so that is what determines its meaning.

Expressivism is compatible with this basic metasemantic theory. The meaning of ‘torture is wrong’ could be determined by being associated with the procedure to speaker-mean that torture is wrong, whereby speakers intend their audience to believe that torture is wrong (or to believe that the speaker believes that torture is wrong) and so on. Expressivists will add that the belief that torture is wrong is adequately characterised as disapproval of torture, but no extra trouble is caused by the paraphrase. Intending my audience to recognise my disapproval and to join me in disapproving of it makes sense and is intuitive. I am reminded of Stevenson’s analysis of ‘this is wrong’ as “I disapprove of this; do so as well.”⁷

The ease of expressivism slotting into this mentalist theory in metasemantics is general. Since mentalist theories begin with thought, all the expressivist needs to do is characterise moral thought in a way that distinguishes it from the way cognitivists characterise moral thought. (This task is harder than it seems, as we shall see when we take it up in §4.4.) Regardless of whether something Gricean is correct or whether the relation between mental states and meaning is more to do with conventional pairings of sentences and beliefs (Lewis 1969) or words and thoughts (Davis 2003), the expressivist can agree with the basic metasemantic story and add a chapter about the relevant mental states associated with moral terms. This is what Ridge does. He prefers the ideationalist theory of Scott Soames (2010) who has an understanding of propositions as types of mental event. We will discuss it in §3.3.

What about basic metasemantic theories that are not mentalist in nature? One set of examples includes what we might call ‘interpretational’ theories. Donald Davidson (1973) proposed that we assign meanings to sentences and beliefs simultaneously rather than deriving one from the other, and that the correct assignment was the one that maximised the truth and rationality of what people were saying and believing. This “principle of charity” to assign meaning works along the following lines. Since we English speakers tend to say ‘it is raining’ only when it is in fact raining, it is more charitable to assign the meaning that it is raining to that sentence rather than, say, that a tiger approaches. The principle, then, is that sentences mean what they do in virtue of that being the most charitable interpretation of what people say. To solve problems about indeterminacy

⁷ Stevenson (1944: 267).

of reference, David Lewis (1984: 227) added the dictum that the referents of terms should “respect the objective joints in nature.”

Expressivists can in principle adopt these metasemantic theories too. Just as the realist might say that the most charitable interpretation of ‘what you did is wrong’ is to assign the meaning that what the audience did was wrong, since the sentence is typically uttered only in cases when the audience did in fact do something wrong, the quasi-realist can agree. The principle invoking the notion of truth-maximisation is open to minimalist conceptions of truth. One noticeable aspect is that the assignment I just mentioned involves a moral judgement that, in general, people say the sentence only when their audience *in fact* has done something wrong.⁸ As such the expressivist will construe someone assigning this meaning as being in a state of disapproval. But this is perfectly consistent. A huge problem with these theories is that they are *too* charitable. Intuitively it seems possible for a community to get things wrong more often than not. However this is a specific problem with the metasemantic theory and not with its combination with expressivism – the same problem would arise were metaethical realism true.

Another kind of metasemantic theory makes use of the dispositions of speakers to apply terms. Here is Johnson and Nado (2014: 81): “A linguistic expression E means some object, property, kind, relation etc., X, in the mouth of speaker S, in virtue of the fact that S would be disposed to apply E to X if S had all the relevant information.” They later clarify (2017: 720) that having all the relevant information includes having all the true beliefs that would influence the speaker’s disposition to apply the term. The fact that Cliff is disposed – if asked and if willing to sincerely respond – to apply the term ‘cat’ to the fox outside doesn’t determine the meaning of ‘cat’ and this is because he *falsely* believes the fox is a cat. Once his mistaken belief is updated, his dispositions will change accordingly. These kinds of theories should once again be fairly simple to combine with expressivism. Ben T Ham is currently disposed to apply ‘good’ to all and only pleasurable

⁸ The supervenience of moral properties on natural ones may give rise to a concern. Suppose rightness supervenes on maximising pleasure. People’s attributions of rightness would therefore be true in equally as many cases if ‘rightness’ were assigned the property of maximising pleasure as it would if it were assigned the property of rightness. Therefore it is either indeterminate which of the two assignments is correct, or they are in fact identical and reductive naturalism is correct. One option would be to embrace naturalism as I do in chapter 3. The other would be to use whatever method non-natural moral realists use to get out of this problem.

actions, but that is due to his false moral belief that ‘good’ means pleasurable, and once we update this mental state of his, his dispositions will determine the meaning of the word.⁹

One family of basic metasemantic theories I will not try to square with expressivism are causal theories of meaning and reference. The reason is that they “outsource” morality to what seem like morally irrelevant causal factors in our language use, something I discuss in more detail in §5 of chapter 3. There are two main ways these theories work. On the way most directly inspired by Saul Kripke (1980), a term is introduced by the first person to use it, who “baptises” the referent of the term in an ostensive way, and then the term gets communicated to other people. The true referent of the word ‘water’ is determined by tracing its causal history back in time to whatever the first person to use the word was metaphorically “pointing at” at the time. The other kind of causal theory holds that a term’s meaning is determined by what features of the world causally regulate our actual usage, and is associated with Cornell realism (Sturgeon 1985, Boyd 1988). Yet causal theories of reference are unpopular for moral terms, regardless of the truth of expressivism. For the Kripke-inspired versions, who wants the correct moral theory to be beholden to the first person to use moral terms? Given the age of moral language the first people to use the words may well have had outdated and potentially abhorrent moral frameworks. Secondly, “causal regulation of use” theories face well known objections that I will not rehearse here, such as Terence Horgan and Mark Timmons’ (1991, 1992a, 1992b, 1996, 2015a) *Moral Twin Earth* objection. This has led to David Brink (2001), one of the Cornell realists, dropping this metasemantic theory and defending normative theories of meaning instead, which we have already discussed in §3.1.1 as compatible with expressivism.

My strategy in this subsection has been to show that expressivism is compatible, not in competition, with basic metasemantic theories, and as such is *not* a metasemantic theory. However, Ridge (2014: 110) *agrees* with my first point: “Although ideationalism is the most natural home for Ecumenical Expressivism, there are ways of transposing the main ideas of Ecumenical Expressivism into an impressively wide range of other metasemantic frameworks.” So why does Ridge continue to hold expressivism to be a metasemantic theory? In the sentence following the one I just quoted

⁹ The issue of supervenience seemingly implying reduction rears its head once more. I echo what I said in the previous footnote.

he claims the “key commitment is that the theory in *some* sense explains the meanings of sentences in terms of a state of mind with which it is systematically associated.” I am not sure why this is the case. Of course if ideationalism is true and meanings are assigned in virtue of relevant mental states, then this won’t change after combining it with expressivism. But that is purely due to ideationalism, not expressivism. On the other hand, if Ridge and I are correct that expressivism is compatible with metasemantic theories that *deny* meaning is to be explained in terms of mental states, then Ridge’s “key commitment” is false, whether or not expressivism is true.¹⁰ I thus conclude that expressivism is not a metasemantic theory.

3.3 Propositions

Here I address the issue of moral propositions. If expressivism can indeed make use of different theories in semantics – especially truth-conditional programmes – then we had better be able to make sense of their talk of propositions. After all, the most common meanings put forward for declarative sentences are propositions (or, for context-dependent sentences, functions from circumstances to propositions): ‘snow is white’ and ‘la neige est blanche’ have the same meaning because they both mean the same proposition, the proposition that snow is white. But what *are* propositions, and are they expressivist-friendly?

Recently, Schroeder (2015: ch.3) and Köhler (2017) have respectively constructed and deflated propositions in order for expressivists to use them as semantic contents. I have no objections to either of these projects. But, as may have become apparent by this point, I like to hedge my bets. If it turns out that a particular metaphysics of propositions is correct, it would be both a shame and quite odd for this to refute a theory of what we’re up to when we do ethics. Thankfully, I do not think any metaphysics of propositions proves to be problematic. Once again my strategy will be to discuss some of the main theories and show what an expressivist can say if they turn out to be true, and once again this cannot be an exhaustive list. I merely hope to shift the burden of proof to those who would excise us from standard conceptions of propositions.

¹⁰ Chrisman (2010: 199) makes a similar point: “We can recognize that certain sorts of ethical claims conventionally express desires, intentions, and/or plans without holding that they mean what they do in virtue of expressing those things.”

Propositions might be sets of possible worlds (Lewis 1986a), namely the set of worlds in which they are true. On these theories, 'snow is white' and 'la neige est blanche' both mean the set of possible worlds containing all and only the worlds where snow is white, and they are true if and only if the actual world is a member of that set. What stops an expressivist from using the same theory? 'Torture is wrong' means the set of possible worlds where torture is in fact wrong, and it is true if the actual world is a member. I can imagine a possible world where torture generally has moral benefits that outweigh the moral costs, and that world would not be included in the set. Figuring out which set of possible worlds should be assigned to a moral sentence will be a matter for moral theorising, but this is not a problem for expressivism any more than it is for realism. For instance, if we decide that Bentham's utilitarianism is correct, then the proposition we will assign to 'one ought to maximise pleasure' is the set of all possible worlds, and its negation will be the empty set. Expressivists can adopt this theory by maintaining the *assignment* of propositions is desire-like in nature: to assign the set of all possible worlds to 'one ought to maximise pleasure' is to endorse hedonistic utilitarianism.¹¹

One unpalatable consequence of possible worlds theories is how they interpret all sentences that are necessarily true as meaning the same thing. The sentences 'all dogs are mammals' and 'all cats are mammals' are true in all possible worlds (let us assume cats and dogs are essentially mammals), but presumably they mean different propositions. To combat this, with inspiration from Gottlob Frege and Bertrand Russell, propositions might be abstract structured entities (Cresswell 1985, Salmon 1989, Menzel 1993): the former sentence is assigned a proposition composed out of constituents which include dogs, the latter is not. Consider the proposition that grass is green. In the literature, this structure of this proposition is usually written a little like this: $\langle \langle o \rangle, P \rangle$ where, in our case, o is the object grass and P is the property of being green. The proposition is true if and only if o has the property P . I see no obstacle to expressivists using these kinds of accounts of propositions. When o is torture and when P is the property of being wrong we have a moral proposition, and I believe it to be true. Expressivists can happily characterise my belief in this proposition as disapproval of torture. The proposition is a robustly real, abstract entity; believing the proposition to be true is a matter of moral commitment. What

¹¹ We can view Gibbard's (1990, 2003) solutions to the Frege-Geach problem as taking a similar, albeit extended, route.

is wrong with this line of thought? Here is why Ridge rejects these kinds of propositions:

One answer *not* available at this stage would be to begin with propositions understood roughly as Russell or Frege understood them. In spite of their differences, Russell and Frege both understood propositions as abstract entities which intrinsically represent the world as being a certain way. Once we understand propositions in this way, there is no reason a normative proposition could not figure as the content of a robustly representational belief. Clearly, such a belief would be a normative belief. In short, such a conception of the normative proposition would lead very quickly to a representationalist order of explanation of the semantic contents of normative claims, and hence to a form of cognitivism. (Ridge 2014: 125.)

Ironically, a similar objection is levelled against Ridge's use of Soames' propositions by Andrew Alwood (2016), which we shall discuss in the next paragraph but two. The same objection is also levelled by Matthew McGrath and Frank Devin (2018: §10.1) against deflating propositions for expressivist use: "moral propositions exist, and so why can't they be believed independently of having any intrinsically motivating states?" The answer, first, is that moral propositions *do* represent the world as being a certain way. That was the lesson of creeping minimalism. If there is a way to distinguish "robust" and "minimalist" senses of representation, nothing stops expressivists from interpreting these theories as saying propositions are structured and *minimally* representational. (To reiterate, propositions themselves are perfectly real in a robust sense. To believe there are such things as moral propositions is to have a robustly representational belief. It is when we characterise someone thinking a moral proposition is *true* that the expressivist theory comes in. My belief that *the proposition that torture is wrong exists* can be viewed realistically while my belief that *torture is wrong* is desire-like.) Secondly, the contention that moral propositions cannot be believed independently of having intrinsically motivating states *just is* a core thesis of expressivist theory. To believe a moral proposition *is* to take up a practical stance. Anyone who says propositions must be possible to believe absent any motivational state is begging the question.

Finally, we come to theories of propositions as types of mental event. The most fully worked out account is found in the work of Scott Soames (2010, 2014), and Ridge (2014: 127-31) adopts it so expressivists can make

use of propositions in their theories. Ridge (2014: 127) writes that the “proposition that *p* should be understood as the cognitive event type of entertaining that *p*.” But what is ‘*p*’? A proposition? That won’t work. We’re trying to understand what propositions are. The best way to understand the theory is to begin with judgements. I judge that a particular object is red, and this is a type of mental event – in Soames’ terms, it is a mental act of predication – which, we assume, can be cashed out in (psychological, neurological, physical?) terms that make no reference to propositions. Judging an object to be red is one way of entertaining that the object is red, but we might entertain it without believing it. We might imagine that it is red, or we might suppose that it is red in order to reason on that hypothesis. Ridge construes mere entertainments as *simulations* of judgements, which seems plausible. The idea, then, is that the proposition the object is red *just is* the type of mental event described in these entertainments.

Alwood has two complaints about Ridge’s use of Soames’ theory of propositions. First of all, he claims Ridge is using the theory in order to avoid mentalist semantics, according to which meanings are mental states – perhaps ‘grass is green’ means the belief that grass is green and ‘I ought to bathe’ means the plan to bathe. But, “if he both says that normative propositions are the meanings of normative claims and then reveals normative propositions to be attitudes (or events in which an agent bears these attitudes), he has committed himself to the mentalist semantics that he was trying to avoid.”¹² Alwood is right that this is a route to mentalism, but I’m not sure why he thinks Ridge wants to avoid it. All Ridge wants to do is vindicate normative propositions. Indeed, the implied mentalism is something Ridge (2014: 131) explicitly endorses as it closes the apparent gap between expressing a state of mind and a proposition, since to “express a proposition just *is* to express a mental *type* of a certain kind.” Absent any deep and substantive argument that propositions cannot be combined with mentalism, Ridge should not be worried.

Alwood’s second complaint is, as promised, analogous to Ridge’s complaint about an expressivist adopting Russell and Frege’s conception of propositions. Soames’ theory is couched in terms of mental *representation*. Not just any type of mental event will do: “to entertain the proposition that *o* is red is to *predicate* redness of *o*, and thereby to *represent* it as red.”¹³

¹² Alwood (2016: 17-18).

¹³ Soames (2014: 95). Even to think it is not true that the object is red involves representing that object as being red (*ibid*: 97): “To entertain the proposition *that it is not true that o is red*

Alwood thinks Ridge is therefore not permitted to make use of Soames' account because it "excludes the possibility of nonrepresentational propositions."¹⁴ My reply, as you might expect, is equally analogous: moral propositions *are* representational. They represent the world of moral facts as being a certain way. Entertaining the proposition that torture is wrong is to *predicate* wrongness of torture, and thereby to *represent* it as wrong. If expressivism is true then to predicate wrongness of something is to disapprove (or to simulate disapproval) of it. Perhaps Alwood would insist that once we have figured out how to distinguish realism from expressivism, Soames' theory is making use of representation in a robust sense which is available to realists and unavailable to expressivists. I see nothing in Soames' work that suggests he has such a fine-grained distinction in mind, but if he has, then Ridge can be seen as proposing a modified theory with a more liberal understanding of representation. Without any deep and substantial argument that propositions *must* be understood in the robustly representational manner available to realists and unavailable to expressivists, such an insistence begs the question just as much as the insistence that all Russellian and Fregean propositions *must* be believable absent desire-like states. If propositions are types of mental event, the burden of proof once again lies with those who wish to prohibit the use of moral propositions in expressivist theory.

Let us take stock. The aim of this chapter is to characterise expressivism. We began in §1 with the idea that moral sentences express desire-like states. This is a good start, but it could not be the whole story because desire theorists and hybrid realists agree with this. How about characterising pure expressivism as the view that moral sentences do not express beliefs? This, we saw in §2, will not do. On the hypothesis that minimalism is correct, expressivists can agree with the core claims of realism: that there are mind-independent moral facts, that moral judgements are beliefs about these things, and that moral sentences express these beliefs. Furthermore, moral beliefs are expressed in exactly the same way that 'grass is green' expresses the belief that grass is green. In this section we explored the possibility that expressivism is a theory of the meaning of moral terms or of how moral

is (i), to predicate redness of *o*, and thereby to entertain the proposition that *o* is red (ii), to negate the property being true, and (iii) to predicate the resulting property not being true of that proposition."

¹⁴ Alwood (2016: 19).

terms get their meanings. I argued that expressivism is in fact compatible with the majority of semantic and metasemantic theories as well as different metaphysics of propositions, not in competition with them, so expressivism is not best understood as a semantic or metasemantic theory. This is huge. As Richard Joyce (2016: §1) writes, expressivism “is usually defined as a thesis about moral language.”

If expressivism cannot be distinguished via its distinctive semantics or metasemantics, the problem we must return to is that of creeping minimalism. In this next and final section we will examine some of the main proposed solutions to the problem and I will show why they all fail. I will then argue for a solution which yields the following answer to the chapter’s title question: expressivism is the view that moral beliefs are *necessarily* desire-like states. As Ridge (2006a: 307) suggests, hybrid expressivists are usefully divided such that “ecumenical cognitivists” think the truth of a moral belief necessarily aligns with the truth of its constituent belief whereas “ecumenical expressivists” deny this.

4. Halting Minimalism

We are trying to figure out what expressivism is. In the Good Old Days it was viewed as the theory that moral declarative sentences are not truth-apt assertions of moral propositions but are used instead to express desire-like states. Due to the considerations in the previous two sections we should not subscribe to this characterisation. On our working assumption that minimalism is the correct theory of truth, moral declarative sentences *are* truth-apt assertions of genuine moral propositions and are used to express *beliefs* in those propositions. We have also seen how it is open for expressivists to believe in moral facts that are mind-independent, and so we are still stuck with the problem of how to distinguish between an expressivist such as myself and a realist.

There are several proposals I will refute. The first is the idea found in Blackburn's work that the realist and the quasi-realist end up in the same place but it is their different starting points that distinguish them. I then discuss Dreier's "explanation" explanation, Asay's appeal to truthmakers, and Sinclair's functional distinction in the psychology of moral judgements. Once these four solutions have been rebutted I provide my own and end by speculating how I envisage this to affect the dialectic between expressivists and their opponents.

4.1 The Journey is Not the Destination

Dreier states that 'Blackburn's famous slogan is, "It's not what you say at the end of the day, but how you got there that matters."' ¹ The theme can be felt throughout Blackburn's work but especially in his (1993) essays where he attempts to close the gap between realism and anti-realism in several domains, not just in metaethics. If he's successful in reducing the gap to zero, has Blackburn not become the very thing he was against in the beginning? In response to this problem he tells us his "reaction is to distinguish what we start with in the construction, and what we finish with" ² – he caricatures the realist as beginning her enquiry with moral facts taken for granted and then proceeding to defend them from arguments against their existence and against our ability to know about them, whereas

¹ Dreier (2004: 32).

² Blackburn (1993: 9).

the anti-realist begins his story with human beings in their natural environment and gradually earns the right to speak of moral facts on the basis of, for example, all we have discussed in this chapter and in the two that follow. But if they end up in the same place saying the same things, why should we take sides? Blackburn (1993: 34) writes “there may be reason to sympathize more with the anti-realist. He has earned the concepts associated with objectivity, while his opponent merely stole them; he has founded our practices on known facts about human capacities, while his opponent invents more.” In Blackburn’s eyes, it is a *methodological* difference that separates the realists and the anti-realists, and the realist methodology assumes too much at the outset.

I agree with Blackburn’s assessment that the methodologies differ and that the anti-realist takes the more humble route to vindicating realistic-sounding moral practice. However I think the destination *does* matter, and it matters more than the journey when it comes to making distinctions between theories. For example, I believe the general theory of relativity. My journey is less impressive than Einstein’s – I worked nowhere near so hard. There is good reason to “sympathise more” with Einstein than with me. Nevertheless, despite our routes being different, we both believe (or believed, in Einstein’s case) *the exact same theory*. Catholics may base their religious judgements on testimony, faith, philosophical argument or combinations of these, yet it is natural to describe them as believing the same tenets. The task at hand is not to distinguish methodologies but to say what makes it the case that expressivism and realism are different *theories*, rather than different ways to justify the same theory. Indeed, when people justify the same theory in different ways they are normally seen as working toward the same goal, whereas realists and expressivists are seen as rivals. Intuitively there is not only a difference in their journey but in where they have arrived as well. I turn now to Dreier’s own attempt.

4.2 The “Explanation” Explanation

Dreier, drawing on hints in the works of John O’Leary-Hawthorne and Huw Price (1996), Kit Fine (2001) and Gibbard (1990, 2003), argues that the best way to distinguish realism and expressivism involves the metaphysical explanation of what makes it the case that a person has a particular moral judgement. According to Dreier, realists will cite moral facts in their explanations whereas expressivists will not. Why does Sandy have the

moral belief *p*? “Because,” says the realist, “there’s a moral fact out there and a mechanism via which Sandy detected it.” “Because,” says the expressivist, “Sandy is a feeling being, brought up with a certain sensibility and has directed her conative attitudes thus and so.” Dreier’s solution to the problem of creeping minimalism is to contend that expressivists think “to explain what it is to make a moral judgement, we need not mention any normative properties.”³ Realists, according to Dreier, must think we *do* need to mention normative properties to explain such beliefs.

The first thing I want to draw attention to is the availability of moral properties and facts for expressivist explanations. Why does Sandy think we should avoid misgendering people? “Because,” says the expressivist, “it’s *true* that we should avoid it; Sandy’s moral sensibility is on point, so if it weren’t a fact then she wouldn’t believe it.” Is the expressivist being insincere when citing moral facts in their explanation of Sandy’s moral belief? No. Expressivists will understand such an explanation to be an expression of some desire-like attitudes. First of all, it involves agreeing with Sandy in her disapproval of misgendering people. Secondly there is an approval of Sandy’s moral sensibility – roughly, approval of forming moral judgements in the way that Sandy does. But the idea that acceptance of explanations can be desire-like in nature is perfectly benign. At least as benign as expressivism, anyway.

Dreier’s thesis is not that expressivists are barred from citing moral facts and properties in their explanations of moral beliefs, but that realists *do* need to cite them. I dispute this. Of course the realist *can* explain Sandy’s belief in the exact same way the expressivist did in the paragraph above, but I see no reason to suppose they have to *always* use this explanation. They might use a sociological explanation, choosing to cite Sandy’s upbringing in liberal, left-wing circles. In the distant future they might choose a neuro-psychological explanation, scanning Sandy’s brain and identifying the neural pattern that explains why Sandy believes we should avoid misgendering people. If all reduces to physics, perhaps even a bottom level physical explanation involving fundamental particles is possible in principle. To put the point another way, robust realists, with the exception of Graham Oddie,⁴ believe in the causal closure of the non-normative

³ Dreier (2004: 39). See also Dreier (2015b: 289).

⁴ Oddie (2005: 2) writes that “mind-independent, irreducible value facts are not idle bystanders, but are fully paid-up contributors to the causal network. Values can affect us, causally, and it is through their causal impact on us that we have knowledge of value.”

world.⁵ Given that causal explanations of our mental states are possible, it must also be possible to give a non-normative explanation of those mental states.⁶

Since expressivists and realists can both use normative and non-normative explanations it may seem Dreier's "explanation" explanation is doomed, but there is one final thing to consider before we abandon it. Dreier (2004: 35) asserts "we want the most illuminating explanation," or "the best explanation" (2015b: 289). Perhaps the thought is that even though both kinds of explanation are *available* to both the realist and the expressivist, the realist will think normative explanations are better than non-normative ones, and *vice versa* for the expressivist. Unfortunately this won't do. I can imagine agreeing with a realist about which explanation is better, and the choice of explanation is going to be context-dependent. Sometimes it might be much better to explain someone's judgement by citing the moral facts, perhaps so that we convey to our audience that her judgements are indeed correct ones. Conversely, when I explain the perverse moral judgements of the alt-right, I am unlikely to cite any moral facts in my explanation but instead the socio-political conditions that gave rise to their movement.

Even if we suppose for a moment that the best explanations for moral belief differ according to the realist and expressivist frameworks, this leaves something unanswered: what features of the theories *make* one explanation better than the other? If there are such features, the distinction is already found. If explanation rankings are brute instead, it is unclear why this is of any theoretical importance, especially given how realists and expressivists can agree that both normative and non-normative explanations of moral beliefs may be *true*. I conclude that Dreier's "explanation" explanation fails.

4.3 Moral Truthmakers

Asay (2013) argues that the distinction between realism and expressivism lies not in the semantics or the psychology but in the metaphysics. More specifically, the distinction is to be found in truthmaker theory. What is a

⁵ This gives rise to the "cosmic coincidence" worry for robust realists (Bedke 2009). If our moral intuitions are caused by purely non-moral states of affairs, how can we be justified in relying on them? Thanks to Stephen Ingram for this reference.

⁶ Indeed, if it is true that the error theorist and the robust realist agree about everything except the existence of moral facts, then moral beliefs must be explainable without referring to moral facts – unless error theorists are unable to explain moral beliefs.

truthmaker? Truthmakers are the facts that *make* true sentences true. Roughly, for any true sentence S there is a fact F which is responsible for its truth. It is not an uncontroversial thesis, but let us accept it here for the sake of argument. Why is the sentence ‘I am in Sheffield,’ as I currently write it, true? Because of the fact that Graham Bex-Priestley is in Sheffield. Let us see how this idea might be utilised in metaethics.

Except for the error theory, which we shall put to one side, all contemporary metaethical views allow that some moral sentences are true. Let us suppose we all agree that ‘torture is wrong’ is true. Asay (2013: 220) first notes that some metaethical theories can be distinguished via what fact they believe *makes* that sentence true. Subjectivists might think the fact that they themselves disapprove of torture is what makes ‘torture is wrong’ true; contractualists might say it is the fact that under certain idealised conditions we would agree not to torture each other; natural realists will say it is the fact that torture has some natural property such as failing to promote happiness; and robust realists will say it is the fact that torture has the irreducibly moral property of wrongness. What will quasi-realist expressivists say? Asay thinks they will say something very much like the natural realist. Drawing from passages such as “[the expressivist] thinks that what makes it wrong to kick dogs is that it causes them pain”⁷ Asay argues that expressivists believe moral sentences are made true by natural facts. This sits well with my own view. In chapter 3 I will argue that expressivists *must* believe moral facts are reducible to descriptive facts such as natural ones. What, then, is the difference between expressivism and natural realism?

Asay (2013: 227) proclaims that “even if naturalists and quasi-realists can agree on what makes moral judgments true, they may disagree as to *why* those objects and the truths they make true stand in the truthmaking relation. In other words, they may disagree as to how those objects get to be truthmakers for moral truths.” On his understanding of the expressivist position (*ibid*: 228), “the reason why certain natural facts stand in the truthmaking relation to moral facts is due to our projections,” our own sentiments. So, while natural realist Nigel and expressivist Erica might agree that the truthmaker of ‘torture is wrong’ is the mind-independent natural fact that torture fails to promote happiness, only Erica thinks the

⁷ Blackburn (1981: 179).

reason this particular fact is assigned as the truthmaker of that particular sentence is due to how we feel.

The problem with this is that it threatens to collapse the expressivist position into the subjectivist one they were trying to avoid. Asay (*ibid*: 230) is aware of this consequence but seems to embrace it: “Moral realists are those who (i) admit that there are moral truths, (ii) accept that those truths are made true by a realism-relevant (i.e., mind-independent) reality, and (iii) hold that the nature of the truthmaking relation that obtains between those truths and that reality is itself of a mind-independent variety ... and quasi-realists reject (iii).” But why would expressivists accept that the truthmaking relation is mind-dependent? He thinks that having mind-independent truthmakers is good enough. I disagree.

Consider someone who, due a projection of their peculiar passion for pineapples, thought the truthmaker for ‘torture is wrong’ was the fact that it failed to produce enough pineapples. Would they have assigned the correct truthmaker? Of course not. The wrongness of torture has nothing to do with pineapples. Importantly, the *correct* assignment of the truthmaker has (or might have) nothing to do with our sentiments. What we in fact assign as the truthmaking facts of moral sentences has everything to do with our sentiments, but our assignments could be, and often are, wrong. In short, accepting mind-dependence at the level of the truthmaking relation is going to have similar subjectivist consequences as accepting mind-dependence at the level of truthmakers. So, while Asay’s proposal would differentiate the expressivist from the realist, it is not one that expressivists would be happy to accept. We should look for a solution that avoids troubling commitments to mind-dependence in our moral metaphysics.

4.4 The Function of Moral Belief

I argued in §3 that expressivism is not a semantic or metasemantic theory. I then argued at the beginning of this section that expressivism should not be distinguished via the way its proponents justify it, for intuitively there is a difference in the end thesis itself. In the previous two subsections I argued against proposals to define expressivism metaphysically. The only realm left is the psychology of moral judgement. Sinclair (2006) searches here for a solution, and although he is looking in the right place, it is not quite right.

Sinclair (2006: 254) begins with a terminological distinction between ‘minimal beliefs’ and ‘robust beliefs.’ Minimal beliefs are what I described in §2.3 – they are the mental states expressed by declarative sentences in a discourse that is syntactically disciplined. Robust beliefs are the kinds of states realists think are expressed with moral sentences and pure⁸ expressivists *deny* are expressed with thin, atomic moral sentences. I will adopt this terminology.

Rightly, Sinclair rejects the idea that robust beliefs are those that are criticisable when they fail to be true whereas minimal beliefs are those that are not. Expressivists can and should allow that the minimal belief that, say, abortion is wrong is not true and thereby criticisable on (at least) that basis. Instead, Sinclair proposes to understand robust beliefs as those that aim at truth, and “a mental state aims at truth when it is the product of a mechanism whose function is to produce states of mind whose content matches the state of the world.”⁹ We could put the idea in terms I introduced in §1.1: pure expressivism is the view that moral beliefs are not BELIEFS but DESIRES. But this is wrong.

Recall the two main characterisations of BELIEFS. On descriptive views such as Smith’s, the BELIEF that *p* is such that the apprehension that not *p* tends to cause the extinction of that state. This happens with moral beliefs as much as any other. If you believe that it is wrong to decriminalise drug use and then it becomes apparent to you due to further investigation and philosophical argument that it is not wrong after all, then you will have a tendency to change your mind. On normative views such as Gregory’s, the BELIEF that *p* is such that the fact that not *p* gives an objective reason to abandon that state. To use the same example, expressivists can and should accept that the fact it is not wrong to decriminalise drug use gives you an objective reason to stop believing it *is* wrong. Therefore expressivists should not only think that moral judgements are beliefs in the common minimal sense, but they should also think they are BELIEFS in the technical Humean sense.

The implication of this is that pure expressivists must think of moral beliefs as single, simple mental states with both directions of fit, correctly characterisable as BELIEFS and as DESIRES; the belief that torture is wrong

⁸ Hybrid expressivists think moral sentences express minimal beliefs that have robust beliefs as a component, which only complicates matters here. We will get to them later.

⁹ Sinclair (2006: 256).

might be described as both the BELIEF that torture is wrong and the DESIRE to blame people for torturing, for example. In other words pure expressivists, who began their journey with the Humean Theory of Motivation, must end up denying it and proclaiming there *are* BELIEFS that motivate after all – moral beliefs are besires! Insofar as besires are unpalatable, pure expressivists could avoid this by going hybrid. Hybrid expressivists agree that moral beliefs are both BELIEFS and DESIRES but this is only due to their being composed out of simpler mental states each with only one direction of fit. (That being said, I find besires much easier to swallow after discovering this argument, and I imagine many pure expressivists would rather reject the Humean Theory of Motivation than accept that robust beliefs form constituent parts of all moral judgements.)

Is there another way to differentiate the function of robust beliefs that does not rely on the characterisation of 'BELIEF'? Sinclair (2006: 257-8), in saying that robust beliefs are those that function to match the states of the world, also draws upon the notion that part of this function is to *guide* our actions whereas the function of desires is to *drive* our actions. According to this idea, robust beliefs can be thought of as a kind of map. Robust beliefs tell you the lay of the land; desires tell you where to go. Why did I go to the fridge? I was driven by my desire for hounous and guided by my belief that hounous was in the fridge. Unfortunately this proposal won't work either. Moral beliefs can guide us just as much as robust beliefs. In Brian Bolland's (1996) short story 'An Innocent Guy,' a man struggles with the thought that he might only be avoiding wrongdoing out of fear and not because he's a good person. Because "anything done out of fear has no moral value," he declares "I've decided that just once I wanna do a really bad thing. I mean a *really seriously bad thing*." In this case, his desire to do something bad is *guided* by his moral belief that killing the Batman would be really seriously bad. That's not to say that his moral belief doesn't also drive him (in the opposite direction) by itself, and indeed the tugging in different directions is apparent in the comic. Expressivists should admit that moral beliefs have both directions of fit. But when we desire to do something good or bad or courageous or vengeful, our moral beliefs undoubtedly provide pieces of the map with which we guide our actions. The function of moral belief fails to halt minimalism.

4.5 The Necessity of Desire

Sinclair was right to look at the psychology of moral judgement, but he was wrong to think that expressivists must view the function of moral beliefs any differently to the kind of realist who thinks moral beliefs have the power to motivate us to act. It is true that the expressivist and this kind of realist arrive here from different starting points. This kind of realist began with the thought that moral judgements are beliefs and noticed they motivate us like desires, and so called them 'besires'; the expressivist began by noticing how we tend to act in line with our moral judgements which was evidence of their being desires, and then earned via minimalism the right to call them beliefs, and so we might call them 'deliefs'. I wondered whether there may be some way to distinguish besires and deliefs, perhaps to do with grounding. The idea was that besires are desires *in virtue* of being beliefs of a certain kind whereas deliefs are beliefs *in virtue* of being desires that behave in a certain way. Alas, I cannot see a way to spell this out convincingly so I have abandoned that line of thought. Instead, I make use of a creature called Amy.

Imagine a godlike character (this is Amy) who is motivationally inert. Perhaps she exists outside of time, but wherever she resides she doesn't act. She has no DESIRES. At the same time, Amy *does* have BELIEFS, and they're as plenty and as accurate as can be. I am not sure if such a being is possible. Nevertheless, we can imagine she is. Amy is incredible – she knows everything there is to know about black holes and DNA and climate change and so forth (not that she would ever tell us). The question is this: can inert Amy make moral judgements? Expressivists of all kinds will say 'no'. For expressivists, moral judgements are necessarily desire-like states, so a being without the relevant desires does not make moral judgements. There may be others who would answer 'no', such as the sentimentalists who think certain desires are necessary for the formation of moral beliefs but the moral beliefs themselves are not desire-like.¹⁰ My contention here is that the common commitment among expressivists is that, necessarily, all moral judgements are DESIRES.

What, then, of the besire theorists who would also say our moral judgements are DESIRES? The key word in the previous sentence is 'our'. The fact that our moral beliefs motivate us might be, in their eyes, a peculiar

¹⁰ Arguably Prinz (2004, 2007) and Tappolet (2011) fall in this category.

fact about our psychology. There is nothing in their theory that rules out Amy making moral judgements. It is consistent on their view that Amy has moral beliefs but due to her inertness those beliefs are not DESIRES. It is hard to find textual evidence in support of this. Russ Shafer-Landau (2003: 147) writes that moral beliefs “might be intrinsically motivating without being *necessarily* motivating.” However, by this he only means that our moral beliefs may exert no influence over our actions due to, for instance, “competing beliefs or desires, physical exhaustion, severe depression, etc.” (*Ibid.*) Yet, due to informal conversations with both Shafer-Landau and Philip Stratton-Lake (another moral realist who believes in besires), I believe both would be open to the idea that Amy could “see” moral reality and form moral beliefs about it. Therefore, while *our* moral beliefs are (at least often) DESIRES, on their view moral beliefs do not *need* to be DESIRES. This is what distinguishes their theory from expressivism.

Now that we have defined expressivism, we have the thin end of a wedge. We can use it to “propagate backwards” and introduce terminology that will help us in the remaining chapters of this thesis. Following Sinclair, I will use the word ‘robust’. Robust beliefs are those that Amy can form. While expressivists ought to think moral beliefs are genuine BELIEFS, they should not think that they are *robust* beliefs. Moral beliefs are true and representational, but they are neither robustly true nor robustly representational. However, moral facts are robust facts! As mentioned in §4.3 and as I will argue in chapter 3, moral facts are bog-standard descriptive facts such as the fact that something promotes happiness or is willed by God. Amy *can* therefore “see” moral reality, but she cannot see it *as* moral reality. As an analogy, a thousand years ago people could form true beliefs about H₂O, but not under that description. They knew *de re* that the lake was full of H₂O because they knew it was full of water, which (we now know) is H₂O. Similarly, Amy might know *de re* that torture is wrong because she knows it has descriptive property D (whatever the wrongness property turns out to be), but she does not have the *de dicto* belief ‘torture is wrong.’

In the following three subsections I will discuss some theories that come out as expressivist on this definition yet are *prima facie* distinct. I bite the bullet and admit they are the same or I provide a further subdivision.

4.5.1 Necessary Besire Theorists

If I am wrong about Shafer-Landau and Stratton-Lake, or if I am wrong to generalise from them and in fact there exist besire theorists who believe moral judgements necessarily are DESIRES, then I am happy to say that these metaethicists are expressivists. The reason they appear different is simply because they have a different starting point and have used different arguments to defend their view. I do not think there is anything they can point to in their theory that separates it from expressivism. I would consider them expressivists who have not bothered to solve the Frege-Geach problem – that is, they haven't adequately described the mental state expressed by, say, 'if lying is wrong then getting little brother to lie is wrong' and how it combines with the belief that lying is wrong to produce a brand new DESIRE. I discuss the Frege-Geach problem in the following chapter and, in arguing against the pure expressivist, I believe I am actually also arguing against besire theorists!

4.5.2 Hermeneutical Fictionalism

Moral fictionalism is the view that moral discourse is fictional discourse. All substantive moral sentences, if taken literally at their face value, are strictly speaking untrue. The revolutionary fictionalist says that, in light of this, we ought to engage in moral discourse as a fiction, a pretence. The hermeneutical fictionalist believes this is what we've been doing all along. On this view, we haven't been *believing* that torture is literally wrong; we've been make-believing, or pretending it was the case, or believing what would have been the case if the moral fiction were true.

There are many ways to be a hermeneutical fictionalist, but here is one way that seems to make it a version of expressivism and it is inspired by Mark Kalderon's (2005) work. The idea is that, when engaging in moral make-believe, there are still rules of the discourse. The condition when it is linguistically appropriate to utter 'torture is wrong' (a sentence which is not literally true) is when you have a particular non-cognitive attitude towards torture such as disapproval. (Once again, this is the kind of theory that needs to solve the Frege-Geach problem, as raised by Lenman (2008).) This kind of fictionalism construes moral judgements as necessarily desire-like states.

I am inclined to think that this version of fictionalism is a kind of expressivism, but it is an interesting (and incorrect) subcategory. Here are two possible ways to subcategorise it. Firstly, a difference might be found in the pragmatics of language. Contemporary expressivists think moral language can be genuinely assertoric, but I wouldn't want to build that into the definition of expressivism because expressivists have thought moral language is primarily emotive or is primarily prescriptive. Fictionalists, then, may want to distinguish full blown assertions from quasi-assertoric utterances within the fiction. Secondly, they might distinguish the content – on this view, 'torture is wrong' is a genuine assertion but of something like 'according to the moral fiction, torture is wrong.' I have argued that expressivism is compatible with many semantic theories of the meaning of moral sentences, so as long as the mental state expressed is necessarily a DESIRE, this is a version of expressivism combined with an incorrect theory of what moral sentences mean. Finally, they might accept that moral sentences can be genuine assertions that are literally and strictly speaking false. This theory could be arrived at by combining expressivism with a rejection of deflationary theories of truth. The idea would be that acceptance of 'torture is wrong' is a desire-like state but accepting that it is *true* is to do something more than this, and this extra condition is not warranted. Most expressivists are against the idea that morality is systematically untrue, but the fictionalist might not mind. "For the fictionalist, literal falsity is simply not a defect and literal truth as such is not a virtue."¹¹ So there are possible varieties of fictionalism that are kinds of expressivism. Perhaps this is an interesting avenue of research for those sympathetic to expressivism who also find the arguments against truth minimalism compelling.

4.5.3 Hybrid Expressivism

Even realists can be expressivists. Copp (2001, 2018) raised the possibility of realists positing desire-like states that are expressed by moral sentences *in addition* to robust beliefs. However, since Copp thinks these desire-like states are only *implicated* rather than being a necessary part of the moral judgement itself, Copp does not count as an expressivist proper – he would admit that Amy can make moral judgements despite not having the relevant desire-like states. Due to the way I have argued we should define

¹¹ Rosen (2005: 16).

expressivism, the group Ridge (2014: 82) labels “Implicative Ecumenical Cognitivism” (we’ll get to the term ‘ecumenical cognitivism’ in the next paragraph) does not count. The group that *does* count are what Ridge (2014: 85) calls “Constitutive Judgment-Individuating Ecumenical Cognitivism” according to which “the relevant non-cognitive attitude *partially constitutes* – is a proper part of – the normative judgment in question.”

Ridge (2014: 80) proposes that we divide hybrid expressivist views into two camps: ecumenical cognitivists are those who take the content of a moral claim to be the same as the content of the robust belief it expresses, and ecumenical expressivists are the rest. Here is the simplest way to be an ecumenical cognitivist: take any realist (or other cognitivist) theory you like and add the condition that fully-fledged moral judgements must also involve some specified DESIRES, such as a policy to do right and avoid wrong. Amy, who has no policies at all, may be able to form BELIEFS about moral reality but she cannot make fully-fledged moral judgements. According to these theories, moral sentences necessarily express DESIRES and as such they are varieties of expressivism.

Examples may help. Consider the sentence ‘chemical weapons are bad.’ A possible ecumenical cognitivist might say that it expresses disapproval of chemical weapons and the robust belief that chemical weapons have the irreducible property of moral badness; the sentence is true just in case the robust belief it expresses is also true. Another ecumenical cognitivist might say it expresses the plan to minimise bad things and the robust belief that chemical weapons promote pain over pleasure; the sentence is true just in case the robust belief it expresses is also true. A possible ecumenical *expressivist* (inspired by Ridge 2006a) might say it expresses, in the mouths of some hedonists, disapproval of things insofar as they’re painful and the robust belief that chemical weapons are painful; the sentence is not necessarily true in all and only the cases when the robust belief is true. Another ecumenical expressivist (such as Toppinen 2013) might say it expresses the higher-order state M where M can be composed by many possible pairs of robust beliefs and desires by different substitutions for K: disapproval of things insofar as they have property K and the robust belief that chemical weapons have that property. There is no particular robust belief expressed by the sentence, but the robust belief composing M in a particular person may not be true even when the sentence itself is.

The reason why this divide is a natural one is because the ecumenical cognitivists face the same objections pure (non-hybrid) cognitivists face (such as Moorean open questions and worries about queerness) except for arguments based on moral motivation, whereas ecumenical expressivists face the same objections pure expressivists face (such as challenges to account for disagreement (chapter 4) and fear of error (chapter 5)) except for the Frege-Geach problem (although there are still residual Frege-Geach style worries for ecumenical expressivists (chapter 2)). The lesson I take from chapters 2 and 3 of this thesis is that the best version of expressivism is ecumenical.

Conclusion

This brings me to the end of this section and this chapter. Here is what we have discovered so far. In §1 we discussed the characterisations of BELIEF and DESIRE and I argued that we can make use of Schroeder's assertability expression relation to cash out the notion that moral sentences express DESIRES. In §2 we examined how truth minimalism makes it difficult to distinguish expressivism from realism, since it allows contemporary expressivists the option of embracing several tenets once thought definitive of realism: that moral sentences express truth-apt beliefs about mind-independent moral facts. We then turned to the possibility of defining expressivism as a unique theory of the semantics or metasemantics of moral language, but I argued that expressivism is in fact *compatible*, not in competition, with the majority of theories in these domains. Finally, in this section we tried to halt minimalism. I argued that all previous attempts fail to do so, yet minimalism *can* be halted by motivationally inert Amy. I ended by discussing the key distinction between hybrid expressivist views and why it was important.

What is expressivism? It is the theory that, necessarily, moral judgements are desire-like states. This is what I have argued for in this chapter, and if it seems obvious to you, it has not been obvious to those who believe expressivism has strong (and unpalatable) commitments in the philosophy of language and to those who have struggled with the problem of creeping minimalism. It also means that some views thought to be in opposition to expressivism are not, such as potential desire theorists and

hermeneutical fictionalists. Champions of these views need to start working with expressivists in tackling issues such as the Frege-Geach problem.

What is the *advantage* of expressivism? It is no longer clear. A traditional reason to favour it was the power it has to explain the apparently reliable connection between making a moral judgement and being at least partly motivated to act in accordance with it, but ecumenical cognitivists have that power too. If expressivists can provide an adequate theory of disagreement (a challenge I take up in chapter 4), they may well have an advantage over subjectivist or relativist theories, but most cognitivists seem to avoid this trouble. Another traditional reason to favour expressivism was its rejection of metaphysics, but in chapter 3 I'll argue that we expressivists cannot escape it. In fact we ought to *embrace* metaphysics because as far as I can tell, it is one of our best hopes for an edge over everyone else. To foreshadow, it allows us (indeed *compels* us) to deny the existence of irreducibly normative properties without running into the problems with reference that plague the realists who also deny irreducible normativity. The other advantage in my eyes is the anti-sceptical argument I offer which rests on analysis of the thought that we might be mistaken. This waits for you in the final chapter. In the next chapter I show why pure expressivists have strong reasons to become ecumenical.

~2~

The Problem with Purity

Expressivists believe atomic declarative sentences in some domain express desire-like attitudes such as plans, commitments, goals, intentions, or feelings of approval and disapproval. Expressivists these days tend to talk about the normative in general, but I am mostly interested in expressivism about moral thought and discourse. In this chapter I argue against pure versions of expressivism according to which atomic, thin moral sentences such as ‘torture is wrong’ express *only* desire-like states rather than robustly representational beliefs.

The problem with purity is the Frege-Geach problem. The structure of this chapter is to first examine what the problem is and then to raise objections to several proposed solutions. Some of this is history. Bob Hale and Mark van Roojen have caused plenty of trouble for Simon Blackburn’s “higher order attitude” solution and I believe James Dreier has put to rest Paul Horwich’s “minimalist” solution. It is nevertheless worth discussing them, especially since I will argue that Dreier’s strategy also puts pressure on the later “inferentialist” solutions of Blackburn and Allan Gibbard. To anticipate, it can’t be the case that any old sentence or thought can help itself to inferential logical connectives – the viable sentences and thoughts must have some property that *legitimizes* the use of logical connectives. It could be argued that the moral attitudes expressed by moral sentences do indeed do so because they direct us, and directions really ought to be satisfiable. Unfortunately Blackburn and Gibbard’s anchor of unsatisfiable attitudes turns out to make many apparently sensible moral judgements incoherent, which is implausible. Finally I turn to Mark Schroeder’s “structural” solution (the shadows of which are present in Blackburn, Gibbard and James Lenman’s work) and argue that, first of all, the moral attitude is

mischaracterised, and second, it yields counterintuitive results about people who think they shouldn't believe something they actually do believe. I end by showing how hybrid expressivism can solve the Frege-Geach problem in the face of objections from Schroeder, Lenman and Caj Strandberg, allowing me to conclude that the problem is purely with purity.

1. The Frege-Geach Problem

Expressivism is often characterised as explaining the meanings of terms via the states of mind those terms express instead of, for instance, analysing the truth conditions of sentences in which they figure. Let us consider the proposal that 'torture is wrong' expresses disapproval of torture. The first trouble to notice is that one can take many stances towards this sentence beyond acceptance or endorsement, and presumably one only disapproves of torture when one accepts or endorses the sentence. One may *wonder* whether torture is wrong, or suppose it for the sake of argument, which might be conveyed by asking the question 'is it the case that torture is wrong?' or using the sentence in the conditional 'if torture is wrong then we should stop extraditing prisoners to America.' Peter Geach (1960, 1965) took "the Frege Point" that sentences may have the same meaning in unasserted contexts as in asserted ones (the same content can be given different "force," such as the assertoric force or the interrogative force) and used it to argue against expressivism. Geach's illustration was an instance of *modus ponens*:

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

We first note that ' p , if p then q , therefore q ' is valid whereas ' p , if r then q , therefore q ' is not. For *modus ponens* to work, the meaning of 1 must be the same as the meaning of the antecedent in 2 despite the fact that the assertoric force has been "stripped" from 'lying is wrong' in the second premise. And yet, even if we grant that 1 expresses disapproval of lying, someone who accepts 2 does not necessarily disapprove of lying. So, if the meaning of 1 is exhausted by the expressed mental state of disapproval of lying, and there is no such disapproval expressed in accepting claim 2, then it seems the meaning of 1 does not figure anywhere in 2, and so the

expressivist cannot say this is an instance of *modus ponens*. Since it clearly *is* an instance of *modus ponens*, this counts against expressivism.

The point also applies to descriptive, non-normative cases. When I say ‘if the gardener is innocent then the butler did it,’ where ‘innocent’ is understood descriptively as *someone who didn’t break the relevant law*, I am not expressing my belief in the innocence of the gardener, and yet the meaning of the antecedent is the same as when I assert ‘the gardener is innocent’ independently and thereby express my belief in it.¹ However, in the ordinary descriptive case there are ways to explain the sameness of meaning in embedded sentences that may not be open to the expressivist. The standard method is to posit some entities as the propositional contents of sentences (such as sets of possible worlds, for instance) which explain why ‘I am tired’ has the same meaning as ‘je suis fatigué.’ Sentential operators correspond to operations on the propositional contents (such as a negated proposition being its complement set of possible worlds, and a disjunction of two propositions being the union of the two sets). The final result is that each well-formed compositional sentence has as its meaning a complex proposition as a function of its parts, and one expresses belief in the complex proposition when one asserts the compositional sentence (such as – running still with the possible worlds model – thinking the actual world is a member of it). Furthermore, the inferences we make in our beliefs are explained by relations between propositional contents (such as set-theoretic operations like complementation, union and intersection).

So what is the problem for expressivism? The mental state expressed by an atomic sentence like ‘the gardener is innocent’ or ‘lying is wrong’ is not also expressed by conditionals in which those sentences figure as clauses such as ‘if the gardener is innocent then the butler did it’ or ‘if lying is wrong, getting little brother to lie is wrong.’ Given that the same “problem” applies equally to language that is not given expressivist

¹ Conditionals may not be the best example. There are many different theories of how to interpret the indicative conditional as it appears in natural language. Most reject a purely truth-functional account (with notable exceptions such as Jackson 1979, 1987 and Lewis 1986b: 152-6), and some reject truth conditions altogether (Adams 1965, Edgington 1995). Nevertheless, the Frege Point applies to less controversially truth-functional non-assertoric embeddings such as negation and disjunction. The negation of a proposition certainly doesn’t involve asserting that proposition! (I must also neglect the different ways to interpret a denial in natural language. When I deny your claim, do I reject it as a claim to be used in our conversation, do I assert its negation, or do I do something else?)

treatment, the first lesson we should take from Geach is this: if Frege is right that the meaning of a sentence remains the same whether asserted or not, then the meaning of ‘torture is wrong’ cannot simply *be* the state of mind expressed on given tokens of utterance.² That’s not so bad – the expressivist only needs to say that we learn something interesting about moral practice by looking at the mental states involved in accepting moral claims. As I have argued in the previous chapter, expressivism is not a *semantic* theory and is in fact compatible with a number of different ways to understand the meanings of words. If my arguments in that chapter are correct, then not only can expressivists help themselves to any of most current theories in semantics and metasemantics, they can also help themselves to any theory of propositions too (such as sets of possible worlds). Given that the Frege-Geach problem is usually seen as a semantic obstacle to expressivism, is the problem dissolved?

No. Just as recasting expressivism from a semantic to a metasemantic view shifts the Frege-Geach bump along the rug and a parallel challenge arises at the metasemantic level,³ the move I made in the previous chapter to characterise expressivism as a view of moral *thought* and not language shifts the bump *across* the rug. The problem lies in characterising the mental state involved in accepting, say, a disjunctive moral claim.⁴ Cognitivists have several models for representing the contents of atomic and complex sentences, and to believe a sentence is to take one attitude – robustly representational belief – towards those contents. Things are not so simple

² You might think this is too hasty. Perhaps an utterance of ‘torture is wrong’ in the antecedent of a conditional does express disapproval of torture but this mental state is not necessarily had or endorsed by the utterer. I am open to mentalism about meaning, and it wasn’t my point to dismiss it. Let me clarify: given the understanding of ‘expression’ that we’re working with from §1.3.2 of chapter 1 where sentence S expresses mental state M if and only if it is linguistically correct to assert S only when one has M, and given the fact that one may accept a conditional without accepting the antecedent, the mental states expressed (in this sense) by sentences vary depending on embeddings. If Frege is right and meanings remain constant whether embedded or not, meanings are not the *expressed* mental states. This does not rule out their being mental states at all.

³ As Ridge (2014: 138) recognises: “The move from semantics to meta-semantics does nothing to avoid [the Frege-Geach problem].”

⁴ Importantly, this psychological version of the Frege-Geach problem can be applied to some adversaries of expressivism. The ‘desire’ theorists (Shafer-Landau 2003) and some hermeneutical fictionalists (Kalderon 2005) both think that desire-like states are involved when we make moral assertions, and both give standard semantic accounts of the meaning of complex moral sentences which they think excuses them from grappling with the Frege-Geach problem. I disagree. I consider their theories incomplete until they provide a plausible account of the mental states involved when we make complex moral assertions.

for the expressivist. We are owed an account of the mental states that correspond to acceptance of complex moral sentences, especially in sentences where those states of mind expressed by atomic sentences are not (necessarily) present during acceptance, such as negations and disjunctions. Furthermore, this account ought to explain why certain inferences are licensed. It needs to explain, for instance, why it is inconsistent to accept the premises and deny the conclusion of an instance of *modus ponens*. The Frege-Geach problem is therefore best viewed as a challenge: a challenge to give a plausible account of the mental states expressed by arbitrarily complex moral sentences and to explain why the right logical relations hold between them.

1.1 The Negation Problem

The negation problem was illustrated very clearly by Nicholas Unwin (1999). He explained it in terms of an attitude ascription, but I'll stick with the straightforward moral claim that stealing is wrong. If 'stealing is wrong' expresses disapproval towards stealing, what do the following sentences express?

- 1) Stealing is not wrong.
- 2) Not stealing is wrong.

The second sentence is easy – it expresses disapproval of not stealing. The first sentence is the tricky one. Sentence 1 doesn't look like it expresses any disapproval whatsoever, yet a mere lack of disapproval is not sufficient for judging that stealing is not wrong because someone who has not formed any judgement about the wrongness of stealing may not disapprove of stealing. The judgement that stealing is not wrong is not indifference or undecidedness but a positive judgement that stealing is permissible. To judge that stealing is not wrong is to judge *something*. As Unwin shows, judgements of permissibility cannot be reduced to some function of disapproving of stealing because there simply aren't enough places to insert the negation sign. There are only two places to insert a negation into the claim that I disapprove of stealing:

I [do not] disapprove of [not] stealing.

The first bracketed negation corresponds to my failing to judge that stealing is wrong and the second bracketed negation corresponds to my positive

judgement that it is wrong not to steal. If ‘stealing is wrong’ expresses disapproval of stealing, there is no plausible way to couch ‘stealing is not wrong’ in terms of disapproval of anything.

Expressivists need to give an account of what state of mind is involved in accepting sentence 1. Furthermore, this state of mind should explain why it is inconsistent to accept it while simultaneously accepting the claim that stealing *is* wrong.

1.2 Other Embeddings

It doesn’t stop with negation. Expressivists need to give an account of the state of mind involved in accepting sentences containing other logical connectives too. Conjunctions appear easy at first glance. To accept ‘torture is wrong and stealing is wrong’ is simply to disapprove of both torture and stealing, is it not? Well, this is too coarse. One must distinguish between believing a conjunction and believing all the conjuncts – the former may imply the latter (at least in aware and rational agents if not everyone) but the latter does not imply the former. For example, in a raffle of millions, I believe my ticket won’t win. I also believe your ticket won’t win, and I believe my neighbour’s ticket won’t win, and so on. I believe of every ticket that it won’t win, but I do not believe the conjunction ‘Ticket 1 will not win, Ticket 2 will not win, Ticket 3 will not win, ... and Ticket 5 862 277 will not win.’ Merely disapproving of torture and also disapproving of stealing seems to only capture two individual judgements rather than an acceptance of a single conjunction.

Disjunctions are more difficult still, for it is possible to accept a disjunction without accepting either disjunct. Indeed those are the cases where disjunctions are most useful. They are also puzzling. What could it be to think that either stealing or torture is wrong? One need not disapprove of stealing or of torture to accept that disjunction, just like with the conditional given in Geach’s argument. To overcome this, cognitivists can use a single attitude – robustly representational belief – taken towards a disjunctive content. The expressivist can do something similar or something different. They can either posit a single attitude – *belief* in a wider sense, not robustly representational in nature – which is taken up towards different contents (the structural and hybrid solutions are in this category), or they

can posit different kinds of attitudes that constitute accepting different sentences. Either way, the expressivist has explaining to do.

In addition we also have quantifiers such as ‘all’ and ‘some’ to account for, as well as modal claims and tensed claims such as ‘it might have been obligatory to obey your parents hundreds of years ago, but no longer.’ I will not focus on these tricky cases because I feel pure expressivists have trouble much earlier with negation, disjunction, and also mixed sentences. Mixed sentences are those that combine the normative and non-normative. It is clear that we can form judgements expressed by sentences like ‘either abortion is wrong or foetuses feel no pain.’ This puts pressure on the expressivist to give similar treatments of non-normative language, leading Schroeder (2008) to a form of global expressivism as we shall see in §5. The point I wish to emphasise here is that the expressivist needs to account not only for logical connectives and other embeddings of purely normative language, but also for cases where normative and non-normative language are mixed together.

With an understanding of the Frege-Geach problem in hand, and in particular its psychological variant, we are now in a position to examine some proposed solutions.

2. Minimalist Solutions

Philosophers such as Daniel Stoljar (1993) and Horwich (2004) have argued that the Frege-Geach problem is really no problem at all. The basis for their claim is a particular conception of truth, namely *minimalism* or *deflationism* about truth, which we came across in §2.2 of the previous chapter. The thesis comes in many forms, but the common theme is, “roughly, that the traditional search for its *underlying nature* is misconceived, since truth can have no such thing.”¹ To call a sentence true is not to ascribe some substantial, metaphysically robust property to it. Instead, we understand truth by understanding the equivalence between a sentence ‘S’ and the sentence ‘S is true.’ Not just any sentence will do, of course. The sentence must be in declarative form such that it makes grammatical sense to say ‘it is true that S.’² The truth conditions for ‘it is true that S’ are expressed by S

¹ Horwich (2004: 1).

² We ignore sentences that don’t embed fluently into that construction such as ‘here I am’: it is true that I am here but it is unnatural to say ‘it is true that here I am.’ Nevertheless,

and there is nothing above and beyond truth than this (or something like this); grass is green is true if and only if grass is green, it is true that ravens are fierce if and only if ravens are fierce, and so on for any sentence “embeddable in the contexts ‘it is true that...’, and ‘it is false that...’, where by ‘embeddable’, I mean that placing the sentence in these contexts results in a grammatical sentence of English.”³ What about falsity? Well, it is false that grass is green if and only if grass is *not* green. The search for anything more complicated than this, based on correspondence to facts or maximal coherence for example, is misguided. Deflated truth conditions are taken to come as cheaply as the right sort of sentence structure.

How does this help the expressivist? Most expressivists tend to adopt truth minimalism in order to explain why, even if moral sentences like ‘torture is wrong’ express states such as disapproval, it is legitimate to speak of moral truth. I assent to ‘it is true that torture is wrong’ because I assent to ‘torture is wrong’ and the two are semantically equivalent. And, as we saw in §2 of the previous chapter, we needn’t stop there. Why not deflate other notions such as ‘fact’? I also assent to ‘it is a fact that torture is wrong’ for the same reason I assent to ‘torture is wrong.’ Minimalism creeps until we end up seemingly accepting the very tenets of realist positions. Whether this is a problem or not, it is not the issue here. What we’re currently interested in is the idea that if the expressivist can help herself to the truth of atomic moral sentences, the meanings of complex moral sentences and the states of mind involved in accepting them can be explained in terms of the potential truth or falsehood of the components.

Let us see how it works in the case of disjunction. If we want to understand the meaning of ‘either foetuses feel no pain or abortion is wrong’ we explicate the truth conditions of the whole sentence. The sentence is true just in case one of the two disjuncts is true. We know what the truth of ‘foetuses feel no pain’ would amount to, and thanks to truth minimalism we know what it takes for ‘abortion is wrong’ to be true - it is true if and only if abortion is wrong, and to judge this to be the case is to disapprove of abortion. Granting that we know the meanings of the disjuncts, minimalist solutions to the Frege-Geach problem hold that we can understand the meanings of disjunctions simply by knowing the truth table:

while not every truth-apt sentence *S* can be embedded into ‘it is true that *S*’, the other way around (moving from ‘it is true that *S*’ to ‘*S*’) is fine.

³ Stoljar (1993: 83).

disjunctions are true just in case any of their disjuncts are true. The solution extends to other truth-functional connectives naturally.

The idea then is first to notice that ‘is wrong,’ ‘is good’ and so on are, grammatically speaking, predicates, and so they naturally embed into the more complex sentences we have been discussing. We understand the parts individually and we understand the connectives, therefore we understand arbitrarily complex sentences involving normative terms. In keeping with the general theme of Horwich’s philosophy the problem is not really solved but *dissolved*: “once we have supplemented the expressivist analysis with the principle that ‘rational’ [or any other normative adjective] is a logical predicate, there is no reason to suspect that there are constructions involving that term whose deployment cannot be explained. ... In other words, if our entire use of the term ‘rational’ can be explained on the assumption that the rules of use we have in mind are simply (1) ‘Y believes that x is rational’ means ‘Y has a pro-attitude toward x and ...’ and (2) ‘rational’ is a logical predicate, then we may conclude that it is in following these rules that our knowledge of the meaning of ‘rational’ consists.”⁴ It is as though the explanation we have been looking for is really the wrong way around. We expressivists have been asked to examine our normative concepts and show why they behave according to the same logical rules as non-normative ones. Minimalist solutions begin instead with the surface, and claim that we learn our concepts by learning the rules governing the grammatical, syntactic logic. It is therefore no surprise that declarative normative sentences obey truth-conditional logical rules, because *all* sentences embeddable in the construction ‘it is true that ...’ do.

2.1 Theft over Toil: ‘Bob is Hiyo’

The trouble with these solutions was clearly demonstrated by Dreier (1996). Dreier invites us to imagine developing a custom whereby instead of greeting Bob by saying ‘hi Bob!’ we say ‘Bob is hiyo!’ The two sentences have the exact same meaning and linguistic function. Wherever it would be appropriate to say ‘hi Bob’ it is also appropriate to say ‘Bob is hiyo’ and vice versa. The only differences between the two sentences are superficial - their sounds and their surface grammar. This change does allow for the new sentence to fit (grammatically, at least) into the same embeddings as other

⁴ Horwich (2004: 145).

declaratives: ‘it is false that Bob is hiyo,’ ‘I am unsure whether Bob is hiyo,’ ‘either Bob is hiyo or I shall eat my hat,’ and so on. Dreier’s point is that we have no idea what these complex sentences mean, and if we do not know, minimalist solutions won’t be any remedy. Here’s how the explanations might go.

‘It is true that Bob is hiyo’ means the same thing as ‘Bob is hiyo,’ and therefore the same as ‘hi Bob.’ Maybe we can get on board with this. When I see Bob and wish to greet him, perhaps I can say ‘it is true that Bob is hiyo’ and all will be well.⁵ What about negations? ‘It is false that Bob is hiyo’ means that Bob is not hiyo... and already we’re lost, because it is unclear when we would use this sentence, or what it would mean to “disagree” with the “claim” that Bob is hiyo. And disjunctions? ‘Either Bob is hiyo or I shall eat my hat’ is true just in case ‘Bob is hiyo’ is true or ‘I shall eat my hat’ is true, and again this is no help whatsoever. Even if we grant that we understand what ‘Bob is hiyo’ means when it stands alone, our understanding does not automatically carry over to sentences in which it is embedded merely in virtue of its syntax.

One plausible reason why embeddings don’t work is that ‘Bob is hiyo’ is not a *judgement* of anything. One says ‘hello’ to people not to express a judgement but to greet them. If it were a judgement, we might have an idea of what it would be like to negate the claim, and disjoin it with others, and how inferences might be legitimate between various constructions, but “[i]t is obvious that the idea of inferring is out of place when the conclusion is a speech act of accosting.”⁶ But this is *exactly* what someone might say in opposition to expressivism: people use the sentence ‘stealing is wrong’ to express a bona-fide judgement, whereas booing is merely a speech act with the function of expressing disapproval – the idea of inferring is out of place where desire-like attitudes are concerned (or so the complaint might go).

Of course, the expressivist thinks disapproval (or whatever the relevant attitude is) *is* a judgement, but they cannot assume it outright. One who is sceptical of expressivism may view the expressivist interpretation of ‘stealing is wrong’ as sceptically as we view ‘Bob is hiyo,’ and so giving a minimalist explication of the meaning of embedded sentences is going to be

⁵ It is unclear how to make sense of ascriptions of truth to other people’s utterances of ‘Bob is hiyo.’ Maybe I would only say Jane’s utterance of ‘Bob is hiyo’ is true if I was intending to greet Bob. However an intention to greet can clearly come apart from an actual greeting, so this is very hazy (which is the point Dreier is making).

⁶ Dreier (1996: 43).

no more illuminating than it was in the case of ‘either Bob is hiyo or I shall eat my hat.’ What *would* be illuminating? Well, the expressivist thinks the key to understanding morality lies in characterising the states of mind we express. What we need, then, is a substantial account of the states of mind that constitute accepting negations, disjunctions and other embeddings. All the following solutions involve attempts to do so.

3. Higher Order Attitude Solutions

Blackburn (1984) initially posited higher order attitudes to explain inconsistency and inference. “Higher order attitudes” in this sense are attitudes taken *towards* other attitudes or combinations of other attitudes (as opposed to the “higher-order states” we shall discuss in §6 which are mental states multiply realised by compositions of more familiar mental states). I feel good about loving progressive metal, and I believe that my fear of death and interest in cryonics are related. These are higher order attitudes. The proposed solution to the Frege-Geach problem posits higher order attitudes as the mental states expressed by logically complex sentences. Perhaps the conditional ‘if lying is wrong, telling your little brother to lie is wrong’ expresses a higher order attitude of approval of: upon disapproving of lying, forming disapproval of telling your little brother to lie. Alternatively, as Hale (1993) prefers, it might express disapproval of: disapproving of lying while simultaneously not disapproving of telling your little brother to lie. Accepting the conditional along with the antecedent ‘lying is wrong’ exerts pressure (to conclude it is wrong to tell your little brother to lie) in the following way.

Anyone holding this pair must hold the consequential disapproval: he is committed to disapproving of getting little brother to lie, for if he does not his attitudes clash. He has a fractured sensibility which cannot itself be an object of approval. (Blackburn 1984: 195.)

The conclusion of the argument is supposed to follow from the premises because, in virtue of accepting them, we would disapprove of ourselves if we failed to disapprove of telling your little brother to lie, and that’s not a very nice state to be in. A problem with this, which Hale raises, is that this doesn’t seem to be a logical failing but some other kind of failing, perhaps a moral or prudential failing or a kind of practical incoherence exemplified by Moorean sentences like ‘it is raining and I do not believe it is raining.’

Someone sincerely asserting a Moorean sentence would be guilty of some pragmatic incoherence, but the sentence itself is not contradictory. This shows that merely being guilty of a “fractured” mind is insufficient to have committed a logical mistake. The point was nicely elucidated by van Roojen (1996) who looked at how validity might erroneously be attributed to invalid arguments. Consider the following inference from (1) to (2).

- (1) It is wrong for me to believe my daughter will fail.
- (2) My daughter will not fail.

This is invalid. (1) may be true and (2) may be false simultaneously. Still, to accept (1) and to accept the negation of (2) (which I’ll call ‘ $\neg(2)$ ’ hereafter) would mean I disapprove of my own judgement, because to accept (1) is to disapprove of accepting $\neg(2)$. Once again this shows that the mere fact that one disapproves of one’s own mind is not enough to ensure one is being logically inconsistent, because it manifestly is not logically inconsistent to accept (1) and $\neg(2)$. Perhaps due to these sorts of problems, Blackburn moved away from this approach and towards characterising logically complex judgements as dispositions to infer.

Might he have been too quick? Derek Baker and Jack Woods (2015) argue that fractured sensibilities – or “discordances” in attitude, as they put it – have the potential to explain why we take particular sets of sentences to be logically inconsistent, without this having to imply that *every* possible instance of a set of sentences expressing discordant attitudes is a set we take to be logically inconsistent. Their paper is extremely rich and I will have to omit the details, but here it is in a nutshell. They first note that discordance in attitude is familiar. Someone with intransitive preferences, or who likes and dislikes the same object in the same respect, or who puts a credence of 0.9 in a proposition and a credence of 0.4 in its negation, would seem confused and fractured to us. Second, they argue that the correct way to distinguish between logical inconsistency and merely practical incoherence or other forms of irrationality is via the orthodox method: the syntax of the language. After all, this is how we distinguish between the necessarily false ‘water is not H₂O’ and the logical contradiction ‘water is not water.’ We look at the *logical vocabulary* involved and ignore the specific interpretations of the other terms: ‘ x is not y ’ could be true for many interpretations of x and y , whereas ‘ x is not x ’ could not be true for any x . Baker and Woods argue that expressivists are entitled to give a parallel story. Some sentences express discordant attitudes partly in virtue of what the non-logical terms

mean, perhaps such as (1) and $\neg(2)$. These are not the logically inconsistent sentences. "Some sentences are such that their assertion expresses discordant attitudes simply in virtue of their logical vocabulary and other structural features. These are the logically inconsistent sentences."¹

If Baker and Woods are right, we might be in a position to rescue Blackburn's early approach. The logical vocabulary and sentence structure of (1) and $\neg(2)$ do not guarantee that the sentences express discordant attitudes. The consistent set [(1a) It is easy for me to believe my daughter will fail; $\neg(2a)$ My daughter will fail] has the same structure and logical vocabulary but doesn't necessarily express discordant attitudes, and the same goes if we change the pronouns in (1) and $\neg(2)$ from first to second person. Therefore, Blackburn does not have to say that the inference from (1) to (2) is valid. The higher order attitude solution needn't misdiagnose invalid arguments.

As it happens, I *do* think Baker and Woods are right. An approach to explaining inconsistency and to solving the Frege-Geach problem that relies on discordance of attitudes seems viable in principle. Yet God is in the detail. While Blackburn doesn't have to use the word 'validity' he does seem to have to admit that accepting (1) and $\neg(2)$ is discordant. You might think that's not too bad. There is something a little bit jarring about a person who thinks it's wrong to ϕ and proceeds to ϕ . Is this really *discordant*, though? Is such an agent irrational or incoherent in a manner similar to having intransitive preferences or believing a Moorean sentence? I find it easy to imagine rational people who purposefully do (or believe) what they think is morally wrong to do (or believe). There are all sorts of reasons people might commit what they take to be wrongdoing.

Perhaps we should not interpret the kind of higher order disapproval as *moral*, then. It does seem strange to *morally* disapprove of Percy who believes the antecedent and denies the consequent of a conditional he accepts, especially if it's a particularly benign one like 'if Julius Caesar was tall, he wore boots.' Percy might be lovely. How about a grand *all things*

¹ Baker and Woods (2015: 423). Ridge (2014: 156) makes use of the distinction between logical and non-logical terms in a very similar way: "An argument is valid just in case any possible believer who accepts all of the premises but at one and the same time denies the conclusion would thereby be guaranteed to have inconsistent beliefs, *where this remains true on any acceptable substitution of the non-logical terms of the argument.*" If 'inconsistent beliefs' is read minimally enough to include all discordant attitudes instead of restricting us to robust beliefs with inconsistent contents (*contra* Ridge!), it's the same view.

considered disapproval? This won't do. Once again there are all sorts of possible reasons Percy might have to believe the antecedent and deny the consequent of a conditional he accepts. (We can always invoke some evil demons.) Instead, Blackburn might plump for a particular kind of *rational* disapproval. While 'stealing is wrong' expresses disapproval_{moral} of stealing, 'if stealing is wrong then taxation is wrong' expresses disapproval_{rational} of disapproving_{moral} of stealing without disapproving_{moral} of taxation.² This makes a bit more sense but not enough. Sally, who accepts the conditional, need not think it is irrational for Chen, who does not accept the conditional, to fail to infer 'taxation is wrong' from 'stealing is wrong.' There is only rational pressure to make the inference *for those who accept the conditional*. Could we make room for some self-reflexivity in the mental state? Let us try the following. All conditionals 'if p then q ' express:

State C: disapproval_{rational} of simultaneously accepting p , denying q and being in *State C*.

Notice that I have decided to describe *denying* the consequent rather than failing to believe the consequent. This involves understanding negation. Let us apply the same technique. Negations ' $\neg p$ ' express:

State N: disapproval_{rational} of simultaneously accepting p and being in *State N*.

Likewise, disjunctions ' p or q ' express:

State D: disapproval_{rational} of simultaneously denying p , denying q and being in *State D*.

Once we secure these we're up and running. The finish line will still be a long way away but at least we'll be running. I'm unsure how comfortable readers will be with reflexivity in mental states, but I quite enjoy these kinds of things. Besides, it is not without precedent. Just as 'erase this sentence' can be a legitimate instruction written on a chalkboard, Luca Barlassina and Max Hayward (manuscript) argue that pain is a mental state M that

² We could push Blackburn further here, since it needn't be irrational to fail to infer the consequent from the antecedent of a conditional one accepts. What is irrational is *denying* the consequent while believing the antecedent of a conditional one accepts. This requires us to answer the negation problem: what is it to deny that taxation is wrong? Blackburn (1988) postulated a separate attitude of *tolerance* as the attitude expressed, but this forces on us a nasty choice: either 'grass is not green' implausibly expresses some kind of tolerance, or the word 'not' is ambiguous despite there being no linguistic markers of ambiguity. Due to its unpopularity I will not discuss this view further.

motivates the agent to eliminate that very state M .³ (We might say that *State M* is disapproval_{pain} of being in *State M*!) Their proposal is controversial of course, but the idea that functional characterisations of mental states may refer to themselves is not outlandish.

A consequence of this approach is that *all* inconsistency, even between ordinary descriptive sentences, is explained in terms of attitudinal discordance. The fundamental explanation of why ‘grass is green and grass is not green’ is logically inconsistent is because, given only its logical vocabulary and structure, it necessarily expresses a fractured sensibility, namely *State N* and the acceptance of p . While some might see this as a *reductio* I do not. But the finessed higher order analysis I’ve given on behalf of Blackburn is incomplete. While it *might* be the case that the formal structure of the solution will satisfy the Frege-Geach requirements, this doesn’t show that *we* have such mental states as C , N and D . What we want is a credible gloss of disapproval_{rational}. It seems more plausible to me that these states are better understood as dispositions than the kinds of affective feelings that come to mind with normal uses of the word ‘disapproval’. The idea would be that to rationally disapprove of a combination of mental states is to be disposed not to be in those states. If one is in *State D*, for example, and also denies p and denies q , one will have the tendency to stop denying p or stop denying q or eliminate *State D* by no longer having the relevant disposition. However, this is now starting to sound *exactly* like the inferentialist solutions to the Frege-Geach problem. Rather than explicate the current proposal any further I will jump right into these inferentialist solutions. The problems they face will transfer to this one.

4. Inferentialist Solutions

Treatments of the Frege-Geach problem given by the later Blackburn (1988), Gibbard (1990, 2003) and Terence Horgan and Mark Timmons (2006) can be viewed in a similar manner to each other. They all present an abstract,

³ Does pain not motivate us to prevent whatever is damaging our bodies? No. Instead of removing the bad tooth, we might take painkillers. The existence of pain can come apart from bodily damage, as with painful phantom limbs or painless organ failure. If Barlassina and Hayward are right, pain is a kind of “suicidal” state telling us destroy it (or lessen it), which only *contingently* might be best achieved by treating the bodily harm that generated it. On the happier side of things, Barlassina and Hayward characterise pleasure as a mental state that tells us to keep it around or intensify it.

formal structure, in terms of Hintikka sets or sets of fully opinionated “hyperplanners” for example, and use it to model moral thought. The structure effectively gives us a recipe for finding out whether any two sentences in the language are inconsistent. The contents of complex sentences are explained in terms of the relevant bits of the structure.

Let us consider Gibbard’s (2003: ch.3) model. Gibbard thinks normative judgements are plans. In the previous chapters in his book he introduced hyperplanners – idealised creatures who, for every proposition p and for every possible action ϕ , either believe p or believe $\neg p$ and either plan to ϕ or plan not to ϕ . They believe no contradictions and do not intend the impossible. My judgement that grass is green is represented as the set of all the hyperplanners, or the mental “hyperstates” of the hyperplanners, I do not disagree with (in virtue of this particular judgement anyway) – it rules out all those hyperplanners who think grass is not green. My judgement that I ought to dance is likewise represented as the set of all the hyperplanners I do not disagree with in virtue of that plan – it rules out all those hyperplanners who plan not to dance.¹ Set-theoretic operations for logical operators are then defined in the same way as in possible worlds semantics. The content of the judgement that either grass is green or I ought to dance is the union of the set of hyperplanners who don’t believe grass is not green and the set of those who don’t plan to not dance. How does this help with inference? Just as we should always believe the actual world is in a non-empty set of possible worlds, Gibbard maintains we should always keep open the possibility of agreeing with a hyperplanner. *Don’t rule out everything*. So, if I accept the disjunction just mentioned, and if I also believe grass is not green (which is to rule out all the set of hyperplanners who believe grass is green), then it would be inconsistent to think I ought not to dance – I will have ruled out everything. Michael Ridge is not impressed with this general strategy.

However, this account does not tell us *anything* about the state *in virtue of which* one disagrees with the relevant set of hyperplanners. That, though, was pretty much the original challenge. Given Gibbard’s stated aim of explaining meanings in terms of states of mind expressed, what we need is a systematic explanation of what the relevant states of mind could *be*, such that the right pattern of disagreement is secured. All we really get from Gibbard are criteria

¹ Whether divergent plans constitute disagreements is an issue I take up in chapter 4.

for a satisfactory ideationalist/expressivist theory, and not the hard work of spelling out such a theory. (Ridge 2014: 139.)

Ridge then quotes the following passage from Schroeder, who had made the same point.

What it is really saying is merely that ‘murder is wrong’ must express a mental state that is inconsistent with all and only the hyperdecided mental states that ‘murder is not 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 states. (Schroeder 2008: 52-3.)

The complaint is that all Gibbard (and Horgan and Timmons etc.) have provided is a model, and models come cheap. They need to characterise our mental states when we accept the bits of the structure corresponding to complex sentences. Gibbard and company may respond thus: the formal model *is* the characterisation. If we adopt inferentialism about meaning, a list of legitimate inferences is all we are expected to provide. In this case we are only concerned with the meaning of logical connectives, whether or not we adopt it for other terms as well. “An inferentialist theory of meaning holds that the meaning of a logical operator can be captured by suitably formulated rules of inference (in, say, a system of natural deduction).”² If inferentialism is true, then giving a list of what is to be (and not to be) inferred from what is a fully exhaustive and comprehensive account. Let us consider a conditional. According to Blackburn, “[o]ur firm assent to ‘Whenever *p* is true and *if p then q* is true, *q* is true’ corresponds to an inferential disposition.”³ On the current proposal, this is all there is to accepting conditionals. If you are disposed to make the relevant inference, you accept the conditional in virtue of that; not the other way around. Being disposed to infer that abortion is wrong from foetuses feel pain (i.e. being disposed to come to disapprove of abortion upon believing foetuses feel pain) *constitutes* accepting ‘if foetuses feel pain then abortion is wrong.’

² Tennant (2007: 1056).

³ Blackburn (2010: 36). The same idea is in Hare (1970: 19): ‘we can easily perform the standard maneuver for letting the consequent of the hypothetical out of its cage. Thus, if I am prepared to say that it is a good movie, and that if it is a good movie it will make a lot of money, I can go on to say “It will make a lot of money.”’

Throughout various pieces of expressivist literature, the assent to inferentialism is clear:

The proposal, as I understand it, is that we can characterize the sense of compounds whose components may be evaluative, not in terms of their truth conditions, but in terms of the inferential commitment involved in endorsing them. (Hale 1993: 347.)

Here, then, I think Blackburn is right about what is needed, and my treatment of normative directives in the web of judgment proceeds the way he suggests: I give primacy to inferential relations, with a conceptual role picture of logic. ... We explain the meanings of normative directives by the normative governance that accepting them brings. Other statements get their content from their place in a web of implication. (Gibbard 1993: 72-3.)

A sincere utterance of 'if [B!(l)] then [B!(b)]' will be said to indicate that a speaker possesses what may be called an 'inferential disposition' – a mental state such that if the speaker were to adopt the mental attitude associated with the utterance '[B!(l)]', she would be led to adopt the mental attitude associated with '[B!(b)]'. (Price 1994: 139.)

The suspicion just aired is, in effect, that the prospects for success in this case may be arguably no worse than the prospects for an inferential role semantics for the logical connectives. (Lenman 2003: 36.)

To follow up on the final comment by Lenman, we may worry the prospects for success are no *better* than the prospects for inferentialism about logic. As was apparent in §3 of the previous chapter I dislike the idea that expressivism is committed to a controversial theory of meaning. I would prefer not to be tied down if I can help it. Nevertheless, it is not the worst thing to be tied down to. Insofar as expressivism and inferentialism are forms of pragmatism they are happy bedfellows. Furthermore, inferentialism about logical connectives is independently plausible. If you know the rules for disjunction introduction and elimination, what more do you need to know to understand the inclusive 'or'? The rules of inference appear sufficient. So, while I am officially non-committal on this issue, I am prepared to grant that adopting inferentialism is a bearable cost.

Inferentialist solutions to the Frege-Geach problem involve, in the first instance, a formal model that maps inferences. The meanings of logically complex sentences are explained in terms of the positions they occupy in

the inferential web. To accept a logically complex sentence is to “tie oneself to a tree”⁴ such that one is *committed* to making the relevant inferences, whether or not one ever does.⁵ ‘If lying is wrong, getting little brother to lie is wrong’ therefore expresses the inferential commitment (might we call it ‘approval_{rational}’?) to come, upon disapproving of lying, to also disapprove of getting little brother to lie.

4.1 ‘Bob is Hiyo’ Revisited

If I told you that the conditional ‘if Bob is hiyo, cats are greasy’ expresses the disposition to infer cats are greasy from Bob is hiyo, would you be satisfied you understood the conditional? I imagine not. So, if inferentialism fails to solve the ‘Bob is hiyo’ Frege-Geach problem, it can’t solve the expressivist one either. But perhaps this is unfair. Expressivists at least give characterisations of the atomic mental states involved. Since there is no obvious state to assign as the “assertability condition” of ‘Bob is hiyo,’ the two cases aren’t analogous. So let’s invent another one. Headaches are a mental state. (If you don’t like this example, substitute some other state. Maybe love, or ecstasy, or the qualia of seeing strawberry fields would do the trick.) If you have a headache it is linguistically correct for you to say ‘head is achey-O.’ This sentence is distinct from ‘I have a headache’ which is linguistically correct to say when you *believe* you have a headache.⁶ I’m sure you see where this is going: ‘if head is achey-O, cats are greasy’ expresses the disposition to believe cats are greasy upon getting a headache. A strange disposition to be sure, but we could discuss a more common one. ‘If head is achey-O, I have a headache’ expresses a very normal disposition,

⁴ The phrase comes from Blackburn (1988).

⁵ The notion of ‘commitment’ in the expressivist literature seems to be dispositional, so we could specify whether an agent accepts a logically complex sentence in purely descriptive terms. However, it is open to the expressivist to interpret ‘commitment’ normatively. The idea would be that someone might accept ‘if *p* then *q*’ without being disposed to make the inference – all that needs to be the case is that they *ought* to be so disposed, or are perhaps *responsible* for being so, or something along these lines. (Brandom 1994, 2000 defends a normative conception of inferential role.) Given expressivism about ‘oughts’ we are going to get into the kind of circularity familiar from §1.1 of chapter 1. I do not think much hangs on this, so I will not discuss the normative option further.

⁶ Recall from §1.3.2 of the previous chapter that conditions of linguistic correctness are not truth conditions. It is linguistically correct for me to say ‘I have a headache’ when I believe I have a headache, but it may still be false that I have a headache. (Perhaps I have been confused by a pain somewhere else in my body.)

namely the disposition to believe one has a headache when having a headache. Yet I have no idea what this sentence means.

With Dreier's 'Bob is hiyo' sentence we learned that a sentence being in declarative form does not guarantee that logical constructions involving it will be meaningful. Likewise, say I, with underlying mental dispositions or commitments. The mere fact that I understand – or indeed have – the disposition to move from one mental state M to another M^* does not guarantee I'll understand a sentence 'if p then q ' where p expresses M and q expresses M^* . In the end I think it boils down to the issue of justifying the relations of consistency and inconsistency which are at the heart of logic. We can have inconsistent beliefs but we cannot have inconsistent headaches or Bob greetings. We need to earn the right to the use of logical connectives by independently showing *why* the relevant mental states are the right sort to ground inconsistencies, inferences, and so on. Why believe the desire-like states proposed by expressivists to be moral judgements are of this sort? The reason both Blackburn and Gibbard allude to involves realisability.

4.2 The Anchor of Realisability

Plans, intentions, commitments, aims, goals – these are the sorts of mental states Blackburn and Gibbard cite as the relevant attitudes involved in making moral judgements. These kinds of attitudes are very plausibly subject to constraints of consistency, and this is true regardless of inferentialism about logical connectives. Consider a trekker who plans to fight the bear and not fight the bear, or a confused revolutionary who aims to eliminate all forms of government as well as to set up a computerised government. These are examples of agents with internally inconsistent goals.

Can we use this to anchor our logic? Blackburn thought so. In his (1988) paper he grounded deontic logic with the concept of a (theoretically) realisable morally ideal world where every obligation is fulfilled. Gibbard followed suit. 'At this point in the treatment, then, I too would use a notion something like Blackburn's one of "consistently realizable attitudes".'⁷ In Gibbard's earlier (1990) terms, the norms we accept must be able to govern us, and in his later (2003) terms, the plans we make must be possible to carry

⁷ Gibbard (1993: 71).

out. There is a clear parity here with robust beliefs.⁸ Constraints of consistency guide us in both arenas: don't believe or intend the impossible. This idea may seem enough to get us going – it certainly sets the relevant mental states apart from headaches – although even if it is on the right track there are obstacles, such as accounting for mere permissions.⁹ Unfortunately it is beset by problems right from the starting block.

4.3 Implausible Ascriptions of Irrationality

Grounding the inconsistency of normative judgements via their incompatible commitments has counterintuitive consequences for various seemingly consistent and rational moral judgements. I'll discuss three cases. These involve people with agent-relative values, who believe in moral dilemmas, and who doubt that 'ought' implies 'can'.

Dotty's daughter and Sonny's son are dying of poison. The only available dose of antidote is within reach. Here's a claim someone might sensibly make: the right action for Dotty is to grab it and save her daughter, and the right action for Sonny is to grab it and save his son. These are inconsistent aims – only one of them can be satisfied – but the judgement itself doesn't seem inconsistent or irrational, whether or not we agree with it. There are two ways pure expressivists might object. First, perhaps the more natural judgement isn't that both Dotty and Sonny ought to grab the antidote but that they both ought to *try* to do so. Everyone should *try* to save their child; succeeding is a bonus. In reply, this seems *ad hoc*. When someone asks about the right thing, the usual answer is in terms of what to *actually do* rather than merely try to do. Here is a second objection pure expressivists might put forward: we're not saying it is right that Dotty and Sonny both save their child. That's impossible. What we're saying is, when it comes to Dotty, the right thing is to save her daughter. That's possible.

⁸ Stevenson's (1944: ch.1) views on disagreement are similar. Roughly, two people disagree if their beliefs are incompatible or if their desires are. If we understand inconsistency as disagreeing with oneself, the views are the same. The possibility of grounding logic in inconsistent plans is reminiscent of Hare's "logic of imperatives" (1967) and Dummett's (1981: 327-54) "obedience" conditions for commands (but see Hurley (1989: 180-85) for a convincing objection to Dummett).

⁹ For Gibbard (2003: 57), "hyperplans leave nothing indeterminate as to what to do," but this is a bad model for moral views which allow for several equally permissible options at once. See Dreier (2009: 95-7) for discussion. For a promising solution using the decision-theoretic notion of conditional weak preference, see Silk (2015).

And for Sonny, the right thing is to save his son. *That's* possible. What this amounts to is planning (say) to save your child, no matter who you are. Whoever you turn out to be, your plan *can* be carried out (at the expense of the other). While it is guaranteed to be the case that a non-specified *someone* fails to do the right thing, it is not the case that someone *in particular* is guaranteed to fail. Thus, in no situation would we be intending to do the impossible.

This brings us to the next problem. If the second objection works it is only because the guaranteed wrongdoing is “split” between two agents such that each agent may possibly do right. Yet some people believe in the existence of genuine moral dilemmas. Suppose you promised Isha you'd lie to her girlfriend about where she was last night. Here are your exhaustive options: keep your promise and lie, or don't lie and thereby break your promise. Some people may think that in this situation there is no way to act morally. Lying is wrong and breaking promises is wrong. (We can put it in terms of rightness or obligation too: you are morally required to keep your promises and to only tell the truth.) Doing what's right and avoiding what's wrong may not always be possible, or so it is often judged. This judgement may be incorrect, but the point is that it is surely not *inconsistent* despite being “unsatisfiable” and ruling every option out.¹⁰ Given the discussion in §3 leading from Baker and Woods' (2015) paper, I admit that the pure expressivist doesn't have to say that the defender of moral dilemmas is *logically* inconsistent. However, they do have to say the dilemma defender is discordant, or incoherent, or fractured, or irrational in the kind of way that a believer of a Moorean sentence is. This is something a realist does not have to say. On the realist's view, one can robustly believe all available options have the property of moral wrongness without any form of incoherence whatsoever. Since it is plausible that one *can* coherently believe in moral dilemmas, this counts against the pure expressivist.

The cases we've been discussing stem from the moral principle of 'ought' implies 'can'. If moral judgements are plans, goals, or other states that are inconsistent when they're unrealisable, then we are committed, on pain of incoherence, to the principle of 'ought' implies 'can'. But there are many people who deny this principle and they don't seem incoherent.

¹⁰ Van Roojen (1996: 324-25) makes a similar point about moral dilemmas and consistency. It is mentioned in a single paragraph almost in passing. Immediately afterwards he writes “I don't want to make too much of this,” but I do; it is a strong objection and merits further analysis. Thanks to Jimmy Lenman for this reference.

Those who believe in moral dilemmas are one such group, as they can be interpreted as thinking that one ought to avoid wrongdoing and that sometimes wrongdoing is impossible to avoid. There are many other ways to argue against the principle that 'ought' implies 'can'. Peter Graham (2011) considers a surgeon completely psychologically compelled to cut up two healthy patients to save ten others and argues the surgeon is doing something morally wrong. This is because, first, it would be permissible for the janitor to shoot the surgeon, and second, because every adequate explanation of *why* it would be okay cites the prevention of wrongdoing. Whether or not we agree with Graham's argument (I don't), the point is that he seems coherent and not in any way like the believer of a Moorean sentence. Once again, pure expressivists implausibly ascribe irrationality.

I should mention that biting the bullet is an option. Perhaps all this should be seen as a first-order discovery. Once we see what moral thinking is, we realise we are committed to thinking 'ought' implies 'can'. We therefore have a transcendental argument in favour of the principle and against the existence of moral dilemmas. Now, I am not against making use of transcendental arguments for first-order claims. I do so myself in chapters 3 and 5. But here it comes at a higher cost: while the commitments I bring out on behalf of the expressivist do not fly in the face of ordinary folk practice, judgements about obligations and abilities, as well as about moral dilemmas, *are* part of ordinary folk practice, and there is evidence that a substantial portion of the folk deny 'ought' implies 'can' and believe in moral dilemmas.¹¹ To write their views off as incoherent is to undermine the Wittgensteinian conviction (that folk practice is by and large fine as it is, and if we interpret it as irrational then our interpretation is likely to be wrong) that drives many to expressivism in the first place. We would do better to continue our search.

5. Structural Solutions

Here is a possible diagnosis of the faults in the previous section. Pure expressivists, especially Gibbard (2003), are interested in all practical reason and not just moral reason. They are interested in the *all things considered* 'ought' in the first instance and only derivatively the moral 'ought'. Moral considerations are only part of what goes into ordinary decision making. It

¹¹ Buckwalter and Turri (2015).

might be the case that moral considerations count against all the available options, but we still may decide which option we ought to choose on the basis of other reasons such as self-interest.¹ Unlike moral dilemmas it is far less clear whether there is any support for *all things considered* dilemmas. Perhaps we can imagine thinking to ourselves “Given that both options are morally wrong, I should probably tell Isha’s girlfriend the truth.” Similarly, while it is not unreasonable to believe we have moral obligations we are unable to fulfil, the idea that one ought, all things considered, to perform an impossible action is much more implausible. Perhaps, then, expressivists are only committed to there being no *grand practically normative* dilemmas. Maybe it *is* irrational to take all options off the table when it comes to your final planning decisions. For all I say in this chapter, pure expressivism about the all things considered ‘ought’ is open. But I’m interested in morality, and in this section I argue against one final attempt for pure expressivists to solve the Frege-Geach problem in the moral domain.

Recall the negation problem discussed all the way back in §1.1. There we saw that a one-place disapproval, e.g. disapproval of stealing, did not have enough structure to account for negation. The solution explicit in Schroeder’s (2008) book is to add more structure, and there are allusions to this idea in other expressivist writings. The basic non-cognitive attitude for Schroeder’s purposes is “being for” something, which is a general pro-attitude, and moral sentences allow for two distinct places: “to borrow a proposal from Gibbard (1990), we might say that ‘wrong’ corresponds to *being for blaming for*, so that ‘murder is wrong’ expresses FOR(*blaming for murder*).”² ‘Murder is not wrong’ therefore expresses being for *not blaming for murder*, which is distinct from not being for anything to do with murder. Thus, the negation problem – distinguishing positive judgements of permissibility from lack of any judgement at all – is tentatively solved.

¹ I must admit I don’t fully understand the *all things considered* ‘ought’. I can make sense of oughts in different domains – I prudentially ought to steal the bread, I legally ought not to, and morally it’s permissible either way – but I cannot see how to *combine* them. In my eyes they appear incommensurable. I find it difficult to make sense of questions such as “Yes, I *morally* ought to do this, but should I *really* do it?” without indexing the ‘should’ to another domain. Copp (1997), Tiffany (2007), Baker (2018) and others join me in my scepticism about an ought *simpliciter*. Yet, since I grant that I do make plans based on considerations in many domains, perhaps I should allow the creation of a term which expresses those plans. I don’t catch myself ever using such a concept but there’s nothing stopping me learning what Gibbard and others were apparently taught growing up.

² Schroeder (2008: 58).

Schroeder explicitly appropriates the idea of linking judgements of wrongdoing to blame from Gibbard (1990). Gibbard (1990: 47) also refers to other reactive attitudes: “to think an act *morally reprehensible* is to accept norms that prescribe, for such a situation, guilt on the part of the agent and resentment on the part of others.” To judge an act to be *not* morally reprehensible is to accept norms that *don't* prescribe those feelings. The extra structure is also present in Blackburn (1995: 244-45) who says the judgement that being uncharitable is not wrong is a “positive disposition” “not to bring pressure on persons to make them charitable, or to be angry with those who are not, or shun their company, or preach against them or encourage people not to be like them” as well as the later Gibbard (2012: 242): “To believe an act wrong is to plan, in case one contemplates doing it, to feel obligated not to do it.” We also see enough structure for negation in Lenman (2013: 400), who writes that ‘torture is wrong’ expresses “unwillingness to accept any candidate sets of rules for my community that permit torture” – we could then take ‘torture is not wrong’ to express unwillingness to accept community rules that do *not* permit torture.

How does all this help with the troubles raised in the previous section? Consider the person who believes you were in a genuine moral dilemma regarding Isha’s girlfriend. They can now be interpreted coherently. Sticking with Schroeder’s view, the dilemma defender is both *for* blaming for lying to Isha’s girlfriend and *for* blaming for breaking your promise to Isha. These are realisable ‘*for*’ attitudes: it’s perfectly possible to blame you no matter what you do. Thus we can now draw upon the anchor of unrealisable attitudes to ground our logic without having to believe, on pain of incoherence, that the moral ‘ought’ implies ‘can’.

5.1 The Wrong Attitude

Whether or not this idea can solve the full Frege-Geach problem and capture the full extent of our linguistic expressions beyond simple logical connectives (Schroeder thinks it cannot), my first issue is with the moral attitude itself. It is implausible that being in favour of blaming for something is judging it wrong. The two can pretty clearly come apart, especially for anyone who thinks consequences are sometimes morally important. I might believe Jess did the right thing by torturing the terrorist to prevent a mosque being blown up, and yet be in favour of blaming her for that very action in order to thwart the propagation of a more casual

attitude towards torture. I might believe eating meat is wrong without being for blaming for it because I think blame is unlikely to do any good. (Ironically, while many of the folk are for moral dilemmas and thus against the moral ‘ought implies can’ principle, they are very unlikely to want to *blame* someone for doing wrong in a moral dilemma because of course it couldn’t be avoided: “unlike attributions of obligation, attributions of blame are informed by considerations of ability, which lends support to the related principle that blame implies can.”³)

The problem generalises to the other proposals. To judge something wrong is not to plan to feel guilty about it. Indeed, some people think guilt is not necessary for morality at all and may even hinder it.⁴ Likewise, it seems possible to be unwilling to accept community rules that permit, say, taxation of earnings above £100 000 at 80% without believing said taxation to be *morally* wrong. Similar things can be said about shunning the company of wrongdoers or preaching against them. In short, for any substantive characterisation of the moral attitude in the first of the two structural places, we seem to be able to imagine someone with that attitude who does not make the relevant moral judgement and vice versa.

What we have ended up with is a kind of open question argument. Consider the following question: Angela has mental state *M*, but does she make moral judgement *m*? Whenever ‘*M*’ contains no moral vocabulary this question will seem open. Cognitivists who believe in irreducibly normative concepts – robust and relaxed realists and the Cornell natural realists, for example – have an easier time: Angela has the robust belief that stealing is wrong, but does she believe that stealing is wrong? Quite plainly yes. However, expressivists have two lines of reply. First, even if realism is true there will be another characterisation of Angela (plus elements of the world, if mental externalism is true) that uses no moral vocabulary and yet guarantees she will be in the same mental state, such as a purely neurological or biological or physical description, so the realists will be in the same boat further down the line. Second, expressivism need not be thought of as a purely conceptual truth. It is informed by empirical evidence, either in the form of experiments in moral psychology or by our own *a posteriori* engagement with moral practice in our society.

³ Buckwalter and Turri (2015: 16).

⁴ Harman (2009).

Nevertheless, this “mental state open question argument” is a way to test proposals for the moral attitude. If it seems not just conceptually open but, due to compelling counterexamples, outright *false* that Angela makes judgement *m* when and only when she has *M*, then this is evidence that we have identified the wrong attitude. This is what has happened for the given pure expressivist proposals. But what if we take our cue from the non-reductive cognitivists and posit an irreducible moral concept that is desire-like in nature rather than robustly representational? Since I favour a kind of affective, emotional feeling, I’m going to call the relevant attitudes ‘moral approval’ and ‘moral disapproval.’ They are somewhat analogous to love and hate. Importantly, they cannot be easily *reduced* to terms that don’t involve moral ones. We can point towards some tendencies but they won’t be constitutive. If I morally disapprove of some action or character trait I might tend to be outraged when people do it or exemplify it, but this tendency is by no means a strong one. The same is true in other areas. How can I explain why I aesthetically love certain music and hate others?⁵ I can only gesture towards some dispositions but at the end of the day, I just *like* some music in a certain way, and if you don’t get it I can’t communicate it. Perhaps one day with more science we’ll understand the underlying features better. In the meantime let us continue with the idea that the moral attitudes cannot be reduced – not from the armchair, anyway.

The proposal is that the relevant moral attitudes are not about blaming for one action or feeling guilty for not doing another, but are kinds of approval and disapproval with their own moral character that’s pretty hard to describe. I know it when I feel it – I can tell when I morally approve of something rather than like it in some other respect – and I can only hope you know it too, lest you mistake the lovable rogue for the saint. Now, since this section is about the *structural* solutions to the Frege-Geach problem, what we have at bottom is a pro-attitude *for* the relevant moral attitudes towards the action or object to be evaluated. So, on this picture, ‘stealing is wrong’ expresses being for morally disapproving of stealing, ‘stealing is not wrong’ expresses being for *not* morally disapproving of stealing, and someone who does not accept either sentence is not *for* either of the two.

⁵ I am not discussing an expressivist theory of aesthetic value here. I qualified the love as aesthetic in order to distinguish it from loving music in other ways, for instance for political reasons. For all I say, the judgement that *Blackwater Park* is aesthetically valuable may be a robust belief. What I’m concerned with are the affective feelings we experience towards music, regardless of their relationship to aesthetic value judgements.

5.2 Mixed Sentences and Pascal's Wager

Structural solutions also help us to understand sentences that combine the moral and the non-normative, but at the cost of becoming global expressivists.⁶ That is, they require us to posit the same conative attitude that is expressed by *all* natural language declarative sentences otherwise mixed sentences raise the Frege-Geach problem all over again. For if 'grass is green' merely expressed a robust belief, what would 'stealing is wrong or grass is green' express? Neither the robust belief nor the non-cognitive attitude is necessary. This is why Schroeder (2008: 89-92) goes global. Moral sentences express being for some moral attitude towards something; non-normative sentences express being for something else. Exactly what this is could vary between theories. Schroeder always puts some *action* in the first place of his structure – in the moral case, 'stealing is wrong' expresses being for *blaming* for stealing – so he posits some act-type we express being for with non-normative sentences: "I say that believing that *p* is being for proceeding as if *p*. ... On my best gloss, to proceed as if *p* is to take *p* as settled in deciding what to do."⁷ Here is an alternative: believing that *p* is being for buying some bets on *p* where the price divided by the stake is greater than 4. Here is another: (robustly) believing that *p* is being for (robustly) representing *p* as actually true.

We can now make some sense of mixed sentences. 'Either abortion is wrong or foetuses feel no pain' expresses being for blaming for abortion or proceeding as if foetuses feel no pain, or alternatively, being for morally disapproving of abortion or robustly representing it as true that foetuses feel no pain. This is meant to be different to either being for one or being for the other. It is a little unclear how to characterise someone who accepts the disjunction without accepting either of the disjuncts. What is it to be in favour of either ϕ -ing or Ψ -ing without being in favour of ϕ -ing or being in favour of Ψ -ing? Perhaps it is to be disposed to favour ϕ -ing after coming to favour not Ψ -ing. This already looks like we're moving into the higher order attitude and inferentialist solutions we looked at earlier, but I argued that perhaps they weren't so bad except for the counterintuitive consequences when it comes to moral dilemmas and such. Since these

⁶ Price (2015) does not see this as a cost but a natural and indeed inevitable extension of Blackburn's "localised" expressivist programmes.

⁷ Schroeder (2008: 93). It quickly becomes apparent that believing that *p* must also be a case of being for not proceeding as if not *p* as well. This is just one of many Ptolemaic epicycles Schroeder ends up having to make in his semantic programme.

consequences can be avoided with the extra structure, things are starting to look more promising.

The *new* trouble is that it comes out with the wrong results whenever people are in favour of a view they do not currently hold. Pascal thinks it's a great idea to believe in God. He plans to go to church, hypnotise himself if he can, spend time with religious people in the hopes of absorbing their beliefs by osmosis, but right now he does not assign a high credence to the existence of God. He only thinks that, given the small positive credence he assigns God's existence combined with the infinite reward of believing in God if he exists, one *ought* to believe in God. So, it seems to me that Pascal is *for* proceeding as if God exists and *for* robustly representing as true that God exists. But intuitively it is not the case that Pascal believes the sentence 'God exists.' Not until he's managed to convince himself.

In the moral case we have a similar situation. There may well be moral truths we ought not to believe, and even if there aren't, some people (such as myself) believe there are. Here's an example. *Sac*: one ought to sacrifice one's child to save the lives of ten others. Millian Millie thinks *Sac* is true because the damage that would be done to your child and yourself is not as great as the harm that would be caused by the deaths of ten other children. Nevertheless she thinks we ought not to believe *Sac* because, first, believing *Sac* may have adverse consequences for our relationships with people and any children we might have, and second, because the potential benefits probably will never manifest – it's pointless to prepare for making the right choice in a horrendous situation when such a situation is so unlikely to ever arise. It seems most natural to me to describe Millie as someone who morally approves of the child sacrifice but is not in favour of so morally approving. This contradicts the structural account on offer. According to the proposal under consideration, Millie, in accepting *Sac*, is actually *for* morally approving of child sacrifice. The bullet could be bitten in principle but I'd rather not have to do so. I am unconvinced that this is a good characterisation of Millie and I am even more unconvinced that Pascal can be captured within this framework.

To sum up, the problem is that the extra structure ends up construing our judgements to be about the wrong things. When I judge that torture is wrong, my attitude is towards *torture* or some aspect of it, and not towards taking up some other attitude towards torture or some action to those who torture. Likewise, when I believe aliens exist, this is a judgement about *aliens*

rather than about how to proceed. My objection to the solution proposed in this section is that adding structure in this way changes the subject.

Let's take stock. Pure expressivists have trouble assigning mental states to logically complex moral sentences. When I defended the higher order attitude solution as extensively as I could, it turned out to behave just like inferentialist solutions. But inferentialist solutions, just like minimalist solutions, fall foul of the 'Bob is hiyo' objection. We cannot help ourselves to truth tables and inferential dispositions. We have to *earn the right* to them by showing – or at least gesturing – why the relevant mental states are of the right sort to make the language of inference and inconsistency appropriate. The anchor of unrealisability turned out to require extra structure in order to avoid ascribing incoherence to some sensible moral judgements, but in doing so, it interprets our judgements to be about the wrong subject matter.

This is not the end of hope for those committed to purity. Another anchor could be argued for that doesn't imply apparently sensible moral judgements are discordant, or perhaps the bullet can simply be bitten. In defence of structure, maybe there is a way to argue Pascal is not *quite* for proceeding as if God exists but is in fact for *pretending* that God exists, or some other distinction can be drawn out such that Pascal is interpreted correctly as a current disbeliever who wants to believe. But we should also keep in mind the other costs along the way. While I admitted that logical inferentialism is not an implausible theory, it is nevertheless controversial. The idea that all inconsistency, whether normative or not, is grounded in discordant mental states is a very controversial idea too. And of course taking expressivism global is something most people, including Blackburn and Gibbard, want to resist. Furthermore, we have not even scratched the surface of the myriad complexities of moral language.

My conclusion, then, is not that purity is *doomed*, but that it faces very difficult problems that have not yet been given satisfactory answers. It is generally agreed that hybrid expressivists have a much easier time. Here's why.

6. Hybrid Solutions

Hybrid expressivists believe moral sentences express hybrid mental states involving both desire-like attitudes and robust beliefs. How this is done varies from theory to theory. Here's one way. Take a standard cognitivist view such as robust realism and then postulate that utterances of moral sentences also express certain desire-like attitudes, such that one may only assert a moral sentence linguistically correctly if one has that desire-like attitude (Boisvert 2008). Perhaps 'stealing is wrong' expresses a robust belief that stealing has the wrongness property as well as disapproval of stealing. Importantly, 'stealing is wrong' has the same content as the robust belief it expresses – this is the defining feature of “ecumenical cognitivism” as opposed to “ecumenical expressivism.”¹ The attitude tacked on serves to solve the motivational problem for those who are attracted to judgement internalism, and potentially the open question argument too.²

The metaethical theory I find most promising is a form of ecumenical expressivism which Teemu Toppinen (2013) calls the “higher-order state view.” Schroeder's (2013) essay outlines and promotes a view called “relational expressivism” which is practically the same, and Ridge in his (2014) book defends a specific and more detailed instance of both higher-order and relational expressivism. Toppinen's theory is best illustrated with an example.

‘There is reason to ϕ ’ expresses the higher-order state of having (1) a desire or desires (of some appropriate kind) for actions in so far as they have some properties K_1, \dots, K_n , and having (2) a belief that ϕ has one or some of those properties. (Toppinen 2013: 254.)

Let us apply this to a couple of individuals, calling the relevant desire ‘moral approval’ since we are restricting ourselves to moral language. Luna morally approves of actions insofar as they accord with God's will and she robustly believes God wills us to donate to charity. Jeffrey only morally approves of doing as you please without actively harming others, and he neither robustly believes he wishes to give to charity nor that failing to do so will actively harm others – in fact he robustly believes the contrary. In virtue of these described lower-order states, Luna is in the higher-order state expressed by ‘there is reason to give to charity’ and therefore she

¹ Ridge (2014: 80), discussed in §4.5.3 of the previous chapter.

² See §5.2 of the next chapter for a brief discussion of how hybrid expressivist forms of natural reductivism might address open question worries.

believes there is reason to give to charity. Jeffrey, on the important stipulation that he *only* morally approves of what I described, is not in such a higher-order state and therefore does not believe there is moral reason to give to charity.

This is a form of ecumenical *expressivism* rather than cognitivism because the truth of the sentence does not necessarily align with the truth of the robust belief component of the judgement. In the example above, Luna's robust belief is not true (which I assert here without argument) and Jeffrey's robust belief is true (we can presume), and yet Luna's overall moral judgement is true while Jeffrey's is not. There *is* moral reason to give to charity. In saying so I am expressing the same higher-order state as Luna. If I am speaking linguistically correctly, I will *be* in the same higher-order state as Luna. However, my higher-order state may be realised by different lower-order states, such as moral approval of actions that promote net pleasure over pain and a robust belief that donating to charity has that property.

Toppinen's theory leaves open the kinds of desires involved as well as the mode of presentation of the content of the belief. Ridge spells out the details in terms of normative perspectives. "An agent's normative perspective just is a set of relatively stable self-governing policies about which standards to reject and accept. More specifically, a normative perspective in my sense is a set of relatively stable policies *against* accepting certain kinds of standards of deliberation."³ He adds that to count as having a normative perspective one must also plan to act and deliberate in accordance with the standards one has not ruled out. One of my worries about Ridge's theory is similar to my worries about Gibbard's theory. If we apply it directly to moral language it is too strong: just as I might coherently believe something is morally right without planning to do it, I might also coherently fail to plan to act and deliberate in accordance with my moral perspective. But Ridge, like Gibbard, is interested in the normative as a whole. I may think something is morally right yet not *all things considered* right to do, which explains why I may not always do what I take to be morally right.

The issue is now to explain moral judgement in particular. What separates thinking there is a *moral* reason to ϕ and thinking there is merely a prudential one? How do we partition the normative perspective? Perhaps,

³ Ridge (2014: 115).

following Gibbard, moral judgements are ones involving a normative perspective that rules out standards for forming certain reactive attitudes. In that case, my argument against structural solutions in the previous subsection is going to apply. This is why I prefer to speak in terms of (admittedly vague⁴) moral feelings such as approval and disapproval as the relevant desire-like states, rather than plans and intentions. The idea would be that each of us has a collection of these feelings which make up what we might call our *moral* perspective. If we agree with Gibbard that our perspectives are determined negatively via what we rule out, then we can believe our moral perspectives to be determined via our disapproval. If we prefer something positive (which you might think is required to explain motivation), then we can say our perspectives will be determined via our approval. And there's nothing stopping us from taking on both positive and negative aspects simultaneously like Ridge does. How exactly this works and how it links to motivation is presumably an empirical question. If I feel an emotion such as happiness when I play the drums, this does not mean I will *plan* to play the drums, but it is prone to play some role in my planning and increase the likelihood that I will play the drums in future. In a similar fashion, moral approval of actions that, say, are in accordance with either hedonistic or preference utilitarianism will, all other things equal, tend to (but not necessarily) have some positive weight in my planning. I will return to my favoured specification of the higher-order state view in §5 of the next chapter.

Hybrid expressivism has gone through a boom since Stephen Barker's (2000) paper and David Copp's (2001), yet the fruit of adding descriptive content into moral judgements was gestured at by earlier expressivists. The following quote from Gibbard seems to me to be an instance of the higher-order state view. He considers Cleopatra saying "It made sense for Anthony to give battle."

She expresses her acceptance of a system *N* of norms and a set *B* of factual beliefs with this joint property: that *B* includes the beliefs that *N*, as applied to Anthony's circumstances, permits giving battle. If she is sincere, the analysis says, she has some specific combination of factual beliefs and normative commitments that has

⁴ My view of moral attitudes is in agreement with what Gibbard (1990: 100) says about beliefs: "As we experience them, beliefs simply are what they are; they have a gestalt-like quality we can say little about. What we can hope is to give a logical-philosophical account of their structure. That is not to display their intrinsic psychological nature; it is to see how they fit into a system."

this property. We listeners, though, learn much less; we learn only the joint property. (Gibbard 1990: 93.)

Gibbard, now hailed as a leading pure expressivist, can be read in this passage as advocating for something very hybrid. What is expressed by a normative sentence involves desire-like acceptance of norms (or, as he goes on to prefer, desire-like *rejection* of norms) along with robust beliefs (simply called ‘beliefs’ in his (1990) book); and like the relational and higher-order state views, no *particular* norms or beliefs are expressed – a sincere speaker could in principle hold any of an infinite set of beliefs and attitudes so long as they are related in the specified way. The prescriptivist R.M. Hare and the emotivist Charles Stevenson also thought moral judgements had descriptive content as a kind of “secondary meaning,”⁵ secondary to the prescriptive or emotive primary usage. And while A.J. Ayer did not go so far as to bring descriptive content into the meaning of ethical language, the following passage seems only a step away. Here it looks as though Ayer, like Barker (2000) and Ridge (2006a), was proposing that expression of moral disapproval is directed towards a given *class* of actions that share a given descriptive property or properties, and that this combines with robustly factual beliefs to yield the moral judgement:

If our opponent concurs with us in expressing moral disapproval of a given type *t*, then we may get him to condemn a particular action A, by bringing forward arguments to show that A is of type *t*. For the question whether A does or does not belong to that type of a plain question of fact. (Ayer 1936: 111.)

I am not saying that hybrid expressivism is old hat. Early suggestive comments were made but there was nothing particularly substantive when it comes to making use of the hybrid nature of moral judgements to do theoretical work. This was the major difference when Ridge’s (2006a) ecumenical expressivism was introduced: the Frege-Geach problem was promised to be solved in a relatively cheap way. Because of this I am far more confident that *some* form of ecumenical expressivism is true than I am of how the details are to be spelled out. The general tactic is to offload the relevant linguistic complexities to the robust belief components.

⁵ Stevenson (1944: 22) and Hare (1952: 188). See Eriksson (2009) for a development of a hybrid expressivist theory inspired by Hare.

6.1 Putting Beliefs to Work

The Frege-Geach problem requires us to make sense of complex sentences involving moral terms and explain why the relevant logical relations hold between them. Let us address both at once by considering an instance of *modus ponens* with the expressed psychological state and a particular agent's potential realising lower-order states below each sentence.

S1: Torture is wrong.

M1: The higher-order state of having a moral perspective that forbids actions with properties K_1, \dots, K_n and a robust belief that torture has one or some of those properties.

R1: Moral disapproval of failing to maximise net pleasure over pain and the robust belief that torture fails to do so.

S2: If torture is wrong, extradition to America is wrong.

M2: The higher-order state of having a moral perspective that forbids actions with properties K_1, \dots, K_n and a robust belief that if torture has one or some of those properties, so does extradition to America.

R2: Moral disapproval of failing to maximise net pleasure over pain and the robust belief that if torture fails to do so, so does extradition to America.

S3: Extradition to America is wrong.

M3: The higher-order state of having a moral perspective that forbids actions with properties K_1, \dots, K_n and a robust belief that extradition to America has one or some of those properties.

R3: Moral disapproval of failing to maximise net pleasure over pain and the robust belief that extradition to America fails to do so.

The mental state expressed by the logically complex sentence S2 involves a logically complex robust belief component. Other logical operators are accounted for in the same way, i.e. by putting them into the content of the robust belief: 'either there is moral reason to reassess the case or he is guilty' expresses the higher-order state of, say, having a moral perspective that assigns positive weight to properties K_1, \dots, K_n and a robust belief that either reassessing the case has one or some of those properties or he is guilty.

Two questions arise. Does this explain why the correct logical relations hold, and what is it that makes *S3* *validly* follow from *S1* and *S2*? It is important to distinguish between them. The second question can simply be answered in the same way anyone else answers it. Perhaps *S1-3* is a valid argument because it is truth-preserving (necessarily, if *S1* and *S2* then *S3*) purely in virtue of the logical vocabulary and structure (it remains truth-preserving after any consistent substitution of the non-logical terms). But to be entitled to talk of truth we must show the correct inferences are licensed, and so we must answer the first question. Why are competent speakers licensed to infer *S3* from *S1* and *S2*? At first glance this seems very easy. Anyone accepting *S1* and *S2* will have robust belief components with content that implies the content of the robust belief component of the state expressed by *S3*, and so will be licensed to make that inference however we are normally licensed to make inferences in our robust beliefs; and, since the desire-like attitude expressed by *S1*, *S2* and *S3* is the same, there is no need to require an extra “logic of attitudes” to explain how the attitude in the conclusion is generated by those expressed by the premises. Looking “negatively” at validity, we see that the conjunction of the premises and the negation of the conclusion contradict. Ridge (2014: 156) explains this by pointing out any possible believer who accepted the premises and denied the conclusion simultaneously will have inconsistent robust beliefs. In short, according to the higher-order state view moral attitudes remain constant when interpreting arguments, and inferences are made entirely due to the robust belief components.

6.2 Problems with Validity

Schroeder (2009: 294-97) raised the issue that arguments such as *S1-3* are only truth-preserving if ‘those properties’ in *M1-3* refers to the same properties in each of the sentences. Of course it is part of the theory on offer that the phrase ‘those properties’ *does* in fact refer consistently, since it refers to the relevant descriptive bits of the speaker’s moral perspective which we assume remains constant throughout.⁶ However, this information is only believed by those who believe the theory itself. That’s not many people!

⁶ What if our moral views change during our consideration of the argument? Suppose we are utilitarians when we assent to *S1*, then we change to a deontological view when we assent to *S2*. It is clear that we ought not to infer *S3* until we have re-examined *S1* from our new moral vantage point; but this is true whether or not we are hybrid expressivists.

What about the majority of the masses who haven't even *heard* of ecumenical expressivism, let alone believe it? To use Schroeder's analogy, they may be like early Lois Lane considering the argument 'Superman is strong, if Clark Kent is strong then I am a walrus, therefore I'm a walrus.' The argument is truth-preserving but Lois doesn't know it because she needs the extra information that Clark Kent is Superman. Likewise, says Schroeder, non-believers of ecumenical expressivism would not be able to recognise the truth-preserving nature of S1-3 without extra information. The required extra information according to Schroeder (*ibid*: 296) is that "every sentence containing the word 'wrong' in fact expresses the very same desire-like attitude."

My response to this objection is to deny that ordinary speakers need *that* information in order to draw the inference. All they need to know is that the same properties are picked out in each premise. Luckily they do know this. They might not have thought much about it, but it is fair to assume that implicitly they will believe the same properties are picked out because this best explains why they make the inference. Better yet, just ask someone whether the same properties are referred to and I predict they'll say yes. The expressed psychological states may be opaque to them, but most metaethical theories attribute non-obvious psychological states to moral thinkers and it is no obstacle to recognising validity. Compare Lois Lane. All Lois needs to know is that Clark Kent is Superman. She needn't know anything about her own psychological states or about the underlying alien nature of Superman in order to make the inference in Schroeder's example. The same is true of ordinary speakers. They need the background knowledge that the same properties are picked out throughout moral arguments, and this they plausibly have.⁷

Lenman (2018) raises a number of problems for Ridge's (2014) theory. One of them in particular is pertinent to our current discussion of validity. Using 'P' and 'Q' for the objects of robust belief and 'X' for the normative perspective, he presents (2018: 3-4) the ecumenical interpretation of a moral *modus ponens* as follows.

X; P
 X; If P then Q
 Therefore: X; Q.

⁷ See Ridge (2014: 157-9) for a further and more detailed defence.

Since P and Q are bog-standard objects of robust belief they obey the right logical rules non-controversially. But even though the desire-like attitude is just hanging around not doing much, Lenman is worried about it.

Whatever X is, it better be logically tractable enough for *iteration* to be valid for it. X has to at least entail X. And that means we need a logic of attitudes after all. Of course iteration is a very trivial and simple logical principle. ... Nonetheless, iteration is not minimal enough for us to solve the Frege-Geach Problem so very simply. For if we characterize validity in terms of contradiction between the premises and the negation of the conclusion the validity of $X \vdash X$ will consist in the inconsistency of X and not-X. And for that we need to know what 'not-X' means. And explaining negation is right at the heart of the difficulty constituted by the Frege-Geach Problem. (Lenman 2018: 4.)

It is worth noting that X is not a proposition; it is a desire-like state and it is not expressed by any sentence. (The mental states expressed are higher-order states realised in part by states such as X.) Therefore I would deny that Lenman's presentation is "the logical form of the argument"⁸ – the logical form of a moral *modus ponens* is '*p*, if *p* then *q*, *q*.' Nevertheless, Lenman makes an important point. Given that we are earning the right to speak of the logical properties of moral sentences by showing the relevant mental states behave in the predicted ways, it seems as though we cannot exclude the desire-like X from our analysis of consistency. And indeed Ridge describes the self-governing policies that make up normative perspectives as "subject to the same strong demands of consistency and the like."⁹ Lenman therefore maintains that not much is gained from going hybrid if we have to account for desire-like validity too. As I discussed in §4.2 and §4.3, I am happy to accept that desire-like states such as plans are subject to consistency constraints, but I find it difficult to believe moral judgements *are* planning states. I think they are instead composed in part by moral attitudes of approval and disapproval.

My response to Lenman's objection is to deny that we need to understand not-X. 'Not-X' doesn't even make sense, for it is like saying 'not disapproval of causing pain.' To think that X can be iterated is a category mistake. Moral feelings don't logically follow from others or even themselves. They just *are*. To be sure, they can change and be influenced in

⁸ Lenman (2018: 3).

⁹ Ridge (2014: 115).

many different ways, and doing so is an important part of ethics. To persuade you of a moral claim I might aim to change your robust beliefs or I might tug on your heartstrings. But when assessing validity we keep the context fixed. If we didn't, 'Abe is fun therefore Abe is fun' would not be valid since we might be talking about a different Abe in the conclusion. In the context of the moral argument in question, X is the moral perspective of a given agent, and given that "an agent can at any point in time have only one [moral] perspective,"¹⁰ X is guaranteed to be constant throughout the context of the argument. Lenman is unconvinced. "What might be more plausibly claimed is that a *consistent* agent can at a given point of time have only one [moral] perspective."¹¹ Why is this more plausible? Moral perspectives are defined as *all* the moral attitudes an agent has, so I don't see how anybody can have more than one. X, then, is just a mental state that partly realises the judgements in the argument, serving to fix the content of the robust belief components and sometimes provide motivation for action.

I know from discussion with him that Lenman's hankering for consistency stems from the possibility of some whacky moral perspectives. He raised the prospect of a person hearing Nelson Mandela's perspective on the equality of races and thinking "Yes, I'll adopt that perspective" while also thinking of Adolf Hitler's perspective on the inequality of races and thinking "Yes, I'll adopt that one too." This person looks like they have two distinct (and inconsistent) moral perspectives. To get around this, we must clarify what this strange person is doing. Are they agreeing with Mandela that, say, race is not intrinsically morally significant while agreeing with Hitler that race *is* intrinsically morally significant? This poses no challenge for the ecumenical expressivist. These are *judgements* rather than perspectives in our technical sense, and as such they involve robust beliefs. Someone may have one moral perspective X and form both of these moral judgements due to inconsistent robust beliefs. Instead, we might imagine someone who adopted all the desire-like states comprising Mandela's moral perspective and all the desire-like states comprising Hitler's. Is this a case of an inconsistent agent with two moral perspectives? No. To see why it is not, let us examine a clearer case.

Imagine someone who morally approves *and* disapproves of actions that maximise paperclips. This person might seem inconsistent. However I

¹⁰ Ridge (2014: 152).

¹¹ Lenman (2018: 5).

would urge us to think of this person as merely discordant. If their robust beliefs are accurate they will believe instances of paperclip maximisation to be simultaneously right and wrong, but there is nothing logically inconsistent about that, though of course it is morally incorrect. Admittedly there does appear to be something going wrong with this person beyond having a false moral opinion. Someone who believes abortion to be morally wrong is mistaken but not jarring in the way the moral lover and hater of paperclip maximisation is. Expressivists are in an excellent place to explain why: the agent's attitudes are discordant in the sense discussed in §3.¹² Yet as we noted there, discordance is not inconsistency. Furthermore, this strange person would still have only one moral perspective: approval and disapproval of actions that maximise paperclips.

To summarise my reply to Lenman, we must distinguish between moral judgements, which can be inconsistent, and the attitudinal moral perspectives which partly realise those judgements, which are neither consistent nor inconsistent. Agents morally like what they like and dislike what they dislike, and there are no logical constraints on *what* they like and dislike. In order to solve the Frege-Geach problem we just have to show why the proposed mental states expressed by moral sentences are expected to obey the same rules we take moral sentences to obey. This is achieved by the fact that the realising robust belief components are guaranteed to have content that obeys those rules, combined with an unchanging non-cognitive perspective. (This contrasts with the pure expressivists who argue our non-cognitive attitudes change in accordance with logical rules of inference.) As Schroeder noted, merely having content that is truth-preserving is not sufficient for speakers to *realise* that fact and be licensed to make the right inferences. I contended that ordinary speakers do in fact have the relevant background knowledge. This explains why moral judgements behave logically as we expect them to.

¹² This may be an advantage over some pure cognitivist views such as robust realism. Robust realists should agree that there is nothing logically inconsistent about assigning the rightness and wrongness properties to a single action. Presumably the strangeness of our agent would be explained by their conceptual misunderstanding – conceptual analysis should reveal to us that nothing can be simultaneously right and wrong. But I find the agent stranger than someone merely conceptually confused, such as someone who thought bachelors might be married. I prefer the “discordant attitude” explanation. Our agent seems to me more analogous to someone who likes and dislikes the same thing in the same respect. This could be an interesting line of argument to pursue against metaethical theories that exclude the involvement of non-cognitive attitudes.

6.3 Options for Hybrid Expressivism

As I charged against pure expressivism in §5.1, capturing logical structure is not enough for a plausible theory; the mental states themselves must be plausible in their own right. Strandberg (2015) has recently challenged all hybrid theories with a trilemma regarding the non-cognitive attitude involved in moral judgements. There are three theoretical options for the hybrid expressivist to choose from. The attitude expressed by a moral sentence might be directed towards the subject of the sentence, towards actions that have a certain descriptive property, or towards the descriptive property itself.¹³

The first option is defended by the early Copp (2001) and Eriksson (2009). The mental state(s) expressed by ‘stealing is wrong’ would include a desire-like attitude towards *stealing*, whereas the desire-like state relevant to ‘taxation is wrong’ would be directed towards *taxation*. As such, defenders of this theory are required to tell us a story about how these new attitudes are generated and why they obey the right logical rules. Since they need a logic of attitudes they are subject to the Frege-Geach problem in the same ways as pure expressivists. If the problem with purity really is the Frege-Geach problem, we would be wise not to go hybrid in this manner.

The third option comes out in a couple of passages by Ridge (2009: 197 and 2014: 7), but I would be hesitant to attribute the view to him based on these. Nevertheless, it is an option that agents may approve and disapprove of *properties*, and these properties are what fixes the reference of the robust belief components of moral judgements. Since the attitudes can remain constant throughout evaluation of an argument this has the advantage of not requiring a logic of attitudes. Yet the idea itself is rather implausible. Consider the property of maximising pain. I don’t care very much about the actual property, which many consider to be causally inert in Plato’s Heaven. What I want to make sure of is that this property is rarely *instantiated*. My disapproval is better understood as directed towards *having* a certain property rather than towards the property itself. This is the second option, and in my view the correct option, defended by hybrid theorists such as Ridge (2006a, 2014), Boisvert (2008) and the later Copp (2014, 2018).

¹³ Strandberg (2015: 95) notes the option to combine some of these, as Eriksson (2009) does. I will not discuss it because a combination inherits all the problems of the wrong views.

Strandberg has three complaints against the second option. One is a problem with validity similar to Lenman's objection discussed in the previous subsection. Strandberg (2015: 104) worries that, in relation to the desire-like attitudes expressed, the second option "has difficulties to capture the mentioned notion of commitment in relation to valid arguments." My response is the same. Having the desire-like attitude involved in the premises of a valid argument does not logically *commit* us to having it in the conclusion, in the sense that we have made a logical mistake if we fail to have it. (We *couldn't* fail to have it.) The desire-like attitudes are not eligible for this talk of commitment and inference. Inferences occur in the judgements as a whole, which are in turn explained via inferences in robust belief.

Another of Strandberg's complaints is that expressivism loses its intuitive appeal if what we approve and disapprove of is something general rather than specific. He writes (*ibid*: 97) that if someone asserts that lying is wrong, "we immediately presume that she has a negative attitude towards the subject of the sentence, viz. lying. ... [I]t seems much less evident that we would presume that she has an attitude towards all actions that have a certain property." Regarding his first point, I agree that we tend to assume speakers have attitudes towards the subjects of their sentences. But this is easily explained: if we have an attitude A towards acts insofar as they have property B and we believe that lying has property B, there will be a fairly reliable connection with having an attitude towards lying. Admittedly this leaves open the possibility of someone who hasn't "put A and B together"¹⁴ in the right way and, for instance, judges that stealing is wrong yet has no negative attitude towards stealing. Consequently, hybrid expressivism of this sort yields a weaker form of motivational judgement internalism than pure expressivism. Weaker forms are often thought to be more intuitive but to evaluate whether they in fact are would take me too far afield.¹⁵ It is sufficient here to note that many (if not most) philosophers in metaethics have denied that any linguistically correct assertion of 'lying is wrong' *guarantees* the speaker has a negative attitude towards lying, and so it cannot be assumed outright that this is a disadvantage.

Finally, Strandberg (and also Lenman 2018: 8) argues that this option for hybrid expressivism cannot account for people with agnostic, sceptical,

¹⁴ This idea is found in Smith (1994: 92), referred to in the first footnote of chapter 1.

¹⁵ See Björnsson et al (2015) for a variety of arguments regarding motivational internalism.

unclear or non-existent views in general moral theory. Here are the four examples he uses (2015: 99) to illustrate his point.

1. If lying were wrong, people would be less inclined to lie.
2. If lying is wrong, then lying is wrong.
3. It is not the case that lying is wrong.
4. Either lying is wrong or it isn't.

Someone who accepts any of the four sentences need not think there are any wrong actions, so it seems strange to presume they have any non-cognitive attitudes towards actions with a particular property or set of properties. Yet that is what the hybrid expressivist view states. Strandberg asserts that a moral sceptic could accept all four of these sentences, and it would be incorrect to suppose they must have a moral perspective.

To defend hybrid expressivism I propose we examine the cases more closely. Take sentence 1. Why would a moral sceptic believe that? It seems more likely to me they would believe that if lying were *considered* to be wrong then people would be less inclined to lie. Sentence 3 looks easier. An error theorist can believe 3 while, and in virtue of, thinking there are no wrong actions. I respond to this more fully in §4.1 of the following chapter where I interpret adherence to error theory as a moral view. Roughly, error theorists have a moral standpoint that is more demanding than what they think reality gives us. Therefore in this case I bite the bullet: error theorists *do* have a moral perspective, and in my view they should change it and direct their moral feelings towards things insofar as they have more down-to-earth descriptive properties like promoting pleasure and autonomy. Alternatively, someone might believe 3 but simply be unsure about other moral questions. In this case I would again say that the person has a moral perspective. There is some feature she morally disapproves of and she believes lying doesn't have it, but she may not know exactly what this feature is. This is the case with most of us. We are usually more certain that particular things are right and wrong than we are about *what it is about them* that makes them right and wrong.¹⁶ I contend that anyone who fully believes that lying is not wrong will have a moral perspective in order to have an opinion about the particular descriptive features absent from lying.

¹⁶ Ridge (2006a: 315) describes this as a kind of *je ne sais quoi* feeling, analogous to when we hear a piece of music and we know we like it (or not), but we don't know quite what it is about the music that we like; yet we still maintain that there is some descriptive feature of it that we like.

I admit that it is possible for somebody to have no moral feelings, perhaps due to neuropsychological issues or a very poor upbringing. We expressivists have to believe that this person without moral feelings doesn't make genuine moral judgements. I find it difficult to imagine anyone accepting sentences 1 and 3 without understanding what they mean, but I can imagine them assenting to 2 and 4. Given that 2 and 4 also express hybrid states involving the non-cognitive attitudes we are assuming this person lacks, this looks to be a problem. However, the problem dissipates as soon as we imagine an analogous case in the descriptive realm. Suppose Tom has no knowledge of physics and no understanding of mass. He then considers the sentences 'if Socrates has mass, then Socrates has mass' and 'either Plato has mass or he does not.' If we tell Tom that these sentences are meaningful and ask him to bet on their truth or falsity, if he is rational then Tom will bet on their truth. Yet Tom does not have a fully-fledged *belief* in these sentences because he does not understand all the terms. Likewise, someone with no moral attitudes whatsoever may assent to the truth of 2 and 4 without really understanding them. You can spot a logical truth in virtue of its form, and that's all that's really going on here. With the distinction between full understanding of a sentence and mere recognition of its logical structure, I can agree that amoralists would bet on the truth of 2 and 4 while maintaining that *fully-fledged* moral judgements will always involve moral attitudes of a general kind. I therefore embrace Strandberg's second option for hybrid expressivists.

Conclusion

In this chapter I have considered several pure expressivist solutions to the Frege-Geach problem and argued they face severe challenges. In §3 I carried the higher order attitude solution as far as I could and ended up with a kind of inferentialist solution, but in §4 I showed why Dreier's 'Bob is hiyo' objection to minimalist solutions (discussed in §2) also causes trouble for inferentialist solutions: we have to *show* that the relevant states of mind are the right type for inferences rather than assuming it from the offset. One attempt utilised the concept of unrealisable attitudes, but this ended up implausibly ascribing incoherence to many moral judgements that seem sensible. While this problem was tentatively solved by adding extra structure in §5, I argued that this extra structure construes our judgements

to be about the wrong subject matter (as well as committing us to a highly controversial global form of expressivism).

The Frege-Geach problem is a rather ghastly one for pure expressivists. Fortunately the problem can be overcome with the addition of robust beliefs. In §6 I explained the hybrid expressivist solution to the Frege-Geach problem and I argued that the objections from Schroeder, Lenman and Strandberg do not succeed. The problem is therefore purely with purity, and expressivists should embrace robust beliefs as constituent components of moral judgement.

~3~

Metaphysics for Expressivists

Moral metaphysics is inquiry into the nature and structure of moral reality and its relationship with non-moral reality. Are there moral properties and facts? If so, what sort of things are they? If moral facts reduce to facts about natural things like pain and pleasure or supernatural things like the will of the gods, what is the nature of their reduction and why do moral claims feel so radically different to bog-standard descriptive claims? If instead moral facts are not reducible to anything we can refer to in non-normative terms we face metaphysical worries about queerness and supervenience, as well as epistemological challenges. Expressivism is often thought to include a rejection of moral metaphysics. Indeed it might be seen as an advantage to avoid all those difficult questions about the nature of moral properties. It is tempting to think moral practice is not in the business of making metaphysical commitments if all that is going on is the expression of non-cognitive attitudes like approval and disapproval. I argue this is a mistake: if moral facts exist (as most expressivists, being quasi-realists, agree), it makes sense to ask what they are like, and if the question can be formulated, we should look for a sensible answer.

One answer given in the literature on behalf of expressivists (Dreier 2015b, Ingram 2017, Toppinen 2018) is that moral facts exist and are not reducible to bog-standard descriptive facts. Simon Blackburn and Allan Gibbard do sometimes appear to aim their quasi-realism at mirroring a non-natural brand of realism. Yet they also speak of moral sentences being “made true” by natural facts, or of moral facts being “constituted” by them, which is at least suggestive of some kind of natural reduction. I argue it is difficult if not outright impossible to interpret claims of reduction or non-reduction of moral properties on a pure expressivist picture. Hybrid forms

of expressivism are better equipped. I provide a transcendental argument that the only sensible position for expressivism is one that reduces moral properties to bog-standard properties that can be referred to in non-normative terms. I then plump for a form of quasi-realism that mirrors – or perhaps just is – synthetic natural realism, where we discover just how close this view is to Cornell realism. Although the difference is minute given the advent of hybrid forms of realism, I end by showing why expressivist varieties of synthetic natural realism are superior to cognitivist ones.

1. Robust, Relaxed, or Quasi?

Here's a moral fact: it is wrong to torture innocent people merely to pass the time. This fact doesn't depend on what we think of it, and it seems very different to the fact that torturing innocent people merely to pass the time causes lots more pain than pleasure, or any other descriptive fact. (Descriptive facts are those that can be referred to without using normative terms.) Suppose this is because moral facts *are* distinct from bog-standard descriptive facts. Some realists (Enoch 2011, Fitzpatrick 2016) believe morality is backed up with a robust metaphysics according to which moral properties exist on ontologically firm footing. This requires defence against a plethora of metaphysical arguments. Other realists are more relaxed. Relaxed realists (Dworkin 2011, Parfit 2011, Scanlon 2014) believe there are moral facts but they think this is a moral issue, not a metaphysical one. Quasi-realists (Blackburn 1993, Gibbard 2003) also believe in moral facts while shunning metaphysics (Gibbard 2011, Horgan and Timmons 2015b).

In what follows, I take expressivists and quasi-realists to be the same people. It is of course possible to be an expressivist and not a quasi-realist. One may believe moral judgements are desire-like and yet not be on board with the program of vindicating realist-sounding ideas such as the existence of mind-independent moral facts.¹ It is also possible to be a quasi-realist and not an expressivist. One may be a different sort of anti-realist and still try to vindicate realist-sounding discourse.² Since the project of quasi-realism has traditionally gone hand in hand with expressivism I will only be concerned with that coupling here.

¹ A.J. Ayer (1936) was certainly in this camp.

² Error theorists who argue we should retain, rather than abolish, our current moral practices may fall in this category.

1.1 The Realist Commitments

What are these realist-sounding ideas? Stephen Ingram (2017) attributes the following claims to robust realists, relaxed realists and quasi-realists.

Truth. There are substantive moral truths.

Properties. Substantive moral truths are truths about the moral properties of certain acts or types of act.

Non-Reductivism. Moral properties are not reducible to descriptive properties.

Non-Naturalism. Moral properties are beyond the purview of the natural and social sciences.

Mind-Independence. Moral properties are not constitutively dependent on any agent's or set of agents' actual or hypothetical responses to those properties, or to the world. (Ingram 2017: 490.)

These are what Ingram dubs the Realist Commitments. As mentioned in the introduction, it is not obvious whether quasi-realists must accept *Non-Reductivism*.³ They are clear about their position on moral *concepts*: moral concepts do not reduce to non-normative ones.⁴ But when it comes to moral *properties*, it is possible to interpret the quasi-realist position as one open to natural reduction. The majority of this paper will describe how. Until I show otherwise in §3, let us assume that quasi-realists assent to all of the Realist Commitments. Being expressivists, they think their acceptance of the Realist Commitments is necessarily constituted at least in part by desire-like states (as I argued in chapter 1). Robust and relaxed realists deny this. This is one of the differences between quasi-realism and the two realist positions. However, this is a difference in how to characterise *us* in accepting the Realist Commitments, not a difference in characterisation of the claims themselves. What distinguishes the two cognitivist views – robust and relaxed realism – is whether they interpret the Realist Commitments as “internal” or “external” to moral discourse.

Robust realists, being robust, believe the Realist Commitments are strong, tough, external metaphysical claims. Relaxed realists (and quasi-realists) interpret them as first-order moral claims, internal to moral

³ In correspondence, Ingram tells me he agrees with this. In his (2017) paper he was more interested in the question of whether the Realist Commitments are first-order moral claims.

⁴ “As a device of generalisation, the good predicate will indeed resist analysis, resist any account of empirical or causal access, and bear a relation of supervenience to other properties.” (Blackburn 2010: 43.) Here too is Gibbard (2003: 32): “All *properties* are *natural*, but some *concepts* are *non-naturalistic*.”

discourse and subject to the norms of moral inquiry. Exactly what the distinction amounts to is difficult to pin down. Ronald Dworkin (1996: 88) says external claims “purport to stand outside a whole body of belief, and to judge it as a whole from premises or attitudes that owe nothing to it.” I find this perspective natural when it comes to describing the nature of moral judgements.⁵ It seems as though one may argue about whether motivational internalism is true or false without taking any stand on substantive ethical questions, for instance, or whether atomic moral claims express pure or hybrid states of mind. Perhaps we can imagine a morally detached scientist observing morally engaged humans and describing what’s going on in their heads. But I find it difficult to see how we might be able to take such a morally neutral stand when we evaluate some of the Realist Commitments. An error theorist who denies the existence of *Truth* seems to be making a moral claim, albeit metaphysically informed: to say that, no, it is not true that indiscriminate torture is wrong, is to take a (morally abhorrent, in my view) stand on first-order moral issues.⁶ This applies similarly to the other Realist Commitments – acceptance of them appears to require taking a general moral stance. Here is Blackburn on what we’re up to when we think about what moral properties depend on:

We should agree that there is only one way of assessing propositions of the form ‘the value of X depends upon Y’ and it is, indeed, Moore’s way or Mill’s way. We must contemplate the scenario in which Y is varied and see whether as a result X varies. But as the beginner suspects, we will necessarily be ‘standing within’ as we do this, or in other words, deploying our own evaluative sensibilities. This is in itself a perfectly good thing to do, and it needs stressing that there are perfectly good questions of this kind, *for first-order ethicists*. ... But what about the metaethicist, trying to understand the Place of Value in the World as a whole?

⁵ Although not necessarily true, particularly if the normative does in fact creep down into domains such as the expression relation and characterisations of mental states as discussed in §1.1 and §1.3.2 of chapter 1.

⁶ Error theorists may want to separate the two domains such that one may metaethically believe all substantive moral claims are false but ethically believe torture is in fact wrong. Perhaps this can be done without incoherence by differentiating different mental states: when doing metaethics we believe ‘torture is wrong’ is erroneous but when doing ethics we (say) make-believe ‘torture is wrong’ is true and pretence-assert that torture is wrong (Joyce 2001: 199-204), and hence there is no contradiction. Mackie himself (1977: 16) thought one may be a second-order but not a first-order sceptic, and Olson (2014: ch.9) argues for “moral conservatism” – continued engagement in ordinary morality despite its underlying error. To evaluate these possibilities would take me too far afield.

Can't she use MM dependencies as a guide? No. ... We cannot pretend to escape from using our own sensibility as we use it. (Blackburn 2010: 29-30.)

According to Blackburn, the only way to judge whether moral properties depend on something, such as an agent's responses, is to make a first-order moral judgement. Of course it is clear enough that the Realist Commitments are too abstract to determine the answers to many moral questions – the mind-independence of moral wrongness tells us nothing about *which* acts have this property, if any – but moral concerns and arguments still bear upon our acceptance of them. In other words, the implication goes the other way. We make moral judgements like 'it is true that torture is wrong' which implies *Truth*, and 'torture is wrong regardless of any actual or hypothetical responses to it' which implies *Mind-Independence*. In this way we can understand the relaxed realist's claim that the Realist Commitments are to be accepted "inside" morality, from "within" a moral perspective.⁷

1.2 Is it Possible to Reject Metaphysics?

We have seen some examples of how to interpret Realist Commitments as internal to moral discourse and subject to moral argumentation. Robust realists think this is not enough. They want to interpret the Realist Commitments as metaphysical claims external to moral discourse and "to be assessed primarily from that standpoint by appeal to metaphysical argumentation."⁸ As a moral being I personally find it difficult to imagine assessing the Realist Commitments in some morally neutral way. That said, in contrast to the relaxed realist whom we shall get to shortly, I do believe there are metaphysical arguments against moral reality which do not rest on moral premises, such as the arguments from queerness and disagreement. However, we would have to square the consequences of these arguments with our moral opinions, and it is in this sense that they are not morally *neutral*. In short, certain arguments (e.g. from queerness and disagreement) seem to rest on non-moral premises (e.g. that there exists widespread moral disagreement) yet imply something about morality.

⁷ I believe robust realists need not deny this. There doesn't seem to be anything inconsistent about an external metaphysical claim being accepted internally from within a moral viewpoint. This is discussed in the next subsection.

⁸ Ingram (2017: 491).

Ingram (2017: 503) argues, successfully in my view, that the fact that a claim has moral implications does not mean the claim itself must have been a moral one all along. Here is one of his analogies: the fact that the theological claim *God created the universe in seven days* implies the metaphysical claim *naturalism is false* does not mean the initial claim is part of metaphysical discourse. I therefore concede that one may choose not to believe the Realist Commitments without drawing upon one's moral perspective to reach that decision. What I find far less clear is whether one can *believe* the commitments from a purely external perspective. I think it is fair to say that moral realists come to the metaethical table with certain moral views that they feel cannot be justified on an anti-realist basis. The majority of the realist's philosophical time is spent *defending* realism against objections that seek to undermine it, and this may distract us from the morally charged basis of their belief in a moral reality. And, tellingly, when it comes to offering positive arguments in favour of their view, realists call upon moral premises: the "moral fixed points" Terence Cuneo & Russ Shafer-Landau (2014) use as a starting point are themselves moral claims, for example, and David Enoch's "Argument from the Moral Implications of Objectivity (or Lack Thereof)" also rests "on first-order, moral grounds" (2011: 16).

In light of considerations like these, relaxed realists are emboldened to profess that the Realist Commitments are *only* moral claims, lacking any metaphysical seriousness. This supposedly has the advantage of fencing off morality from any external worries. If you have a problem with the Realist Commitments it can only be a *moral* problem, to be resolved internally with moral argument. Given that the arguments from queerness, supervenience and disagreement are typically understood as external to moral discourse, they simply miss the mark. Apparently, we can relax.

Only I don't think we can. If there are moral facts, as all realists agree, why can't we ask what these facts are like? What is it that blocks any metaphysical examination of these things? Derek Parfit's answer involves distinguishing between different kinds of existence. Actual concrete objects like me and you and the chair I'm sitting on exist in the narrow sense of the word 'exist' whereas merely possible objects exist in a wider sense. Finally, abstract objects exist "not only in the wide sense, but also in a distinctive, non-ontological sense."⁹ It is in this non-ontological sense that moral

⁹ Parfit (2011 vol.2: 480).

properties exist. Therefore, moral claims have no ontological commitments to worry about.

One may wonder if there is anything going on beyond merely using different words to describe the same situation. Consider the following case. Sloan dances sloppily and Ellen dances elegantly. I would find it natural to say that Sloan and Ellen are both dancing in the very same sense of the word ‘dance’ but they are dancing in different ways. Now suppose Parfit insists that, in fact, there are at least two different senses of ‘dance’ – Sloan is dancing in the sloppy sense and Ellen in the elegant sense. I find it difficult to get excited about this. In this case I suspect there is no real disagreement about what’s occurring. We agree Sloan and Ellen are both dancing, and we agree they are dancing differently. Whether we say they are dancing in the same sense or in different senses of ‘dance’ seems to be a merely verbal dispute, not a dispute about the way the world is.

Similarly, it is difficult for me to pinpoint a disagreement I have with Parfit about the existence of moral properties. I find it natural to say that moral properties exist in the same sense of the word ‘exist’ as it appears in ‘I exist,’ but I exist in a different *way* to moral properties. I’m spatio-temporal, for instance, whereas moral properties are not. Parfit, on the other hand, thinks I exist_{ontological} whereas moral properties only exist_{non-ontological}. If he wants to phrase it that way that’s fine, but I cannot see why this helps Parfit with his situation. Why are non-natural, irreducibly moral properties any less queer or any less difficult to gain knowledge of if they exist in a non-ontological sense instead of an ontological one? They still seem pretty queer to me. And on top of that we have *another* queer thing to explain: non-ontological existence and our sense of it. As if that weren’t bad enough, Jussi Suikkanen (2017) raises the worry that relaxed realists are going to struggle to find a theory of truth that allows them to maintain a distinctive position.¹⁰

Continuing with this line of thought, it seems entirely legitimate to ask the relaxed realist what these non-ontologically existing properties are like.

¹⁰ Roughly, relaxed realists cannot adopt minimalism because of how it creeps – ‘killing is wrong’ implies it is true that killing is wrong, which implies it is a fact that killing is wrong, which implies it is an ontological fact sewn into the fabric of reality that killing is wrong – thus collapsing their view into robust realism. Correspondence theories of truth seem to require ontological facts that *make* moral claims true. Since relaxed realists reject these ontological facts, substantive moral claims would not be true and their view collapses into an error theory. (Parfit explicitly rejects correspondence theories of truth for moral claims (2011 vol.2: 746).) Finally, coherence or pragmatist theories of truth do not capture enough mind-independence, collapsing relaxed realism into subjectivism, relativism or idealism.

Do they reduce_(non-ontological?) to natural properties or any other ones? Their answer, in line with their assent to the Realist Commitments, is ‘no’. In which case we can ask the question: why is it that moral properties supervene on natural facts if they’re entirely distinct existences? Indeed not only are they distinctly existing, according to Parfit they exist in completely distinct *senses*, which makes their necessary connection all the *more* mysterious, not less.¹¹ In short, it seems entirely possible to ask the same sorts of questions about moral properties, regardless of whether you label their existence ‘ontological’ or not.

The relaxed realist may insist that the questions I’m asking are still internal to moral inquiry, and are therefore not metaphysical. If this is what they want to say, again I will not be excited. I will reply that even if you say I’m not asking questions_{metaphysical}, there are still questions_{non-metaphysical} (which look suspiciously similar) I want you to answer, and any insistence that you don’t *need* to answer them will not reassure me. The lesson I think we should take from this is that, whether or not the Realist Commitments are best understood internally or externally, *we can still formulate questions* about the nature (whether ontological or non-ontological) of moral properties that, we all agree, very much exist in one sense or another.¹² For my part, perhaps due to my bias towards minimalism, I find it most natural to say that the Realist Commitments are both internal *and* external: they must be accepted from a moral vantage point and they are also metaphysical claims about the moral world. I do not see a contradiction in believing this to be true.

2. Metaphysical Sidesteps

My goal in the previous section was not to argue there are no differences between robust and relaxed realists. My goal was to argue that, no matter whether the Realist Commitments are understood as internal or external to moral discourse, the same kinds of questions can be raised. This is important because of Gibbard’s (2011) response to Sharon Street (2011), as well as Terence Horgan and Mark Timmons’ (2015b) insistence on a modest

¹¹ See Mintz-Woo (2017) for more detailed arguments on the mysteriousness of relaxed realism.

¹² See also McGrath (2014). She argues that relaxed realists, in their frequent appeals to mathematics, cannot evade the kinds of questions we want to ask robust realists.

form of quasi-realism. Street argued that her evolutionary argument against realism also applied to quasi-realism. Gibbard's response was, in effect, to deny that quasi-realists want to "mimic" the strong metaphysical claims of robust realism; instead they are only aiming to mirror the relaxed forms of realism espoused by Dworkin, Parfit and Scanlon, and consequently evade the troubling metaphysical worries generated by robust realism. I hope I have given at least some reason to think the worries remain, even if they are interpreted internally. Street argued quasi-realists can't have it both ways (the benefits of mind-independence without the drawbacks) when it comes to her evolutionary argument and suggested this applies across the board: quasi-realists can't have it both ways *ever*. I have come to accept her point of view. Metaphysical questions may be asked internally, and if an internal answer is possible then we should look for it.

In §1.1 we saw how Blackburn treated *Mind-Independence* as a moral question. In believing it to be true, he takes up an internal, moral standpoint: he considers worlds where the inhabitants approve of torture and engage in it, and he assesses their engagement as morally wrong. In general, Blackburn thinks agents' actual or hypothetical responses are not the only morally relevant features. It is important to stress that this first-order moralising is not his job as an expressivist – expressivists, in making sense of moral claims, only need to provide an account of the mental states that constitute the appropriate judgements (and as quasi-realists, they need to vindicate the realist-sounding claims by showing the mental states they express are sensible – that they can be coherently accepted). This is what James Dreier (2015b: 283) calls "the Expressivist Sidestep." To make sense of one of the Realist Commitments, then, quasi-realists must first assign a sensible desire-like mental state that constitutes accepting it, and then it is a matter for first-order moralising whether or not to accept it.

Expressivists may differ in their moral views, and so by interpreting the Realist Commitments morally, expressivists need not agree on them: one may be an expressivist and reject, say, *Mind-Independence*. But how? Suppose we disagree with Blackburn's conditionals. We think that, were we to approve of torture for instance, torture would be right. This is insufficient to reject *Mind-Independence*. Consider robust realists for illustration. Robust realists do not think that moral properties *constitutively* depend on agent responses. However, their view is consistent with moral properties being assigned in the following way: an act is right if and only if the agent plans to do it and it is wrong otherwise. In other words, rightness may be

correlated with agent responses without being *reduced* to, or constitutively dependent upon, agent responses; agent responses don't *make* something right and wrong but maybe they match up nevertheless. Is there a way for expressivists to make sense of this distinction? In general, can expressivists make sense of claims to do with the nature of moral properties?

2.1 Pure Sidesteps

For expressivists, to “make sense” of a claim is to characterise a coherent mental state that constitutes acceptance of it. For the *pure* expressivist, straightforward atomic moral judgements are only desire-like states. These could be plans, goals, intentions, commitments, states of *being for*, or some kind of affective feeling. In what follows I use ‘approval’ and ‘disapproval’ (and the ‘H!’ and ‘B!’ operators) as placeholders for whatever the correct attitude is. Here is an example to begin with.

Claim: torture is wrong.

State: B! (Torture)

To accept *Claim* is to be in *State*. So far, so simple. In the previous chapter we saw the difficulties raised by introducing logical operators such as negation and disjunction. For the moment, though, we want to discuss claims *about* moral properties, rather than moral claims about other things like torture. Now, we could paraphrase *Claim* to look on the surface to be about a moral property, but because it is a paraphrase of the very same claim it seems right to assign the same mental state:

Claim: the moral property of wrongness is instantiated by torture.

State: B! (Torture)

In general, asserting that a property P is exemplified by E is equivalent to ascribing P of E. The problem arises when we consider claims about moral properties that cannot be rearranged in this way into simple ascriptions of moral properties to other things. Consider the claims below, drawn from the discussion before the beginning of this subsection.

Mind-Correlation: moral properties are necessarily correlated with the actual or hypothetical responses of agents.

Mind-Dependence: moral properties are constitutively dependent upon the actual or hypothetical responses of agents.

The kinds of conditionals Blackburn draws upon in his acceptance of *Mind-Independence* are too coarsely grained to distinguish between the two claims above. I might accept, for instance, that all actual and possible acts of torture performed by agents who plan to torture are morally right, and all acts of torture performed by those who *don't* plan to torture are morally wrong. I might accept something similar across the board:

Plan-Correlation: all actual and possible acts X performed by agents who plan to X are morally right.

State_{P-C}: H! (Enacting one's plans, whatever they are)

Plan-Correlation is a form of *Mind-Correlation*. The trouble is that I cannot see a way for a pure expressivist to distinguish accepting a form of *dependence* over and above accepting *correlation*, and the burden of proof is on the pure expressivist to do so. If I am right about this, then it casts doubt on Blackburn's acceptance of *Mind-Independence* – what he was really doing was rejecting *Mind-Correlation*. Blackburn may insist this is okay because the negation of *Mind-Correlation* implies *Mind-Independence* by contraposition: if moral properties depended on agents' responses they'd be correlated with them, but they're not, so they don't. However, this is jumping the gun. Blackburn is not entitled to make this argument; he has not “earned the right” to speak of dependence over and above correlation. In speaking of the implication from one to the other he is assuming we have a grasp of what is being inferred. I do not. Not on the pure expressivist picture, anyway.

The problem is a symptom of the inscrutability of moral properties – it is the difficulty of assigning pure desire-like states to claims *about* moral properties. We have seen how correlation and dependence are difficult if not impossible to distinguish for a pure expressivist. Let me show you another couple of claims.

Pleasure-Correlation: necessarily, things are good if and only if they are pleasurable.

Pleasure-Reduction: goodness reduces to being pleasurable.

State_{pleasure}: H! (Everything actual or possible that is pleasurable)

Which claim does *State_{pleasure}* constitute acceptance of? Once again, the pure expressivist does not appear equipped to make distinctions fine-grained enough for the task. We could try to give it a positive spin. Perhaps this shows that, after all, expressivism avoids metaphysics. If there are no

possible answers that make any sense, then the questions must have been nonsense to begin with. One simply cannot make any claims about moral properties that cannot be rephrased as ascriptions of moral properties to non-moral things. Indeed the pure expressivist is better off than the relaxed realist – at least we now have an *explanation* of why metaphysical questions are bunk. But whatever we gain from this spin is lost on two fronts. First of all, quasi-realists often *want* to say some things about moral properties, such as their being irreducible (which I'll later dispute), non-natural and mind-independent. Pure expressivists can no longer do so. Second of all, the sentences analysed in this subsection intuitively make sense, and so any theory that cannot make sense of them loses some plausibility.

Worse still, the amount of seemingly sensible claims the pure expressivist cannot recognise extends far beyond what we have considered. For instance, the claim 'the moral property of rightness is the moral property of rightness'¹ appears obviously true to me, but without characterising a pure desire-like mental state that constitutes acceptance of this seemingly trivial claim, the pure expressivist will have to maintain it is nonsense. This is a reason, I think, not to reject moral metaphysics, but to reject pure expressivism.

2.2 Hybrid Sidesteps

Hybrid expressivist theories are far better equipped to deal with the questions raised in this chapter. I suspect all hybrid theories will be able to make sense of metaphysics but I will deal in terms of the view I favour. This is Teemu Toppinen's (2013) *higher-order state* view, heavily inspired by Michael Ridge's (2006a) ecumenical expressivism. Atomic moral sentences express higher-order mental states composed of a desire-like state and an ordinary robust belief which are related in a certain way. Furthermore, each higher-order state is multiply realisable. Let us sidestep *Claim* once more.

Claim: torture is wrong.

State_H: [B! (Actions with property K) & belief that torture has that property]²

¹ Thanks to Lewis Brooks for this example.

² Toppinen (2013: 254) allows multiple different kinds of desire-like states might be eligible to realise the higher-order states, and he allows there may be several properties we have

$State_H$ might be realised by an agent disapproving of actions that fail to maximise pleasure and believing torture has that property, or by a peculiar agent disapproving of actions that are noisy and believing torture has that property. All logical connectives get shoved into the belief component: the negation of *Claim* is accepted by being in the higher-order state [B! (Actions with property K) & belief that it is not the case that torture has that property], and to accept that if *Claim* is true then there will be a revolt is to be in the state [B! (Actions with property K) & belief that if torture has that property then there will be a revolt].

Since we now have a robust belief component that robustly ascribes a robust property, we are much better placed to answer metaphysical questions. Furthermore, this does not preclude an internal reading of the Realist Commitments – evaluation of them still involves taking a moral stance, which on the current understanding is a set of hybrid states composed of desire-like states and robust beliefs. Let me begin with the distinction between correlation and reduction.

Pleasure-Correlation: things are good if and only if they are pleasurable.

$State_{P-C}$: [H! (Things with property K) & belief that things have that property if and only if they are pleasurable]

Pleasure-Reduction: goodness reduces to being pleasurable.

$State_{P-R}$: [H! (Things with property K) & belief that that property reduces to being pleasurable]

These mental states are clearly distinct. An agent may coherently judge *Pleasure-Correlation* while rejecting *Pleasure-Reduction*. Here is an example of such an agent realising those judgements:

Realiser of $State_{P-C}$: [H! (Things that are loved by God) & belief that things have that property if and only if they are pleasurable]

Realiser of $\neg State_{P-R}$: [H! (Things that are loved by God) & belief that that property does not reduce to being pleasurable]

To put it in less formal terms, this agent has a moral perspective that effectively reduces goodness to what God loves and has a theological belief that God loves all and only things that are pleasurable. Being pleasurable,

our pro- and con- attitudes towards. As we discussed in §6 of the previous chapter, we must include the *totality* of the agent's relevant attitudes. I list only one for simplicity.

then, is not the *same* property as goodness but it does correlate with it, at least according to this agent. In this way we can make sense of people who accept reductive theses. Here are two potential mental states of agents who believe *Pleasure-Reduction*.

Realiser 1 of State_{P-R}: [H! (Things that are pleasurable) & belief that that property reduces to being pleasurable]

Realiser 2 of State_{P-R}: [H! (Things that are loved by God) & belief that that property reduces to being pleasurable]

The second agent, *Realiser 2*, has a false robust belief. The property of being loved by God is not reducible to the property of being pleasurable. I include it because people have mistaken beliefs all the time. Furthermore, the truth or falsity of the robust belief component does not tell us whether or not the moral judgement as a whole is true or false. In this case, despite the first agent's correct robust belief component, *both* agents have a false moral belief – they believe goodness reduces to being pleasurable, and it doesn't. But that is my assessment as a moraliser, not as an expressivist! I could be wrong, in which case the second agent has a correct moral belief realised in part by an erroneous robust belief. This isn't as counterintuitive as it might sound. It just means that sometimes we form the correct moral opinion on the basis of a false non-moral belief, which does happen.

We have now seen sensible mental states that constitute accepting and rejecting the reduction of goodness to the property of being pleasurable. We have seen someone who rejects the reduction of goodness to pleasure while still thinking there is a correlation between them. We can similarly distinguish between, on the one hand, a subjectivist who effectively thinks moral properties are constituted by what their ideal self will approve of and who believes their ideal self will approve of a certain set of deontological principles, and on the other, a deontologist who thinks moral principles *just are* these deontological principles.

To summarise, on the hybrid view each moral evaluator will have desire-like states that determine which properties they treat as the moral ones. Robust beliefs ascribing those properties to things will partly constitute judgements ascribing moral properties to things, and judgements *about* moral properties will be constituted in part by robust beliefs about those properties. In this way the Realist Commitments may be vindicated as sensible things to believe. One sensible way to believe *Non-Naturalism*

would be to approve of things with the property of being loved by God and to believe discovering what God loves is not a scientific matter. One sensible way to believe *Mind-Independence* would be to approve of things insofar as they are prescribed by all of a given set of moral theories (hedonistic and preference utilitarianism, say, or all theories that neither prescribe theft nor torture) and to believe that property is not constitutively dependent on the actual or hypothetical responses of agents. Moral metaphysics can go on.

3. The Reducibility of Moral Properties

In this section, and using the framework I have just outlined, I will argue that if moral properties exist then they are reducible to properties we can describe in non-moral terms. This means expressivists cannot “mimic” the robust or relaxed realist positions because those realists believe the thesis of *Non-Reductivism*. Instead, expressivists must accept positions (internally rendered) that mirror such theories as natural realism, subjectivism, constructivism, or even error theory. To see why, let us begin by considering what it would be to accept *Non-Reductivism*.

Non-Reductivism: moral properties are not reducible to descriptive properties.

State_{N-R}: [H! (Actions with property K) & B! (Actions with property L) & belief that those properties are not reducible to descriptive properties]¹

Allow me to clarify what is meant by ‘descriptive properties.’ “A property is descriptive if and only if it can be ascribed with a descriptive predicate”² and a descriptive predicate is a predicate that contains no normative terms. The predicate ‘is yellow’ is descriptive and so the property of being yellow is a descriptive one. Robust and relaxed realists believe that moral properties are *not* descriptive, and so the only way to refer to them directly is via normative terms (or normative concepts) such as ‘is right’ or ‘the property of wrongness.’ On the other hand, a natural reductivist who believes goodness is the property of being pleasurable should think

¹ If there are more than two moral attitudes, these should also be included. For brevity I include only H! and B! for actions, corresponding to judgements of right and wrong.

² Streumer (2017: 3).

goodness is a descriptive property, since goodness could then be referred to in purely non-normative terms such as 'is pleasurable.'

When it comes to indirect reference, things are a little messier. Is the phrase 'what she said' normative? On the face of it this is a descriptive phrase – in theory someone with no normative concepts could understand and use it. Consider an example. Morpheus does not hear what the oracle says to Neo, but he trusts her and he assures Neo that what she said is true. Did Morpheus make a normative claim? If the oracle told Neo it is morally wrong to torture innocents for fun then Morpheus' claim is made true by a normative fact, and the phrase 'what she said' refers in this context to a normative proposition. Given that Morpheus may not have had or employed normative concepts in his judgement that what the oracle said is true, what does this mean for us? It means the descriptive/non-descriptive distinction is best understood in terms of direct rather than indirect reference.³ Morpheus might have used no obvious normative terms in his indirect claim, but since *the oracle* referred directly to the normative fact using normative terms then it is still an open question whether or not normative facts can be referred to *directly* in non-normative terms.

Robust and relaxed realists believe moral properties cannot be referred to directly in non-normative terms. If Enoch (2011) is right about robustly real moral properties, I could refer to them indirectly with the ostensibly non-normative phrase 'the properties Enoch describes and argues for in his book.' This will be useful for later discussion. Of course, since Enoch is *not* right (we are assuming), then the phrase does not refer to any robustly real moral properties because there aren't any such things to refer to. The kinds of moral properties Enoch argues exist are the type we can form robust moral beliefs about without having any particular motivational states, which would imply expressivism is false. Since expressivism is true (we are assuming), the properties Enoch describes and argues for in his book do not

³ I do not mean to get into debates about descriptivist versus causal theories of reference here. In the senses I'm currently employing, a direct reference "travels" straight from the use of a term or concept to the thing that is referred to, whereas an indirect reference "goes through" another referrer. So in this sense, a description could count as a direct reference if it makes no mention of other references: 'the sixth castle ever built' or 'whichever properties caused the magnesium to ignite.' Examples of indirect reference in my intended sense include 'what she said' and 'the property Sharon was thinking about last night.' Perhaps we could call direct and indirect reference 'immediate' and 'mediate' reference instead, but I have chosen language that seems more natural to my ear.

exist. We are now in a position to analyse *State*_{N-R}. What possible realisers can we find?

Realiser 1 of *State*_{N-R}: [H! (Actions with the property of maximising pleasure) & B! (Actions with the property of failing to maximise pleasure) & belief that those properties are not reducible to descriptive properties]

An agent who had the above mental state would thereby believe *Non-Reductivism* to be true. Unfortunately they have a false robust belief, since the properties of maximising and failing to maximise pleasure can be referred to in non-normative terms. Quite clearly, putting *any* descriptive property in the places of K and L is going to result in a false robust belief component. Is there a way to put irreducibly normative properties in there instead? Perhaps we could try something like this:

Realiser 2 of *State*_{N-R}: [H! (Actions that have the irreducibly normative property of rightness) & B! (Actions that have the irreducibly normative property of wrongness) & belief that those properties are not reducible to descriptive properties]

Similarly, we might try to slot in some indirect references, such as desire-like states directed towards actions that have the properties Enoch describes and argues for in his book. The question now is whether we have managed to find a true robust belief component in Realiser 2 of the acceptance of *Non-Reductivism*. If we have, then an expressivist could accept all of the Realist Commitments. *Prima facie* it does seem true that the moral properties Enoch argues for are irreducible. However, once we dig a little deeper, we see we have a failure of reference in the belief component – *those properties* or *the properties Enoch describes and argues for in his book* do not exist (or else expressivism would be false and this exercise in metaphysical sidesteps was on the wrong track from the off), and so anything we say *about* them will miss the mark. Is it true that the present King of France is bald? Clearly not, whether or not we want to say it is *false*. In the same way, the robust belief component of Realiser 2 is not true.

But don't we all agree that the properties Enoch argues for are irreducible? We must be careful in these waters. When an anti-robust-realist says that the moral properties Enoch argues for are irreducibly normative, she is not best understood as literally claiming something about robustly real properties. Instead we could interpret her as either making a

conditional claim such as *if Enoch is right* then moral properties are irreducible, or as a claim about Enoch's position: Enoch (rightly or wrongly) describes moral properties as being irreducibly normative. What we should not do is interpret her as claiming the properties Enoch argues for *actually exist* and are irreducible. Importantly, we have failed to find a realiser of *State_{N-R}* that contains a true robust belief.

We have effectively discovered a transcendental argument against *Non-Reductivism*. Rather than looking at direct arguments against irreducibly normative properties such as arguments from supervenience or queerness, we have been examining the mental states involved in accepting *Non-Reductivism* and determining what we as expressivists are committed to. Pure expressivists could not make sense of it at all, but we have seen that hybrid expressivists can characterise acceptance of *Non-Reductivism* with a mental state that includes a robust belief component about the irreducibility of certain properties. Then comes the dilemma: are the 'certain properties' actually irreducible? If the properties exist but are not irreducible, then the robust belief component is straightforwardly false. But if the robust belief is about the irreducibility of, say, the kind of properties Enoch argues for, then the 'certain properties' do not exist, for their existence is incompatible with expressivism – if they existed they could be the objects of robust moral beliefs absent any motivational states – hence, the robust belief component is not true. Therefore, on the hybrid expressivist picture, there is no way for *Non-Reductivism* to be accepted without having a robust belief that is not true. Now, we have already noted how the fact that a moral judgement rests on or is constituted by a false robust belief does not guarantee that the judgement itself is false. However, once a belief of yours has been *revealed* to you to be untrue, it is difficult if not impossible to continue believing it. Given that we should reject beliefs that have been shown not to be true, no expressivist should accept *Non-Reductivism* after the claim has been made sense of. Expressivists, then, should not aim to “mimic” robust or relaxed realism. They should not accept all of the Realist Commitments.

4. Metaphysical Options for Expressivists

What classical metaethical positions *can* expressivists mimic? In the previous section we saw why robust and relaxed realism are off the table. Still, every other position seems available by either making tweaks to the

desire-like states, or by accepting the consequences of “wanting more” from moral reality than can be found: error theory. Let us begin with it.

4.1 Error Theory

If we define error theory minimally as the view that there are no (and cannot be any) substantive moral truths, then we can realise this position in many ways, such as [H! (Actions that maximise pleasure) & B! (Actions that fail to maximise pleasure) & belief that those properties are never realised]. Why would someone, in this example with a utilitarian perspective, have that robust belief? Perhaps they think there are no actions because all action is impossible due to physical determinism. The reason doesn't matter – as long as they believe there are no substantive moral truths, they are error theorists. A different agent may base their moral foundation on what agents would agree to under ideal circumstances and become pessimistic that any such agreement can be found. Yet another agent may have a theological moral perspective, thinking that the only thing that morally matters is the will of God, and then come to believe God does not exist, perhaps causing him to enter a state of moral crisis. I would hope this new atheist would eventually morally approve of the instantiation of more down-to-earth properties, but he need not.

Error theory is more often seen as the view that *were there moral properties* they would be irreducibly normative, but nothing is irreducibly normative so there are no moral properties. On the hybrid expressivist picture, this can be interpreted as something like the expression of a normative perspective linked to the instantiation of properties as described by robust realists, along with the robust belief that there are no such properties. Error theorists agree with me *robustly* metaphysically (i.e. about the metaphysical topics we can form robust beliefs about), but not morally metaphysically. Morally speaking they *want more* from the world than I do, and they want more than it can deliver. Whereas I morally desire the instantiation of descriptive properties such as being conducive to welfare, they want categorical direction from irreducibly normative properties; but, as both the error theorist and I agree, such properties do not exist. Due to the internal nature of moral metaphysics this is the type of argument that can be won with moral persuasion – appeal to their moral sentiments, change the kind of things they morally care about, and bog-standard

descriptive properties may become enough to ground morality in their eyes.

An objection to this interpretation of the thoughts of the error theorist is that one simply *cannot* approve or disapprove of the instantiation of properties one believes cannot be instantiated. It would be like loving all who draw square circles, and that's absurd. In reply I would point to the previous potential realisers of an error theorist. The atheist with a theological moral foundation may think that God's existence is impossible (perhaps because God would have to be a necessary being which, when combined with God's lack of actual existence, implies no such necessary being is possible); the utilitarian free will sceptic may be a hard incompatibilist. These seem like perfectly comprehensible views to me. The objection might be rooted in the conflation of epistemic and metaphysical modality. In the case of the absurd square-circle lover, it is plausible that it is not just metaphysically but also epistemically impossible to draw a square circle. It might be strange to have desire-like states towards what you take to be *epistemically* impossible, but what we are concerned about are metaphysical impossibilities. I may wish that God existed and loved me despite thinking it is metaphysically impossible so long as I recognise I may be wrong. Likewise, while the error theorist believes robustly real moral properties are metaphysically impossible, she is not 100% certain of that. In short, one can consider epistemically possible worlds and take up moral standpoints towards the goings on in those worlds without believing them to be metaphysically possible.

4.2 Reductive Naturalism

Mimicking error theory is an open option then, but it is unlikely to attract expressivists. Expressivists are usually drawn to their position via its promise of *vindicating* their moral judgements as generally (or at least possibly) *true* on a naturalistically respectable foundation. Other options include subjectivism, contractualism and constructivism, where moral values are something a little more human and mind-dependent. People with these metaphysical views might realise the judgement 'I should donate' as follows:

[H! (Actions I would desire to desire under full imaginative acquaintance) & belief that donating has that property]¹

[H! (Actions allowed by all sets of rules for regulation of behaviour that no one could reasonably reject as a basis for informed, unforced agreement) & belief that donating has that property]²

[H! (Actions entailed from my practical point of view according to the rules of agency) & belief that donating has that property]³

We can also have mind-independent forms of naturalism, of course. The property of rightness could simply be the property of causing at least as much pleasure as all other available options, or of being mandated by the categorical imperative, or any other bog-standard, run of the mill descriptive property. (Without making a separate category, we can understand reductive non-naturalism pretty easily by substituting in a descriptive non-natural property, such as being favoured by the gods, in the relevant place.) The debate between those metaethicists who believe goodness reduces to pleasure and those who believe it reduces to desire satisfaction is to be understood as a moral one. Since it is to do with the nature of moral properties I am happy to call it metaphysics, but I would emphasise that this particular metaphysical debate is internal to moral inquiry and we should expect the interlocutors to draw upon moral considerations in support of their respective reductive views. Indeed, I would find it very difficult to imagine the debate being conducted *without* moral argument.

4.3 Non-Reductive Naturalism

The difficult case is that of non-reductive naturalism. Whether or not an expressivist can “mimic” this depends on the form it takes, particularly the kind of reduction that moral properties are supposed to resist. Take the position advanced by Ralph Wedgwood (2007, especially ch.8 and ch.9). Wedgwood argues for normative accounts of intentional mental states (which we encountered in §1.1 of chapter 1) and an account of normative

¹ This is an “internally rendered” view inspired by Lewis (1989).

² Inspired by Scanlon (1998: 153).

³ Inspired by Street (2010). See Lenman (2012) and Ridge (2017) for detailed discussion on the potential benefits of pairing constructivism with expressivism.

properties that makes essential reference to intentional mental states: for instance, what ought to be done is what is correct to incorporate into one's ideal plans. On his view of normative properties and intentional mental properties, it is not the case that one particular set is reducible to the other but instead the two sets are interdependent, such that their potential for natural reduction stands and falls together. After arguing that mental states are not reducible to descriptive ones, Wedgwood concludes that normative properties are irreducible too. He also believes them to be causally efficacious and natural.

According to Wedgwood's variety of non-reductive naturalism, moral properties cannot be directly referred to without using normative terms, which is just to say that *Non-Reductivism* is true. Since I have shown *Non-Reductivism* to be incompatible with expressivism, expressivists should not "mimic" Wedgwood's view. Expressivists therefore cannot subscribe to his (fortunately unpopular) combination of views about the constitutive interdependence of the normative and the mental in addition with the irreducibility of the mental to the natural. This is an interesting but intuitive result. It means if an expressivist is attracted to normative theories of the mind (Gregory 2012) she is committed to reducing normative properties to ones that can be referred to with non-mental terms, so that they can be specified purely descriptively. A further interesting result of having to deny *Non-Reductivism* is the implication that normative theories of the mind are also descriptive theories of the mind.

Another form of non-reductive naturalism that is clearly off the table for expressivists is Ryan Stringer's (2018) "Emergentist Ethical Naturalism." Stringer (*ibid*: 340-1) contrasts his view with "*identity-naturalism*" (which we have called 'reductive naturalism'), with "*constitution-naturalism*" (which he recognises as the usual intended referent of the term 'non-reductive naturalism' and includes the Cornell realists Sturgeon (1985), Boyd (1988) and Brink (1989) to be addressed in the next paragraph), and plain old non-naturalism. Stringer believes moral properties *emerge* from their natural supervenience bases, are causally efficacious, and are *sui generis* – that is, they are not reducible to bog-standard descriptive properties. Like Wedgwood's view this implies *Non-Reductivism* and is therefore incompatible with expressivism.

Finally there are the views of the Cornell realists. Luis Oliveira and Timothy Perrinne (2017) define Cornell realism as the view that for "any

moral term M, M picks out an ethical natural property that supervenes on, without being reducible to, some distinct non-ethical natural property (or properties) N.”⁴ As we will see in the next section, Blackburn and I feel a strong affinity to the Cornell realists, so it is important to know whether it is a view eligible for us to mimic. Unsurprisingly it all hinges on whether the natural properties taken to be the moral ones are descriptive. And in some cases they clearly are. In Richard Boyd’s famous (1988) paper he argues for “homeostatic consequentialism” whereby goodness consists in the satisfaction of human needs. Boyd (1988: 203-4) urges there will be a “homeostatic” cluster of properties (properties that non-accidentally tend to be found together) that satisfy our needs, but nevertheless these properties will be specifiable with non-moral terms. This kind of view does not commit us to *Non-Reductivism*, which is what we want.

Why does Cornell realism fall under the non-reductive label? Perhaps it is because of their proponents’ emphasis that moral truths are synthetic (i.e. not analytic, not true in virtue of meaning) and *a posteriori* (i.e. knowable only through observation or experience). This could be seen as a kind of semantic and epistemological non-reductivism if not a metaphysical one.⁵ But there are Cornell realist views that may be seen as metaphysically non-reductive too, depending on various contentious metaphysical issues. These views are forms of *functionalism*, which Wedgwood (2007) rejects as both a theory of the mind and a theory of moral properties. Moral functionalists (and functionalists about the mind) have come in analytic (Jackson and Pettit 1995, who therefore do not subscribe to Cornell realism) and synthetic (Sturgeon 2003) varieties, but that distinction will not be important in this section because we are discussing the nature of the properties rather than how we know about them.

Functionalism, most often explicated as a theory of the mind, is the view that what constitutes being a certain state or having a certain property is a matter of playing a certain functional role in a system of related states or properties. What makes mental state D a desire to dance is not a matter

⁴ Oliveira and Perrinne (2017: 1025). As phrased, this doesn’t rule out a straightforward reductive naturalism: if ‘right’ picks out the property of maximising pleasure, this is a natural property that presumably supervenes on various physical facts but may not reduce to those physical facts.

⁵ This is supported by textual evidence from Boyd (2003: 26), who writes there is “no prospect for turning the *definitions* provided by a naturalistic account of reference for terms within a domain into a *reductionist* treatment of the *discourse* within that domain” (with the first and third emphasis added by me).

of what D is made of but the way it relates to other mental states, behaviour and external stimuli. Likewise, according to moral functionalists the most natural way of describing a moral property is with reference to other moral properties. Here's a toy example: *right* actions are those that we *ought* to perform; that we have most *reason* to perform; that *virtuous* people tend to perform; and that generally yield *good* consequences. If this is the most illuminating characterisation of rightness, what is notable is the reference it makes to other moral properties, referred to in italics. Suppose further that this is the case for every moral property. Does this mean moral properties are not bog-standard descriptive properties? Not necessarily, due to a process called Ramsification (described in Railton 1993: 47-50 and Jackson 1998: ch.6). It works as follows. Take the whole network of interdefined moral properties, replace all the moral words with variables, existentially quantify over them with uniqueness, and then we end up with purely descriptive specifications: there exists exactly one of each of the properties/relations *v*, *w*, *x*, *y* and *z* such that actions with *v* are those that stand in relations *w* and *x* to our performing them; people with *y* tend to perform actions with *v*; and actions with *v* tend to yield outcomes with *z*. Rightness can be descriptively specified with reference to *v*.

As with role- and realiser-functionalism in the philosophy of mind, so with moral functionalism: "it leaves open whether rightness, say, is the ground-level (descriptive) property that occupies the rightness role, e.g. the property of maximising happiness, or whether it is the higher-order property of having a property that occupies the rightness role."⁶ If rightness is the ground-level realiser, e.g. the property of maximising happiness, then expressivists can straightforwardly accept this by directing their moral attitudes towards the ground-level realisers of the functional roles. If it's the higher-order property, however, then we need one final assumption – that the functional role is not merely a necessary connection to the moral property but essentially constitutive of it. Here's why.

A robust realist could accept the descriptive Ramsification of moral properties by holding that it is not a descriptive property (e.g. maximising happiness) that plays the role of property *v* but it is the *sui generis* property of rightness. "Sure," says this robust realist, "there is a necessary connection between rightness and the functional role you described, but that doesn't count for much – it was only achieved with your cheap semantic trick of

⁶ Jackson and Pettit (1995: 27-8).

substituting all the normative terms with variables.” Necessary connections are not enough for reductions, as illustrated by the following example: x is the property of being even if and only if x is the property I’m actually thinking about as I type this sentence. Given that I was thinking about the property of being even as I typed that sentence, the biconditional is necessarily true (read *de re*), but this is hardly a reduction of evenness to mental properties. Wedgwood (2007: 146) clarifies the idea: “for a normative property to be naturalistically reducible, it must not only be necessarily equivalent to some naturalistically specifiable property, but this necessary equivalence must also follow from the constitutive essence of the normative property.” Thus, the final condition is that the *essence* of rightness is that of occupying the descriptively specifiable functional role. The descriptive specification can therefore be referred to in our desire-like moral attitudes and the robust beliefs that partly constitute moral judgements. Expressivists can mimic higher-order moral functionalism too.

My working assumption is that expressivism is true. In §1 I argued that moral metaphysics should go on. In §2 I showed how it *could* go on. In §3 I revealed a commitment to rejecting *Non-Reductivism*, taking robust and relaxed realism off the table as potential positions to mimic. Yet in this section I have shown expressivists still have several viable metaphysical options. In the next and final section I plump for one of these options: the synthetic moral naturalism of the Cornell realists.

5. For They Shall be Called the Children of God

In his (2015) paper, ‘Blessed are the peacemakers,’ Blackburn presents us with some major points of similarity of his expressivist view with Cornell realism and some remarks on where they diverge. I agree with Blackburn that expressivists can – and in my first-order ethical view *should* – feel very sympathetic towards Cornell realism, and indeed it is the kind of realism I think expressivists should “mimic” with their quasi-realism. In this section I argue that the two positions are (or at least can be pushed) closer than Blackburn allowed. This is because his main complaint against Cornell realism is that it is purely cognitivist and detached from moral motivation, giving no integral place to the kinds of conative attitudes that seem part and parcel of moral thought and discourse. However, these kinds of attitudes can easily be incorporated by going hybrid. Blackburn’s other complaint is

with the causal theory of reference for moral terms,¹ but this is not an essential part of the Cornell realist picture and it has indeed been dropped by those who describe themselves under this banner.² In what follows I outline what I take to be the correct moral metaphysics and in doing so I make as much peace with the Cornell realists as possible. I then identify the real point of departure with Cornell realism. I argue it tips the balance in favour of expressivism.

5.1 Cornell Realism

In §4.3 I discussed forms of non-reductive naturalism and noted that we can distinguish between semantic and metaphysical non-reductivism. Due to the transcendental argument against *Non-Reductivism* we cannot accept any metaphysically non-reductive views. Cornell realism is best seen as a *semantically* non-reductive form of naturalism. Let me begin with some major points of agreement between its defenders and certain expressivists. Blackburn (2015: 843-44) begins with these three:

1. Moral sentences are truth-apt.
2. The moral truth does not depend on our opinions and theories.
3. We are not hopeless at finding the moral truth.

He then writes something (*ibid*: 844) that at first blush seems peculiar: “I avoid worrying about ‘moral properties’ or ‘normative properties’ altogether, believing myself to inhabit a world in which there are only natural properties, including ones to which we often have moral and evaluative attitudes.” Is this a momentary lapse of quasi-realism for Blackburn, admitting that he doesn’t think there are moral properties *really*? No, for he only says he doesn’t *worry* about them. What he does say is “there are only natural properties,” which – given his prior commitments to the existence of moral properties – implies that moral properties are natural. We can add that to the list. Furthermore, due to Moorean open question considerations, Blackburn and I agree with the Cornell realists that moral terms and concepts do not reduce to bog-standard descriptive ones. Moral truths are not analytic – not true merely in virtue of their meaning, such as

¹ See Sturgeon (1985: 55) and Boyd (1988: 195).

² Brink (2001: 170-76).

(perhaps) the truth that bachelors are unmarried – but instead are synthetic, such as the truth that water is H₂O.³ The list grows:

4. Moral properties are reducible to descriptive, natural properties.
5. Moral truths are synthetic.

The main difference between Blackburn and the Cornell realists is their *focus*. Whereas the Cornell realists are concerned with notions such as truth, reference and methodological observation, Blackburn directs his attention to the kinds of moral attitudes we take up and how they steer our behaviour and practice: a society's morality "is located by means of their choices, their comfort zones and their surrounding causes of discomfort, their practical setting of boundaries and permissions, penalties, tolerations, sources of shame and pride, resentment and gratitude, as well as the way they look on (and distort) their history and the symbols it has left them."⁴ We might put it this way: on metaphysical matters Blackburn and the Cornell realists agree, but on psychological matters they do not. What Blackburn thinks is missing from the Cornell picture is an account of the pushes and pulls of the conative attitudes we expressivists take to be integral to moral thought. But this can easily be added.

5.2 Making Peace over LUNCH

Nicholas Laskowski's (2017) Unanalyzable Normative Concept Hybridism, LUNCH, is a hybrid form of Cornell realism. Via its hybridism it is able to overcome Blackburn's qualms. The theory is compatible with claims 1 to 5 in the previous subsection, but it adds that moral judgements also involve desire-like components. Laskowski leaves open how to characterise the desires but plausibly they could be of the sort that drives our choices, discomforts and all the other things Blackburn looks for in moral practice. Problem solved. If it is that easy, why didn't Blackburn consider it? I can only guess it is because in his (2015) paper he was responding specifically to Boyd's (1988) paper which appeared long before the literature on hybrid realism. Nevertheless, the point is that going hybrid addresses Blackburn's worries. We may detect what people consider to be wrong by looking at what they avoid, or feel guilty for doing, or get angry with others about and

³ "[I]t is a mistake to look for analysis and definitions when we try to locate the cluster of [moral] properties" (Blackburn 2015: 849).

⁴ Blackburn (2015: 849).

so on, because to judge something wrong is in part to have desire-like states that tend to cause those kinds of feelings and behaviours.

One of the most important things that Laskowski draws attention to is the conceptual landscape of moral thought. So far in this thesis and (I dare say) in the literature on hybrid expressivism, we have played fast and loose with the mental states, especially the concepts in the robust belief. We have been clear about what is believed but there are many ways in which to believe something. Alice, Bob and Carla may all believe *de re* that Dot is drinking water, but they may do so in different *de dicto* ways: Alice believes *Dot is drinking H₂O*, Bob believes *my sister is drinking water*, and Carla believes *that person is drinking my favourite drink*. We might describe this as a case where Alice, Bob and Carla all believe something which is made true by the same fact (the fact that Dot is drinking water), but each of them deploys different concepts. Hybrid expressivists have not always been very explicit about the details of moral concepts.

Consider the following hybrid naturalist theory. The property of rightness is the property of maximising happiness, and to judge X right is to believe *X maximises happiness* as well as to desire to do what maximises happiness. This theory may boast an advantage over analytic naturalism when it comes to meaning and Moore's open question: 'helping is right' doesn't *mean* the same as 'helping maximises happiness' because one can accept the latter without accepting the former, which requires a desire. Consequently, the question 'helping maximises happiness, but is it right?' feels open. Unfortunately, an issue arises in the reverse direction. According to this theory, anyone who judges helping to be right *already* believes *helping maximises happiness*, so the question 'helping is right, but does it maximise happiness?' is closed. This seems false. Furthermore, it is plausible that we learn moral concepts before we learn the concept of *maximising happiness*, which would be impossible according to this version of hybrid naturalism. The problem clearly generalises beyond *maximising happiness*. In summary, it is implausible that moral concepts, such as the concept of *rightness*, are composed of familiar descriptive concepts like the concept of *maximising happiness*.

This motivates a challenge to my favoured theory, higher-order expressivism. The robust beliefs I have been describing as composing the higher-order states make reference to 'that property' or 'those properties' picked out by the moral perspective. What concepts are deployed in this

belief? Suppose a particular agent has a utilitarian moral perspective and approves of actions insofar as they maximise happiness. When she judges that helping is right she believes *de re* that helping maximises happiness, but for the reasons given in the previous paragraph it is implausible that this is the *de dicto* belief that *helping maximises happiness*. Similarly, it is implausible that it is the belief that *helping has the property I morally approve of actions for having* – it seems possible to lack this *de dicto* belief while believing helping is right. What if we stick with the demonstrative?⁵ Could it be the *de dicto* belief that *helping has those properties*? I find it hard to imagine we use the same concept expressed by terms like ‘those properties’ in the relevant belief. When we use demonstratives such as ‘that gerbil is bigger than that one,’ we use them intentionally and we make it obvious what we’re referring to, perhaps by pointing. Nothing like this – no pointing, no conversational cue – seems to be going on when we make moral judgements. To clarify, I am on board with the idea of the relevant concepts picking out those properties without describing them, which is what demonstratives (and perhaps proper names) do. What I am unsure of is that we use the concept of *those properties* in our moral beliefs.⁶

Perhaps a good case can be made for the possibilities in the previous paragraph. If it can, that’s great! I am happy to accept anything that works to ensure the content of the robust beliefs matches up correctly with moral perspectives. But given my current scepticism about the above concepts, I want to take a leaf from Laskowski’s (2017) paper and claim the relevant concept is unanalysable. As mentioned at the beginning of this subsection, LUNCH is a form of Cornell realism, a synthetic (rather than analytic) natural realism. Suppose once more that the property of rightness is the property of maximising happiness. This is not true in virtue of meaning but something to be discovered *a posteriori*, like discovering our term ‘water’ refers to H₂O. According to LUNCH, the judgement that helping is right does not involve the concept of *maximising happiness* but an unanalysable concept C. (C is not the concept of *rightness*.) The judgement is composed of

⁵ Schroeder (2009: 280-95) interprets Ridge’s (2006a) view as one that employs demonstratives in the robust belief. Schroeder borrows Kaplan’s (1989) “dthat” operator. Adjusting this idea to the judgement we are considering, the robust belief component would be the belief that *helping has dthat(the property I morally approve of actions for having)*.

⁶ The reverse open question is confusing: ‘helping is right, but does it have those properties?’ One natural response to that question is ‘which properties do you mean?’ Given that ‘those’ can refer to anything, and given that what we want is something that refers *consistently* to whatever is picked out by our moral perspective, it seems as though ordinary demonstratives like ‘those’ won’t do the job.

the robust *de dicto* belief that *helping is C* along with the desire to do whatever is C. LUNCH therefore does not suffer from Moore's open question argument or its reverse. I propose that higher-order expressivists make the same move: the robust beliefs composing moral judgements employ unanalysable concepts.

5.2.1 Waging War over LUNCH

Like Blackburn (2015: 844, 849) and Gibbard (2012: 33), I want to join forces with the Cornell realists in saying moral properties are natural and moral truths are synthetic. Blackburn was unimpressed with the lack of feeling to be found in early Cornell realism, but there is nothing to stop its defenders from going hybrid like Laskowski does. Laskowski has also taught us a valuable lesson – just as Boyd (1988) and Gibbard (2003: ch.2) agree with Moore that the concept of 'good' is unanalysable, hybrid expressivists too can make use of unanalysable concepts in their theories of moral judgement.

As much as I hate to argue against LUNCH, there is a clear point of departure. Laskowski believes concept *C* refers to the same descriptive mind-independent property, such as maximising happiness, in the minds of everyone. Furthermore, he thinks this reference is secured via "standing in the appropriate *causal* relationship to the property of [maximising happiness] that actions instantiate."⁷ There is a solid history of attacking causal theories of reference for mind-independent natural realism, notably Hogan and Timmons' (1991, 1992a, 1992b, 1996, 2015a) papers stemming from Hare (1952: 148). I will not rehearse these arguments in detail but Blackburn puts what is essentially the same point as follows, assuming with Boyd that the correct account of moral properties ends up entailing a form of consequentialism.

[C]ontemplate a society such as our own in which a value of personal autonomy is rather firmly entrenched. A person might, for instance refuse to sacrifice their own conception of their right to live the life they wish even if promoting the social good would require it. They refuse because, they say, they have no obligation to give up their personal projects to become 'servants of the world', or to do what they find distasteful or difficult for the sake of widely but thinly distributed benefits to others. And suppose we find this

⁷ Laskowski (2017: 15).

attitude is almost universally understood, tolerated, or admired and encouraged. Perhaps indeed this is our own society: it is not only that we behave in the light of the value of autonomy, but enough people seem to have thought that Williams's famous cases refuted consequentialism to suggest that we give it that priority even on reflection. It seems to me to require a rather peculiar hermeneutic to interpret people who feel like this, as *really* regulating their use of the term 'good' by the very property that in their actions, doings, and sayings, they overtly reject. (Blackburn 2015: 849.)

The idea is that it is utterly implausible that everyone's moral concepts are causally regulated by the same mind-independent natural properties. This is why, according to higher-order expressivism, the concept in the robust belief component refers to something relative to the thinker's moral perspective. Fortunately this does not commit us to moral relativism because we do not link the truth of what people say to the truth of their robust belief component. However, for cognitivists the content of moral claims is always the same as the content of the robust belief expressed, and so to remain a form of realism *they* must maintain that the robust beliefs composing people's moral judgements always refer to the same mind-independent properties.

The problem with LUNCH and the versions of Cornell realism Blackburn attacks is the causal theory of reference. It is much more plausible that our deployment of moral concepts is causally regulated by our moral perspectives (which vary greatly) than some particular mind-independent natural properties. The causal theory of reference for moral terms is therefore difficult to square with realism. But this problematic theory of reference does not seem to be essential to the view and indeed has been dropped by some Cornell realists. In this next subsection we will tackle the issue of reference, bringing expressivism and Cornell realism ever closer to a peace treaty.

5.3 Normatively Enriched Moral Metasemantics

Expressivists have not provided a theory of reference for moral terms. They explicitly reject causal theories of reference but they have offered nothing in its place. Presumably this is because they think they don't need to. But if there are moral properties (which I have argued must be descriptive), it

seems sensible to think that our moral terms refer to those properties. This raises the subject of the mechanism via which they do so. Given that a consequentialist and a deontologist refer to the same property when they use the word 'right' despite their uses of the term being causally regulated by different properties, what is it that fixes the term's reference? Causal theories are ruled out. We should also rule out "interpretational" theories according to which, for example, the content of our beliefs is whatever successfully predicts and explains our behaviour (Davidson 1973, Lewis 1974, Dennett 1987), and for related reasons: we expect the consequentialist and deontologist to behave differently when they judge something to be right because their use of the term 'right' is causally regulated by different properties. Similarly, functional theories (Harman 1973, Block 1986) won't do the trick; they aver that mental states represent what they do in virtue of causal, computational and inferential relations to other mental states, and it seems equally evident these relations will differ for the consequentialist and deontologist. Teleological theories (Millikan 1984) place importance on a natural function such as the biological function that was selected for by evolution, but this is unattractive because we would not want the correct referents of moral terms to be beholden to what is beneficial for our genes.

I realise the above paragraph is sketchy. All the same, I think an intuitive lesson can be drawn from it: we don't want to "outsource" ethics to non-normative domains. Let us do the expressivist thing and look at the problem of reference side-on. What is it to *judge* that, say, 'right' refers to the property of maximising happiness? It is to take a stand on moral issues. One could not think 'right' refers to maximising happiness without any endorsement of utilitarianism. In short, theorising about the reference of moral terms is internal to moral inquiry. The dominant theories of reference all can be deployed "scientifically" without any engagement of the moral perspective, and this is precisely why they fail. We should therefore turn to *normative* theories of reference. And far from this separating expressivism from cold, detached cognitive naturalism, Cornell realists have been here before us.

In his (2015) paper 'Normatively Enriched Moral Meta-Semantics,' Michael Rubin discusses (and argues against) two theories of reference to natural moral properties proposed by David Brink (2001) and Mark van Roojen (2006). Brink's theory is couched in terms of properties that play a role in interpersonal justification – roughly, that we use moral terms with the intention of picking out properties that can justify our actions – and "a

predicate like ‘right’ picks out whichever property *in fact* plays the relevant role with respect to interpersonal justification.”⁸ Perhaps the property of maximising happiness is what in fact justifies our actions (despite what most people think) – a normative claim – and so ‘rightness’ refers to maximising happiness. Van Roojen’s theory is in terms of what “makes sense to do”⁹ whereby ‘right’ picks out whatever property *in fact* explains why it makes sense to perform a particular action. Brink and van Roojen understand “justification” and “making sense” in a non-moral but still normative sense. However, as Rubin argues, this only shifts the bump along the rug and leaves us wondering why everyone’s normative terms (such as ‘justification’ and ‘making sense’) refer non-analytically to exactly the same natural and mind-independent properties. So whether or not we enrich our theory of reference morally (as I suggested in the previous paragraph) or more broadly normatively, Rubin’s (2015: 401-3) challenge arises: that of avoiding vicious circularity.

For simplicity, let us only consider my own morally enriched theory of reference for moral terms whereby to think ‘M’ refers to *p* is to take up a moral stance, and what makes it the case that ‘M’ *actually* refers to *p* is that this moral stance is the correct one. Suppose I’m a utilitarian. “What does ‘right’ refer to,” you might ask me, “even in the mouths of deontologists?” The property of maximising happiness, I reply. “Why does it so refer?” Because utilitarianism is the true moral theory. The trouble is that, in order to know that we are both talking about the *moral* ‘right’ rather than, say, the direction ‘right’, we have to first know the semantic facts. We have to know we are using ‘right’ in the moral sense and moreover the *same* sense as other apparent moralisers. For now suppose Virgil is, or seems to be, a virtue theorist. “What does ‘right’ refer to?” we ask him. “The property of being what a flourishing human would do,” he replies. Suppose I am correct about the truth of utilitarianism. There are two explanations of what Virgil says. He might mean the same thing I do by ‘right’ but he is mistaken about its referent. (This is what we want.) But alternatively, Virgil might be using a homonym, and perhaps he is correct in his belief about what ‘right’ *in his sense* refers to. This “indicates a problem about the metaphysical grounding of the semantic facts related to ‘[right]’: this circle of dependence makes it hard to see how those semantic facts ever get fixed in the first place.”¹⁰

⁸ Rubin (2015: 393).

⁹ Van Roojen (2006: 178).

¹⁰ Rubin (2015: 402).

I am not sure how deep the problem of circularity is by itself. Consider the term ‘reference’. What relation does that word refer to? If the debates about reference are to be substantial, ‘reference’ refers to the same relation in the mouths of causal theorists, functionalists, teleological theorists and so on, but substantive first-order debate in the theory of reference must be settled before we know what ‘reference’ refers to. Likewise, I see no issue in thinking we must do first-order ethics before we know which descriptive properties are referred to with our moral terms. The real problem is that of Virgil, i.e. the problem of knowing *who* is using moral terms in the same sense rather than as homonyms. And this is where the advantage of expressivism over Cornell realism finally materialises.

5.4 Peace is Unstable and Expressivists Win the War

Let us review. I have taken quasi-realism very seriously. Since I believe some substantive moral truths like ‘torture is wrong,’ I believe there are moral properties like wrongness. I have argued that expressivists are committed to thinking moral properties (if they exist) are reducible to descriptive ones. And due to Moorean open question considerations I have taken for granted that moral terms do not *mean* the same as any descriptive terms.¹¹ This has put me very much in league with the Cornell realists. Blackburn was unimpressed with the lack of any motivational elements in the Cornell realist view, but these were added in LUNCH. He was also rightfully cynical about the causal theory of reference, but this can be switched for a normatively enriched theory (Brink 2001). So here are the points of agreement between myself and the best version of Cornell realism: there are mind-independent moral facts, knowledge of them is synthetic, they are descriptive facts, reference to them is fixed normatively, and moral beliefs about them are hybrid states. The divergence is to be found in the hybrid states themselves.¹² Here is how both theories interpret the mental state expressed by ‘helping is right.’

¹¹ I argued in §3.1.1 of chapter 1 that expressivists could, if so inclined, believe that moral truths are analytic on a suitably normative theory of meaning. Yet I do not think this route escapes some of the issues I have raised in this section. If ‘right’ *means* maximises happiness, the question is why. If it is because it is *right* to infer from one to the other (or whatever the normative theory says), then issues of circularity and semantic grounding arise, exemplified by Virgil. See Gibbard (2012: ch.9) for discussion of such circularity.

¹² Another potential divergence is that Cornell realists have traditionally maintained that ethics is an *a posteriori* business whereas I, along with most metaethicists believe ethics is a

LUNCH: [H! (Actions with C) & robust belief that *helping has C*]

HOE: [H! (Actions with property K) & robust belief *helping has C*]

Recall that C is an unanalysable concept which, on LUNCH, refers directly to whatever the rightness property is, such as maximising happiness; yet according to higher-order expressivism (HOE), concept C refers directly to property K, which is determined by the speaker's moral perspective.

Here is something that may be seen as an advantage of HOE: the concept in the robust belief component may refer to K using any of the far less controversial theories of reference, such as causal theories. However, this takes quasi-realism less seriously than I want to. I cannot ignore the fact that the overall moral belief (i.e. everything in square brackets) is a *real* belief, and this moral belief makes reference to the *real* property of rightness. On both LUNCH (with the theory of reference amended) and HOE this reference is achieved via normatively enriched mechanisms. In short, I don't think normative theories of reference are a bad thing. They are necessary for synthetic naturalists to avoid outsourcing morality to a morally detached theory of reference.

Instead, consider Virgil. Suppose he utters 'killing one to save five is not right.' And suppose that 'right' as *we* use the term in the moral sense refers to maximising happiness. Due to our observations of Virgil, we can see he acts in a manner best explained by him having the appropriate desire-like state we label 'H!' The problem is that if LUNCH is correct, there are now at least two different viable hypotheses about Virgil's expressed mental state:

Virg1: [H! (Actions with C) & robust belief that *killing one to save five does not have C*]

Virg2: [H! (Actions with D) & robust belief that *killing one to save five does not have D*]

According to LUNCH and our supposition, C refers to maximising happiness. Assume killing one to save five maximises happiness. If Virg1 is indeed Virgil's mental state, Virgil's robust belief and therefore his claim is

priori. However I am unsure whether they would maintain this after adopting a normative theory of reference. Causal theories of reference make it easy to see why ethics is *a posteriori* – to find what's right, observe people's use of moral terms and discover what causally regulates it. Since I am unsure what the *a posteriori/a priori* debate would look like after rejection of non-normative theories of reference, I omit discussion of it.

false. (This is what we want.) But the given facts are consistent with Virg2, according to which Virgil deployed a *different* unanalysable concept *D* which might refer to some other property in some other way, and may well render his robust belief and therefore his claim *true*. We have gross semantic indeterminacy. This consequence would not occur with, say, the causal theory of reference. We could in theory observe the patterns of Virgil's use of 'right' and detect that it is causally regulated by maximising happiness. The two hypotheses would then look rather less distinct, for they both propose that Virgil has an unanalysable concept which refers to maximising happiness and which renders Virgil's belief and claim false; the only difference is the letter used to represent it, which doesn't look like a real difference at all. But we are dispensing with causal theories of reference for moral terms and using normatively enriched ones instead. To use this theory of reference in the appropriate circumstances we must be able to identify when someone is *genuinely* using moral terms rather than terms that simply look and sound *like* our moral terms, which LUNCH could not do. Herein lies the advantage of HOE. Consider what similar hypotheses would look like according to higher-order expressivism.

Virg1: [H! (Actions with property K) & robust belief that *killing one to save five does not have C*]

Virg2: [H! (Actions with property K) & robust belief that *killing one to save five does not have D*]

As noted earlier, the robust belief components may refer according to standard theories of reference. Using these theories of reference we can identify what *C* and *D* refer to. If they both refer to property K, then Virg1 and Virg2 amount to the same thing, and it is a genuine moral judgement. If this is what Virgil was up to, we know he was using 'right' in the moral sense and his use is subject to the enriched theory of reference – 'right' refers to maximising happiness, even though *C/D* referred to property K which, in Virgil's own perspective, was *not* the property of maximising happiness. However, if *C* or *D* refers to something *other* than property K, then Virgil was not making a moral judgement and his use of 'right' was not in the moral sense. How non-moral terms refer is not my business.

To summarise, the advantage expressivists have over this form of Cornell realism is that we can in principle identify when someone is making a genuine moral judgement *before* knowing whether or not to apply the normatively enriched theory of reference to the term. Cornell realists are in

the unhappy position of having to *simultaneously* work out (a) what an unanalysable concept refers to and (b) how it so refers, *before* they can tell whether the judgement is a moral one. This looks deeply troubling if not outright impossible.

6. Point Scoring

In this section I list several advantages I believe this theory has with regards to some traditional issues in metaethics. I then explain why this is different to expressivists who have come before me.

Meaning. Moral terms do not mean the same thing as any descriptive terms. This explains why Moorean questions are open.

Embedding. The Frege-Geach problem can be solved by making use of the robust belief components as described in the previous chapter.

Properties. Moral properties are descriptive and (my moral stance tells me) are mind-independent and natural. They are not distinct, queer, irreducibly normative entities as robust and relaxed realists believe. There is no problem of supervenience; moral properties supervene on certain natural ones because they are identical.

Reference. I can avoid outsourcing morality to non-moral theories of reference, where cognitivist synthetic natural moral realists run into trouble.

Motivation. There is a connection between making a moral judgement and being motivated to act in line with it due to the desire-like nature of the moral perspective. This connection can be severed – the fact that you approve of actions insofar as they have K and believe ϕ has K does not logically entail the desire or plan to ϕ (compare Gibbard (2003) where ought-judgements *are* plans) – but it does explain why there is a *tendency* to desire to ϕ .

Disagreement. Naturalists often have trouble accounting for the possibility of moral disagreement persisting after agreement has been reached about all the natural facts. Expressivists can account for this possibility with the common conviction that people's desire-like attitudes can diverge even if their robust beliefs are in

agreement. Exactly what disagreement *is* on the expressivist picture is tackled in the next chapter.

Scepticism. Realists, perhaps even by definition, are haunted by the spectre of radical scepticism, the thought that the moral truth might be beyond reach of our best inquiry. In chapter 5 I argue that we expressivists are committed to believing we must be able to reach the moral truth in principle, thus avoiding radical scepticism.

Of the conclusions I have reached in this chapter, how much is new? First of all, the expressivist position on moral properties has sometimes been unclear. Ingram attributed the thesis of *Non-Reductivism* to expressivists, and this is understandable. Blackburn (1993: 181), for instance, writes there is “no harm in saying that ethical predicates refer to properties, when such properties are merely the semantic shadows of the fact that they function as predicates.” If the most we can say about moral properties is that they are the “semantic shadows” of moral predicates which, as Blackburn agrees, are not reducible to descriptive predicates, then it does seem as though Blackburn commits himself to *Non-Reductivism*.¹ Yet I have argued that moral properties are not merely semantic shadows but are full blown, robust, descriptive properties. Toppinen’s (2018) paper also explicitly assumes throughout that moral properties are *sui generis* and non-natural. It is an (ingenious) answer to the question of why, if expressivism is true, moral properties supervene on descriptive ones; but if moral properties *just are* descriptive properties, there is no need for such ingenuity. Likewise, Dreier’s (2015b) paper on the problem of supervenience for quasi-realism assumes that the supervening moral properties are not descriptive.

At the other end of the scale, Ridge (2014: 224) develops his expressivism in such a way that there are “normative propositions, normative truths, even normative facts, and extensions for normative predicates, but no normative properties.” He notes that he is very much open to the possibility of vindicating talk of normative properties, but he himself does not do so. I hope that Ridge would agree that my arguments in this chapter are compatible with his view.

The expressivist whose toes are most likely to have been trodden on is Gibbard (see especially 2003: ch.5). He, like me, makes what he considers to

¹ At other times, as we saw in §4.3 of chapter 1 and §5.1 of this chapter, Blackburn seems to imply moral properties are natural.

be a transcendental argument in favour of normative properties being reducible to (or “constituted by”) descriptive ones. However, unlike me, he does not make any room for the first-order view that nothing has such properties (as I did in §4.1), and his argument is very different, resting primarily on supervenience considerations. Nevertheless, his conclusion is similar. To illustrate he introduces Hera who thinks maximising her own pleasure at all costs is okay to do. He then supposes (2003: 103) “Hera is right and the property of being okay to do just *is* the property of being egohedonic. ... We then have two concepts of the same property, the concept of being egohedonic and the concept of being okay to do. In a loose sense, therefore, we can “ascribe” this property to an act in either of two ways, by calling it “egohedonic” or by calling it “okay to do”.”

If Gibbard has been here before me, what was the benefit of writing this chapter? Firstly, it’s always nice to have a different argument for the same conclusion in case one of them fails. Secondly, he is a pure expressivist and for the arguments I gave in §2.1 I am sceptical he can make good sense of the metaphysical claims he wants to make. Recall that pure expressivists could not distinguish between correlation and reduction. This is perfectly exemplified by the way he considers (2003: 96) the claim “There is a factual property that *constitutes* being okay to do” to be an “abbreviation” of the claim that “There is a prosaically factual property *F* such that for any act *a* open in any possible situation *s*, act *a* is okay to do in *s* just in case *a* in *s* has property *F*.” The latter claim is merely a biconditional and should not be seen as providing a reduction. Thirdly, even though Gibbard says we can ascribe the relevant descriptive property in normative terms (albeit “loosely”), he does not provide an account of how this is done. I have outlined a theory of reference for moral terms. Furthermore I used this theory to argue that expressivist synthetic natural realists have the edge over cognitivist synthetic natural realists. All in all, I believe I have made a novel contribution to expressivist metaphysics.

Conclusion

In §1 I argued that metaphysical questions (or questions that look identical to them) can be asked whether or not we interpret them as internal to moral discourse. In §2 I showed how hybrid expressivists can make sense of the answers to these questions by providing an account of what it is to accept

the Realist Commitments within a moral perspective. In §3 I argued against *Non-Reductivism* by showing its acceptance involves a false robust belief. This took robust and relaxed realism off the table as potential positions expressivists can “mimic” but, fortunately, I showed in §4 that we still have several metaphysical options. In §5 I plumped for one, namely synthetic natural moral realism. The difference between cognitivist and expressivist versions turned out to hinge on the consequences of how we can use moral terms to refer to descriptive properties, and I argued it tips the balance in favour of expressivism. Whether or not I have been persuasive I hope expressivists engage further in metaphysics. They are not exempt from the metaphysical field, on which important battles can be won.

~4~

An Assertoric Theory of Disagreement

Expressivist theories of normative thought and discourse interpret normative sentences as expressions of desire-like attitudes. 'Stealing is wrong' might express disapproval of stealing, for instance, rather than a robustly representational belief that stealing has the property of wrongness. Some prominent arguments for expressivism rest on various apparent phenomena of disagreement, but there is no consensus among expressivists on what disagreement in desire-like attitude *is*. The aim of this chapter is to provide an elegant and unified account that is friendly to expressivism.

I review the three main expressivist theories of attitudinal disagreement – offered by Charles Stevenson, Allan Gibbard and Michael Ridge – in the first two sections. I argue these proposals face major difficulties. The accounts given by Stevenson and Gibbard classify some attitudinal differences as genuine disagreements when they intuitively seem not to be, or vice versa. In light of this, Ridge proposes we understand disagreements in terms of conflicting prescriptions. I argue this fails for the same reason that prescriptivism about normative judgement fails: it faces deep semantic problems when it comes to interpreting disagreement about compositional sentences.

This motivates the search for another theory. In §3 I outline my assertoric theory of disagreement and argue that it overcomes the problems of prior accounts as well as anticipated objections. I propose we tie disagreement to what is (or would be) asserted. Parties disagree if and only if an agent cannot, without inconsistency, simultaneously adopt all of the attitudes expressed (or would be expressed, were the agents speaking

sincerely) with their assertions. Effectively I reduce the problem of disagreement to the Frege-Geach problem, which I take to have been solved in chapter 2. This theory avoids the difficulties of previous attempts and explains what unifies the attitudes apt for disagreement: they can all be expressed with assertions. I then show in §4 why all expressivists, whether pure or hybrid, have good reasons to be optimistic about this theory of disagreement.

I should note that Yonatan Shemmer and I (2017) have developed a different theory of disagreement to the one presented here. We argue disagreement is a normative matter involving the reasons parties have to change their minds. If expressivism about these reasons is true, this implies sentences like ‘Laurel and Hardy disagree’ express desire-like attitudes. I do not discuss our theory in this chapter because, first of all, I am not sure which theory is better, and second, I suspect they are not in competition. When claims of moral mind-independence were interpreted as normative in §2 of chapter 1 and §1 of chapter 3, this allowed expressivists to debate the merits of mind-independence at the first-order level. In a similar vein, if Shemmer and I are right about disagreement being a normative matter then this chapter can be viewed as first-order normative theorising about the conditions under which people have disagreement-relevant reasons to change their minds. If we are wrong, this chapter is a straightforward conceptual analysis of disagreement.

1. Disagreement in Attitude

If I believe aliens exist and you believe they don’t, we disagree. If I have a headache and you don’t, we do not disagree – at least, not in virtue of *that*. Likewise, if I am in love with Harper and you are not, this is no disagreement either. Beliefs are the right sort of mental state to ground disagreements whereas love and headaches are not. Why? As a first pass, perhaps disagreement indicates that at least one person is incorrectly representing the world. We disagree *in belief* about aliens because one of us is guaranteed to be wrong. We do not disagree *in love* for Harper because, although love may sometimes be unfortunate or even inappropriate, love is not a robustly representational belief about the way things are.

This first pass would be fine so long as cognitivism, or perhaps only cognitivism with no relativist components, is true of all discourses in which

we find disagreement. The problem arises if declarative sentences in a given domain express non-representational attitudes like desires and intentions. A. J. Ayer, one of the forerunners of expressivism, believed moral language expressed emotions rather than truth-apt beliefs. His theory was caricatured as the ‘Boo/Hurrah’ theory: to call something bad is to jeer it, and to call something good is to applaud it. This made moral disagreement difficult to grasp. There is no disagreement when supporters of a football team cheer a goal while others boo it, only a difference in reaction. Since there are moral disagreements, expressivists need a theory of disagreement that extends to mental states beyond robustly representational beliefs. In the rest of this section and the next, we will examine why the three most prominent accounts fail to do the job.

1.1 Stevensonian Theories of Disagreement in Desire

We disagree in our beliefs if the content of my belief is inconsistent with the content of yours. Stevenson extended this to desires – perhaps we also disagree if the content of my desire is incompatible with the content of yours.¹ We might summarise the theory as follows:

Parties disagree if and only if they have beliefs which cannot all be true or desires which cannot all be satisfied.

This theory has been very influential for later expressivists, particularly Simon Blackburn,² and even content-relativists have tried to locate moral disagreement “in a (broadly Stevensonian) conflict of attitudes”.³ It explains why disagreement tends to worry us – it either indicates one of us is incorrectly representing reality or that one of us isn’t going to get our wish. Unfortunately it is too permissive. When I play badminton, both my opponent and I want to win, yet only one of us can. One of us is going to leave the court unsatisfied. But it is absurd to say that badminton opponents thereby *disagree* with one another. What do they disagree *about*? The theory as stated lets in far too many cases because desire conflict is not sufficient for disagreement. Opponents playing games, armies fighting battles, and

¹ When two people disagree about where to go for dinner, their “disagreement springs more from divergent preferences than from divergent beliefs.” (Stevenson 1944: 3.)

² See especially Blackburn (1998: 69).

³ Finlay (2017: 193). Finlay refers to this strategy as “quasi-expressivist” and in my view he should be worried about the problems with Stevensonian theories of disagreement.

people competing for jobs do not necessarily disagree about anything. Stevenson sometimes added a further condition which restricts the scope:

Two men will be said to disagree in attitude when they have opposed attitudes to the same object – one approving of it, for instance, and the other disapproving of it – and when at least one of them has a motive for altering or calling into question the attitude of the other. (Stevenson 1944: 3.)⁴

This extra requirement is meant to prevent cases like those of the badminton opponents from being classed as disagreements. Because I have no motive for altering my opponent's desire to win (I want a good game, after all), and likewise for my opponent, we do not disagree. But what if I care not for fair play? Perhaps I desire ease of victory above all, and would quite like my opponent to *let* me win. I therefore have a motive for altering my opponent's attitude – I want him to desire *my* winning instead of his own. This satisfies the extra motive condition, but surely it still falls short of a disagreement.

Additionally, as Ridge (2014: 175) highlights, the “Stevensonian account seems to leave it mysterious why we do not apply the concept of disagreement to other states with propositional contents.” Agents do not disagree in their suppositions or imaginings, for example. I can imagine *p* while you imagine its negation and we do not disagree with each other. A theory of disagreement ought to explain what is special about the states of mind that are capable of the right sort of conflict. (As we'll see later, I believe what unifies these states is their ability to be expressed with assertions.)

The largest problem for the theory after adding the further condition is its failure to account for a great number of disagreements. Suppose I believe in God and you do not, but neither of us cares for religious debates and neither of us wishes to persuade the other. In the name of diversity we may even prefer that we each retain our beliefs, yet we clearly still disagree about the existence of God. In short, disagreeing parties needn't have any wish to change one another's mind. Stevensonian theories place necessary and sufficient conditions on disagreement that are either too permissive or too restrictive. Expressivists should look elsewhere.

⁴ In his later writings he continued to presume people disagree only “when neither is content to let the belief of the other remain unchallenged.” (Stevenson 1963: 1).

1.2 Gibbard's Theory of Disagreement in Plan

Gibbard (2003) believes normative judgements are plans. Roughly, my judgement that I ought to drink the potion in front of me is my plan to do so. Suppose you think I ought *not* to drink the potion. Your judgement – surely in disagreement with my own – is your contingency plan not to drink the potion if you find yourself in my situation. A similar sentiment can be conveyed with the phrase ‘I wouldn’t do that if I were you.’ But why is the fact that we plan to do different things in the same circumstances sufficient to ground a disagreement? In several of Gibbard’s passages, a natural reading is that he takes disagreement to be primitive:

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. ... 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. (Gibbard 2003: 74.)⁵

Perhaps, then, the concept of disagreement is primitive and no unifying explanation is in the offing. If this is the correct interpretation, it is unsatisfying. To many of us it doesn’t *feel* like a disagreement when people plan to live their lives in different ways, so a non-primitive theory of disagreement would be helpful. Furthermore, it doesn’t tell us why people with other differing mental states like headaches and being in love don’t enter these primitive disagreement relations. But I don’t think Gibbard in these passages is thinking of disagreement as a relation between people at all.⁶ He is considering the state of disagreeing with p as the one that is primitively inconsistent with deciding to, or deciding that, p , which we might call *intrapersonal* disagreement. It is plausible, and plausibly primitive, that a single agent cannot consistently plan to do one thing while rejecting that very same plan. The difficulty is to explain *interpersonal* disagreement. Since Gibbard goes to great lengths later in the book to defend the possibility of interpersonal disagreements in plan, I assume he

⁵ The appeal to partners in crime is echoed in later work: “Fregeans start out with negation. They must either explain it or help themselves to it without explanation. Expressivists start out correspondingly with disagreement, and must likewise either explain it or help themselves to it without explanation.” (Gibbard 2012: 273-4.)

⁶ In his section on ‘Excluding States of Mind’ (2012: 274-7), Gibbard writes of a single agent’s *state* of disagreement as one of excluding certain mental states, rather than as a relationship between agents.

does not take interpersonal disagreements to be primitive. However, his defence fails to extend to all cases we intuitively class as disagreements. Let us examine the problem.

Neil thinks that if you're going to study philosophy, you should read the Ancient Greek classics. Tasmin thinks you shouldn't. They are in normative disagreement. But if these judgements are merely their own contingency plans, it is hard to see what the issue is. Neil plans to read the classics should he study philosophy, while Tasmin plans *not* to read them should she study philosophy. So what? Gibbard anticipates the concern:

This is no disagreement at all, it might be objected. It is just a difference of personal characteristics, like having different hair colors. One person plans to do one thing, and the other to do something else. True, their plans are for exactly the same situation that one could be in – but why does this make their difference in plans a disagreement, a difference of opinion? (Gibbard 2003: 270.)

Gibbard's response to the worry is intriguing. Ridge (2014: 181) describes it as "a kind of transcendental argument that we must treat at least some forms of what he calls disagreement in plan as genuine disagreements." Consider single agents across time. When we accept a plan that we later come to reject, we come to view it in the same way we view our once-accepted-now-rejected beliefs – as a state we have come to disagree with (intrapersonally) after our change of mind. Gibbard justifies the move to interpersonal disagreement by espousing the merits of *treating* other people's diverging plans as disagreements. Identifying areas where our thinking diverges is useful for spotting the beliefs or plans we may want to reconsider. He maintains this approach in later work (2011: 48): "The point of regarding us as often disagreeing with each other, I say, is this: we sometimes need to "put our heads together" about a problem, thinking about it jointly."

It is certainly true that I question my plans whenever I learn that someone, whose planning I admire, has made different plans for the same circumstances. If I plan to take the high road and then see my esteemed friend confidently taking the low road, I naturally ask myself "What does she know that I don't?" But notice the qualification that I respect the judgement of those I purportedly disagree with. If I find out upon closer inspection that the person taking the low road is not my esteemed friend but Terrible Terry who always makes terrible decisions, I would *not*

question my own plan. And, being someone who cannot identify a good planner when he sees one, Terry does not respect my decision-making either. The two of us see no benefit whatsoever in treating our respective plans as disagreements. He can take the low road if he wants; I'm sticking with the high road.

Gibbard again foresees this worry. Agents may have such fundamentally different values that there is no point reasoning with them and, "with such an impasse, with no ways left to work to accord, we could find no point to treating questions of how to live as topics for agreement and disagreement."⁷ This might mean that the agents still disagree, but hopelessly. If this is the case though, we need an explanation of *why* they disagree, because by Gibbard's own admission there is no benefit to them *treating* their difference as a disagreement, which was meant to anchor interpersonal disagreement in the first place. On another potential reading – and this is the reading Ridge takes (2014: 185) – Gibbard thinks intractable deadlocks, because they are not worth treating as disagreements, are *not* disagreements. This is hard to swallow. Disagreement does not disappear when it becomes more entrenched. Suppose some of my judgements fundamentally diverge from those of an ideally coherent misogynist. He thinks women ought to be subservient to men and, due to our impasse, he has no rational means to persuade me of his view. But it is implausible that I do not disagree with the misogynist when I respond, pointlessly or not, with the claim that we ought to treat people of all genders equally. As Ridge puts it (2014: 185), "[q]uite the contrary. The point instead is that our disagreement is too deep for conversation to have a point."

This brings me to the end of this section. Stevensonian theories were either too permissive or too restrictive, and if Gibbard is not to leave the concept of disagreement unexplained, his theory relies on a transcendental argument that fails to account for cases of impasse. None of the theories we have considered explain the full range of normative disagreements. I will argue against one final theory in the following section before proposing my own.

⁷ Gibbard (2003: 283). Lenman (2014: 242) makes a similar point about strange alien creatures from whom we could not hope to learn anything about how to live.

2. Disagreement in Prescription

An important function of moral discourse, if not the primary function, is to regulate behaviour. When we attempt to persuade a person that an act is wrong, it is normally with the hope that they do not perform it. It is an unsatisfying end to a moral conversation when we agree with the vegan's viewpoint but intend to continue consuming animal products. Vegans do not merely wish to be acknowledged as correct – they want us to stop! Normativity *directs* us. Ridge's theory of disagreement (2013, 2014: ch.6) captures this idea rather well. Instead of examining our actual desires or plans, Ridge invites us to consider what advice we would, in ideal circumstances, give to an agent. You and I disagree if we would prescribe incompatible courses of action for this agent, and we agree if we would prescribe the same course of action.¹ Since I would advise agents to treat people of all genders respectfully and the coherent misogynist would give advice that conflicts, Ridge's account correctly renders us in disagreement.

The theory naturally extends to disagreement in belief. Suppose a creationist and I are asked to advise an agent about the age of the universe. The creationist advises the agent to believe the universe is 12,000 years old, but I advise her to believe it is closer to 14 billion years. Now, it may be *possible* for the agent to follow both pieces of advice and believe both propositions, but she cannot do so without acquiring inconsistent beliefs. Ridge's formal definition of disagreement makes room for this:

Two people (or two stages of the same person at different times), A and B, disagree in prescription about D's Φ -ing in C just in case in circumstances of honesty, full candor, and non-hypocrisy, A would advise Φ -ing in C and B would advise Ψ -ing in C, where Φ -ing and Ψ -ing are incompatible, in the sense of being impossible to combine without thereby having inconsistent beliefs. (Ridge 2014: 190.)

There is something peculiar about locating conflict in what agents would prescribe rather than in the judgements themselves. Whereas Stevenson thought moral judgements are desire-like states and his theory of disagreement referred to the potential conflict between them, and Gibbard thinks normative judgements are planning states so his theory is couched in terms of divergent plans for the same circumstances, Ridge does not

¹ This leaves space for a third class of prescriptions that are compatible but not identical. I return to the idea of compatibility in §4.3.

think normative sentences are prescriptions. There is a gap between Ridge's theory of normative judgement and his account of disagreement.

This might mean expressivism loses its main weapon against content-relativism. To illustrate, consider a naïve subjectivist theory that interprets moral sentences as direct reports of the speaker's non-cognitive attitudes: Mary, in saying the war was right, is merely stating she approves of it. Halim says the war was wrong, merely reporting that *he* disapproves. Subjectivists have difficulties explaining why Mary and Halim disagree, since they both assert propositions that neither person disputes. This problem traditionally motivates the move from *reports* of non-cognitive attitudes to direct expressions of them, but this move is unwarranted if Ridge's theory is right. If disagreement consists in conflicting hypothetical prescriptions, the subjectivist can point out that Mary would prescribe the war while Halim would advise against it. Subjectivists would be off the hook. Of course, this is only a dialectical problem. In the next subsection I look at a deeper problem raised by James Dreier and I argue that Ridge's response does not quite overcome it. I then raise an objection of my own that I think is even more devastating, and that leads naturally into my own assertoric theory of disagreement.

2.1 Permissibility

Roe thinks abortion is permissible. What advice would she give to an arbitrary agent about whether to have an abortion or not? Ridge thinks Roe would, in ideal conditions, advise the disjunctive plan to either have an abortion or not have one. Wade thinks abortion is impermissible, so he would advise *not* having an abortion. At first blush, their advice is compatible – I can avoid abortions and I'll be contravening neither Roe nor Wade. But plans are not like beliefs. If I believe '*p* or *q*' and later come to accept *p* on the basis of rejecting *q*, I still believe '*p* or *q*'. If I'm planning to turn left or go straight on, come to realise there's a monster straight ahead, and so abandon my disjunctive plan in favour of the plan to turn left, I have *rejected* my former plan. There is disunity between plans and beliefs. To illustrate, imagine overhearing someone telling their child that Alaska is either the 49th or 50th state. That's true, despite allowing an option that's false. Now imagine them telling their child to either share her sweets with her sister or to punch her in the face. That's unacceptable, *because* it allows an unacceptable course of action.

Ridge (2015a: 481) uses this disparity to explain why advice of the sort given by Roe and Wade conflicts: “Two pieces of advice conflict if and only if one rules out an option the other permits.” Wade’s advice rules out having an abortion, whereas Roe’s advice explicitly permits it. So far, so good. However, Dreier worries that one’s advice may not line up in the right way:

Suppose I say John may skip the party and you say he must go to the party. We disagree. You would advise John to go to the party if you were in the super-duper conditions. ... Ridge expects me to advise John to keep the option of skipping the party on the table as a live option. But that depends on whether I regard keeping the option on the table as required. Suppose I do not. ... I think he could keep skipping the party on the table as a live option. But it’s also fine for him to go to the party, and it’s fine for him to take skipping the party off the table and no longer consider it to be a live option. So I do not advise him ... to keep skipping the party on the table. (Dreier 2015a: 457-58.)

Ridge responds to this objection by differentiating what’s being put on the table *just now* and what can be taken off later. Whenever we advise someone to adopt a disjunctive plan, we don’t expect them to keep all the options on the table forever. At some point they’ll pick one of them. But, writes Ridge (2015a: 483), “the advisor is proposing the agent consider [skipping the party] to be an option *just now* – and in offering disjunctive advice I do seem to be doing at least this much.” It’s certainly true that this happens. If I see John distressed about having to go to the party, I might try to convince him that he’s not *obliged* to go in order to put the choice back on the table. And of course I don’t expect him to be in this “all for grabs” state all the time; I just want him to reassess his options at least once more. But this is not always the case. Suppose Roe, while maintaining that abortion is permissible, thinks it is also permissible for a person to *never* consider abortion as a live option, not even once. Dreier’s objection goes through.

The objection is limited in scope. I imagine people who are pro-choice tend to think the choice should be made, and this requires that agents put the option on the table at some point. Ridge could therefore retreat to saying that his theory of disagreement covers all cases but those at the fringe. This is not ideal, but perhaps he could settle for, in his words (2015a: 478), “the worst going theory of the concept, except for all the rest.” Unfortunately,

the theory also fails to account for a much wider class of disagreements: disagreements about embedded normative claims.

2.2 The Resurrection of the Frege-Geach Problem

Ridge, for very good reasons, is not a prescriptivist about normative language. He does not think that ‘stealing is wrong’ means ‘do not steal.’² Prescriptivism is not a live contemporary option because it struggles to account for compositional semantics, a vital part of solving the Frege-Geach problem (Geach 1960, 1965). It is unclear how a prescriptivist would interpret even simple conditionals like ‘if stealing is wrong, taxation is wrong’ – the sentence ‘if do not steal, do not tax’ makes no sense.

It is no exaggeration to say that Ridge created his ecumenical brand of expressivism in order to solve the Frege-Geach problem. By incorporating representational beliefs into moral judgements he hopes to explain compositionality and inference “on the cheap” (Ridge 2006a: 309). Unfortunately, in making such heavy use of prescriptions in his theory of disagreement, the same problems resurface. This is because disagreement is not always expressed with atomic sentences.

Suppose Ayn accepts the libertarian conditional ‘if stealing is wrong, taxation is wrong.’ John rejects it.³ To assess whether they disagree – and they surely do – Ridge would have us consider what they would advise an agent to do in ideal circumstances. But what would they advise? Ayn would be talking gibberish if she said ‘if do not steal, do not tax.’ Here is one possible option:

Ayn: If you think stealing is wrong, don’t tax!

John: If you think stealing is wrong, do or don’t tax!

This won’t do. The problem with this analysis is its identification of ‘stealing is wrong’ with ‘you think stealing is wrong.’ The antecedent of the libertarian conditional is about the wrongness of stealing; it is *not* about the advisee’s opinion. It is perfectly consistent for John to think an agent should avoid taxation if they judge stealing to be wrong, while simultaneously

² The view that the meanings of moral sentences are best understood in terms of imperatives can be found in Carnap (1935) and Hare (1952).

³ Note that neither Ayn nor John must take any particular stance on the moral status of stealing or taxation.

thinking there is no correlation between the *actual* wrongness of stealing with the wrongness of taxation.

Perhaps Ridge can look at the plans Ayn and John might advise an agent to rule out. Assume the libertarian conditional is a material conditional. Maybe Ayn, under ideal conditions, would advise an agent to either reject ruling out stealing or to reject taxation. John would advise an agent to rule out stealing and to rule out rejecting taxation. I think this is better. It's certainly closer to my own theory, to be introduced shortly. However, there are two reasons Ridge would not be happy with this move. The first is that it relies on it being impossible for an agent to accept both the recommended plans without thereby acquiring inconsistent beliefs – but this is a major part of the Frege-Geach problem as applied to plans, which Ridge doesn't think can be solved. (If it can, there is less reason to go ecumenical.) Secondly, Ridge's theory is couched in such a way that *normative* judgements lead to advice for *action* whereas *descriptive* judgements lead to advice for *thought*. When I tell Rosie she ought to go left and you tell her she ought to go right, what was meant to ground our disagreement was the impossibility of her *going* left and *going* right. Normative disagreement, on Ridge's theory, was not meant to be grounded in the inconsistency of planning states.

If we want to continue understanding disagreement in terms of conflicting advice for *action*, it is unclear how to interpret disagreement about conditionals. This is a relatively simple case, too. We haven't begun to consider more complex embedding, tensed judgements, or all the different varieties of modals and quantifiers. Prescriptivism had great difficulty accounting for these kinds of complexities, and, given that we can disagree with any arbitrarily complex normative claim, basing disagreement in prescription is going to inherit them all. Until we are given a satisfying explanation of compositional disagreement, Ridge's theory is, at best, incomplete.

In the following section I motivate the idea that normative assertions put pressure, first and foremost, on *what to think*. I then present my assertoric theory of disagreement and argue that it overcomes previous and anticipated problems. Finally, in §4 I show why the theory is good news for expressivism.

3. The Assertoric Theory

When extending his analysis of disagreement to standard disagreements of belief, Ridge writes that assertions function as prescriptions to believe. “Asserting that *p* is a way of advising someone to believe that *p*. ... For to assert that *p* is not only to express the belief that *p*; it is also to exert a kind of conversational pressure on one’s interlocutor to adopt the belief that *p*.”¹ He cites Gibbard making the same point:

Conversation is full of implicit demands and pressures. Suppose I confidently expound astrology, and you give no credence. The result will be discomfort: in effect I demand that what I say be accepted, and you will not accede. (Gibbard 1990: 172, quoted in Ridge 2014: 189.)

Ridge and Gibbard don’t mean that assertions *are* prescriptions to believe. Perhaps the prescriptions are merely pragmatically implied, which would make them cancellable. Maybe it makes sense to say ‘the government covered it up, but do not believe they covered it up,’ or at least *more* sense than ‘believe they covered it up, but do not believe they covered it up.’ The link between assertions and prescriptions to believe may or may not be properly labelled implicature, but that’s okay. The proposal is that in *most* contexts we experience pressure to believe what is asserted.

I suggest that we can save what was intuitively correct about Ridge’s theory of disagreement by extending this idea to *all* judgements. When I make a normative assertion, the relevant prescription is not about how to act – it is about *what to judge*. My assertion puts conversational pressure on you to form the expressed judgement. If you do not, you insult my epistemic prowess.

Not only does this have a pleasing symmetry with what Ridge and Gibbard wrote above, it also has clear parallels with what Stevenson thought moral claims mean. Stevenson (1944: 21) believed moral sentences were simultaneously expressions of the speaker’s emotions and imperatives for the listener to feel the same way: “‘This is good’ means *I approve of this*;

¹ Ridge (2014: 189). Interestingly, this leads to a kind of individuation problem. If I say “God does not exist” and you say “One ought to believe God exists,” Ridge’s theory incorrectly implies we have expressed disagreement because I would advise believing God does not exist and you would advise believing God exists. The move I am about to make will solve this problem too, because *thinking* God does not exist while *thinking* one ought to believe God exists is consistent.

do so as well." As I have mentioned, it is unlikely that hidden imperatives are part of the *meaning* of assertions. The idea I have in mind is that any assertion *p*, whether descriptive or normative, is merely *attended* with the widespread but perhaps contingent and cancellable pressure for the audience to judge that *p*.

The move from prescriptions of action to prescriptions of judgement prevents the resurrection of the Frege-Geach problem. Ayn advises the audience to share her acceptance of the libertarian conditional while John advises the audience to reject it, and metaethical theories should explain why the audience cannot consistently adopt both judgements. I will set the theory out more formally in the following subsection. Before we go any further, it will be instructive to consider the nature of assertion.

Assertions are speech acts. They are things we *do*. And to make an assertion is to do more than say some words – the mere locutionary act of uttering ‘snow is white’ is not sufficient for asserting that snow is white. Perhaps the speaker is in a play, for instance, or musing to herself out loud. What distinguishes assertions from other utterances is their unique *assertoric force*, the nature of which remains controversial. The assertoric theory of disagreement does not depend on what the correct theory of assertion turns out to be, but it will be helpful to have an account in mind.² Neri Marsili (2015) reconstructs a Searlian account in his discussion of normative theories of assertion whereby a speaker *S* asserts *P* if and only if:

- A1) *S* expresses the proposition *P*
- A2) *S* presents *P* as an actual state of affairs
- A3) *S* commits himself to the truth of the proposition *P*. (Marsili 2015: 122-23.)

The notion of commitment mentioned in A3 is to accept responsibility for providing justification for the assertion, where this may involve presenting arguments, evidence, or deferring to a testimonial authority.³ I am trying to accommodate disagreements within an expressivist framework, so the cognitivist-sounding language may give us pause. The literature on

² For a different model of assertion argued to be compatible with expressivism, see Pérez Carballo and Santorio (2016: 609) who believe that “to make an assertion is to propose an update of certain belief-like attitudes of the speakers.”

³ ‘Responsibility’ is a normative term. If a normative theory of assertion is true, and if normative expressivism is true, it follows that there are no *robustly realistic* facts about what was or wasn’t asserted. While I’m happy to accept this, Cuneo (2014: 177-87) argues this counts against expressivism. Proper engagement would take me too far afield.

assertion focuses on non-normative declaratives, and this is reflected in the terminology of propositions, states of affairs, truths and beliefs. However, since most contemporary expressivists subscribe to quasi-realism – the project of vindicating realist-sounding discourse (Blackburn 1993) – they will be comfortable using these words. Truth is typically understood minimally (although compare Ridge 2014: ch.7) such that to think it is true that torture is wrong is simply to think that torture is wrong, which on the expressivist view is a practical stance. Work has been done on constructing (Schroeder 2015: ch.3) and deflating (Köhler 2017) propositions for use in expressivist theory. Finally, beliefs are understood to be any judgement whether robustly representational or not. If Gibbard is correct, for example, then moral judgements are plans, which means plans are types of belief.

Some may insist assertions *only* express robustly representational beliefs, and never desire-like attitudes. I think this would only serve to enforce an unhelpful and arbitrary distinction (if not beg the question entirely against expressivists). Assertions are speech acts, and the act of making a normative claim involves the same kinds of commitment as making a non-normative claim. Suppose expressivism is true and I tell you that you ought to eat the mushroom. I am still likely to be sanctioned if I do not provide reasons for my claim when asked, or if I am insincere, or if my judgement is unwarranted. I put forward the claim that you ought to eat the mushroom as one to be accepted, not laughed at or enjoyed as if I were in a play. Whatever the assertoric force is, it is present in normative *and* non-normative claims. The potential for these forces to push in different directions is what grounds my theory of disagreement.

3.1 The Formalisation

I formalise the assertoric theory of (expressed) disagreement as follows:

Parties express disagreement if and only if (i) they make assertions which express attitudes A_1 towards content C_1 , A_2 towards C_2 , ... A_n towards C_n , and (ii) no agent can adopt all of these attitudes towards those contents without inconsistency.

To illustrate, suppose Victor the vegan, Carol the carnivore and Harry the hypocrite are having a conversation about eating animals. Victor says it is wrong, Carol says it is permissible. Both make assertions, perhaps in virtue of making themselves responsible for providing reasons for their views, and

an adequate metaethical theory will explain why Harry cannot adopt their attitudes towards eating meat – be they beliefs, desires, or both – without inconsistency. Victor and Carol have therefore expressed disagreement with one another.

For example, suppose Gibbard's *plan-expressivism* is true. Victor's assertion expresses rejection of any contingency plan to eat animals. Carol's assertion expresses rejection of rejection of any plan to eat animals. Gibbard takes it as primitive that it is inconsistent to reject a plan and simultaneously reject rejecting it. Therefore an agent who adopts the attitudes expressed by the two assertions simultaneously has a guaranteed inconsistency.

Now assume Mark Schroeder's (2008) theory is true. On this view, Victor expresses a state of *being for* blaming for eating animals and Carol expresses *being for* not blaming for eating animals. Schroeder's stipulated non-cognitive state of *being for* is *inconsistency transmitting*, which means the inconsistency of mental states is derived from the inconsistency of their contents. Just as beliefs are inconsistent with one another if their contents are inconsistent, states of *being for* are inconsistent when their contents are too. Since it is impossible to simultaneously blame and not blame for eating animals, it is inconsistent to *be for* them both. It is therefore inconsistent to adopt the same attitude (of *being for*) towards the distinct contents expressed by the two assertions.

My theory shows what unifies the mental states apt for disagreement: they are the ones that can be expressed with assertions. Headaches and fantasies do not come with the right sorts of responsibility to justify, for instance. But we may want to go hypothetical, as Ridge does. This is for two reasons. First of all, my above definition is about *expressed* disagreement. We may think that a liar, who insincerely expresses disagreement by contradicting what someone says, *actually* agrees with that person but doesn't admit it. The second reason is we might want to say I disagree with a flat-Earther even though I've never had the pleasure of meeting one. The real test is what someone *would* assert if they were given a truth-serum and forced to express their judgement. We thus arrive at:

Parties disagree if and only if (i) they would (if forced to, and keeping their current attitudes fixed) sincerely make assertions which express attitudes A_1 towards content C_1 , A_2 towards C_2 , ... A_n towards C_n , and (ii) no agent can adopt all of these attitudes towards those contents without inconsistency.

According to this account, it is almost certain that everybody disagrees. No two people agree about everything. The way to make the concept useful in practice is to restrict the potential contents to a given domain. (We may also restrict the potential attitudes, as when expressivists speak of two agents agreeing in all of their robustly representational beliefs but persisting in their moral dispute.) This domain could be as narrow as a single proposition or as broad as you like – my friend and I are in agreement when it comes to evolutionary biology, but not when it comes to politics.

3.3 Overcoming Prior Problems

The first formulation of Stevenson's theory of disagreement was too permissive. It classified any mutually unsatisfiable desires as disagreements, but badminton opponents who want to beat each other do not thereby disagree. The reason they do not disagree is because they do not express these wishes with assertions. How could the opponents express their desires? Well, perhaps the players could say 'mmm, winning' and look hungry for victory. This doesn't look like a disagreement, and it is exactly what the assertoric theory predicts. The players are making no assertions. They are using exclamatives, and exclamatives come with no assertoric force. When I express my pain by saying 'ouch' I put no conversational pressure on you to adopt my painful mental state and I do not incur the responsibility of providing justificatory reasons for being in that state. So, when both players express their desires through linguistic devices of this sort, there is no pressure for either player to change their mind. And this is precisely what was wrong with Ayer's 'Boo/Hurrah' theory – to boo or to cheer is not to assert. As J. L. Mackie (1946: 80-81) puts it, "we do not think that we are merely ejaculating when we talk in moral terms."

Gibbard's (non-primitive) account and the second formulation of Stevenson's theory were too restrictive. The former required there to be a reason for the interlocutors to take each other's judgements seriously, and the latter required at least one party to have a motive to change the other's mind. Neither requirement turned out to be necessary. Again, this is in accord with the assertoric theory. If one person asserts *p* and another asserts its negation they disagree, even if they are at an impasse and neither has a wish to persuade the other, because no agent can share all of their expressed attitudes without inconsistency. I disagree with the misogynist about how to treat women, and happily coexistent atheists and theists disagree too.

The assertoric theory unifies what is special about the attitudes we can disagree with: they can all be expressed with assertions. I know of no way to *directly* express with an assertion what I'm imagining or supposing. I can indirectly communicate it via a report such as 'I envisage the death of the king,' but the attitude directly expressed with this assertion is a *belief* about what I'm imagining. To assert 'it is the death of his twin brother you envisage' is to disagree in belief.

Lastly, there is no worry about compositional disagreement if the Frege-Geach problem can be solved for compositional *judgement*. An adequate metaethical theory will explain what complex sentences mean, even when they don't easily lend themselves to prescriptions for action, and why some judgements are inconsistent with others. That is all my theory requires.

3.4 Overcoming Subsequent Problems

I foresee three new potential concerns with the assertoric theory of disagreement. The first is the move to linguistic phenomena. Indeed, Teemu Toppinen (manuscript) has independently developed a theory of disagreement similar to mine but without the assertoric aspect – roughly, people disagree if we are unable to adopt their views without inconsistency.⁴ What work is assertion doing in my theory? Why not drop the linguistic element altogether? One of my motivations is the following possibility: moral judgements are not plans, differences in plans are mere differences and not disagreements, yet plans are still subject to constraints of consistency. Suppose *plan-expressivism* is false and consider Terrible Terry from §1.3. He plans to take the low road, which cannot be consistently combined with my plan to take the high road, but I don't think we disagree. Perhaps we agree on all factual and normative matters; perhaps Terry makes no (or very few) normative judgements at all. To put it another way, the fact that no agent can adopt both plans without inconsistency is neither here nor there, absent potential assertoric pressure to accept them.

⁴ Toppinen uses the phrase 'incoherent' but this invites variations of Moore's paradox – it would be incoherent to accept Valerie's belief that it's raining while also accepting Ralph's belief that nobody believes it's raining, but some may have the intuition that Valerie and Ralph don't disagree because what they believe is consistent. For these kinds of reasons my theory is in terms of inconsistency rather than incoherence.

Another reason to accept the linguistic element of my theory is because it is independently plausible that disagreement is something only readily ascribed to language users. Ridge (2014: 188) also notes how “we are hesitant to characterize non-linguistic agents as entering into relations of agreement and disagreement.” What of clever creatures like chimps? Some find it plausible that non-human higher primates can disagree about, say, where the grapes are hidden and whether they should be given to the alpha. I have to admit my intuitions are murky in these cases. But intuitions about language use are murky too.⁵ While chimps may not make full blown assertions they certainly communicate thoughts, and this may be why intuitions are unclear. It is also not outlandish to consider what a chimp would assert were it taught more advanced language use. This makes sense of why some people wish to attribute the capacity for disagreement to higher primates and some do not, and yet very few would attribute the capacity to animals with low propensities for language use, like slugs. Since we do not comfortably speak of cows and goats disagreeing with each other despite their having conflicting beliefs and desires, it is plausible that our concept of disagreement does have a linguistic element. I believe this element is the assertoric force pressuring us to align our attitudes.

This brings me to the second concern. I mentioned that the pressure to adopt the speaker’s attitude can be cancelled. If two speakers express conflicting judgements with assertions but cancel the pressure to judge accordingly, do they disagree? My answer is ‘yes’ – conversational pressure is not a necessary condition for disagreement but an explanation for why the concept is important to us. Creatures that do not care about aligning their attitudes perhaps would not develop a concept of disagreement. (Maybe they wouldn’t bother asserting anything either.) Given that we *generally* care about aligning our normative and non-normative beliefs in order to coexist, it has been useful to identify those cases where our judgements diverge. It has not been necessary to exclude cases where we do not wish to align judgement because they are few and far between. Therefore if I assert *p*, you assert its negation, and both of us cancel the pressure to align judgement, we still both disagree about *p*.

⁵ Savage-Rumbaugh (1994) and Rumbaugh (1995) discuss the abilities of apes to learn, use language, comprehend heard speech and recognise symbols and argue these capacities (including capacities to plan and reason) rest on a continuum across species. For an opposing view, see Wallman (1992). Thanks to Stephen Laurence for these interesting references.

Finally, one may worry I have retreated from what was attractive about Ridge's theory of disagreement in prescription. Ridge captured the *directedness* of normativity. When Victor asserts that eating meat is wrong, he hopes animal slaughter will *stop*. But suppose Harry says "I'm still going to eat meat even though it's wrong." With whom does Harry agree? I think it is clear that Harry agrees with Victor, not Carol, even though Victor will be upset. Harry has bowed to Victor's specifically *assertoric* force by sharing his judgement, if not his diet. However, this does not mean we have retreated from directedness altogether. Many philosophers, and especially expressivists, are motivational judgement *internalists*. In persuading Harry to judge that eating meat is wrong, Victor may have at least raised the likelihood that Harry's carnivorous behaviour will change. The purpose of expressing normative disagreement to regulate action and character is maintained.

An assertoric theory of disagreement overcomes the problems that plagued previous expressivist accounts and does not appear to suffer from any new ones. It is also extremely simple – if we have a satisfactory explanation of why it is inconsistent to simultaneously judge p and not p , then two agents expressing these conflicting judgements with assertions will be in disagreement. I will now in the final section examine the upshot for expressivism.

4. Why This Matters

My theory of disagreement matters because it allows all expressivists to correctly identify instances of moral disagreement. The way in which it does so is different for pure and hybrid expressivists. In the next subsection we see why pure expressivists have it easy, but afterwards we will run into a little snag for hybrid expressivists, identified by John Eriksson (2015). I then show how it can be unsnagged, allowing me to conclude after a discussion of *agreement* that expressivists of all stripes have reason to rejoice.

4.1 Pure Expressivism

An assertoric theory of disagreement is very friendly to pure expressivist theories, which are those theories that take atomic moral sentences to express only desire-like states. The one requirement is that the expressivist

understands the linguistic mode of expression to be assertive, as opposed to emotive (Ayer 1936) or prescriptive (Hare 1952).

Let us apply it to Blackburn's (1984, 1993, 1998) expressivism. Suppose I tell you that Theresa May is a good person and you reply that she is bad. I express approval of her and you express disapproval. Furthermore, we use *assertions*, generally accompanied by conversational pressure for our audience to share our attitudes.¹ Since the audience cannot consistently adopt both desire-like states, we disagree. Of course, some dispute the inconsistency of having conflicting desires – I may desire to eat meat and also desire not to because, respectively, meat is tasty and I wish no animals harm – but explaining why an agent cannot, without inconsistency, hold the mental states involved in conflicting moral judgements is part of the Frege-Geach problem. For pure expressivism, the (interpersonal) disagreement problem is solved if we have a satisfying story about (intrapersonal) inconsistency.

On the other hand, content-relativists may well have a disagreement problem. When Mary said she was for the war, according to naïve subjectivism she expressed her belief that she approves of it. Halim *agrees* with this belief (i.e. he believes the same content), despite being against the war. By locating conflict in what is asserted (rather than, say, a prescription leading from it), pure expressivists retain their dialectical advantage over naïve subjectivists. This is good news. Alas, a complication arises for a subset of hybrid theories.

4.2 Hybrid Expressivism: Answering Eriksson

Since the turn of the century, expressivism has become somewhat adulterated. Hybrid expressivists think moral judgements involve both robustly representational beliefs and desires. Some hybrid theorists (Boisvert 2008, Copp 2014) will be able to make sense of disagreement in standard ways: the representational beliefs expressed by disagreeing parties will be inconsistent. However, some theories allow cases where neither the beliefs nor the desires of disagreeing parties conflict. Stephen Barker (2000) and Ridge (2006a) would fall under this classification.

¹ Why are they assertions? Because the speech acts we perform satisfy the relevant conditions, whatever they happen to be. Perhaps it is because we present our claims to be accepted and we incur responsibilities to justify those claims.

Eriksson (2015) has argued that these theories cannot account for disagreement. I will show why his argument fails, allowing me to conclude that my assertoric theory of disagreement should be welcomed by all kinds of expressivists.

Since Eriksson focuses on Ridge's (2006a) theory as an exemplar to generalise from, I will too. Ridge suggested moral judgements should be interpreted along these lines:

"There is moral reason to X" expresses (a) an attitude of approval of a certain kind toward actions insofar as they have a certain property and (b) a belief that X has that property. (Ridge 2006a: 315.)

Kantian Ant and Millian Bill are discussing the moral status of a particular murder. Ant says the murder is wrong and – according to Ridge – in saying so he thereby expresses his disapproval of actions forbidden by the categorical imperative and his belief that the murder is one such action. Bill says the murder is right, and in saying so he expresses his approval of actions that promote utility and his belief that the murder will do so. First of all, let us examine the belief components of their judgements.

Ant: The murder is forbidden by the categorical imperative.

Bill: The murder promotes utility.

These beliefs do not conflict. Indeed, we can suppose that both Ant and Bill agree about all the non-moral facts of the murder. Ant fully admits that the murder will lead to good consequences and Bill agrees with Ant that the categorical imperative forbids it. But, quite bluntly, Bill doesn't care about the categorical imperative whereas Ant doesn't wish to disregard it in order to promote utility. Since a prominent argument for expressivism invokes the possibility of moral disagreement persisting despite agreement on non-moral matters, Ridge had better be able to locate disagreement in the non-cognitive component of moral judgements. Let us examine them.

Ant: Disapproval of actions forbidden by the categorical imperative.

Bill: Approval of actions that promote utility.

Eriksson's complaint is that these attitudes are not in conflict at all. It is perfectly possible for actions permitted by the categorical imperative to promote utility. "An instance of optimificity satisfies [Bill]'s attitude, but it does not necessarily frustrate [Ant]'s attitude and vice versa. There is not a clash in attitudes in the required sense. Hence, it seems that [Ant] and [Bill]'s disagreement is merely apparent." (Eriksson 2015: 46.)

There is no conflict between Ant and Bill in their expressed beliefs, nor in their expressed conative attitudes. Accordingly, Eriksson thinks that Ridge and sufficiently similar expressivists do not render this as a case of disagreement. However, to recognise the disagreement, we must consider the *complete set* of all expressed mental states to see that they are mutually incompatible. Ridge's theory is that two mental states are expressed with each moral assertion, not one. We must therefore consider all four states expressed by Ant and Bill at once. When we do, the disagreement becomes apparent: an agent cannot, without inconsistency, adopt all four attitudes expressed with Ant and Bill's assertions. One ought not to, for instance, plan (for all circumstances) on never breaking the categorical imperative *and* on always promoting utility while believing there is a situation in which one cannot do both. Eriksson's argument fails.

4.3 A Quick Remark About Agreement

Agreement is not the negation of disagreement. The negation of disagreement is compatibility. If you say the grass is green and I say the sky is blue, we neither agree nor disagree. The belief I expressed is compatible with yours because both *can* be true together, even if they needn't be. Agreement requires something stronger. I contend that agents agree when they express (or would express) *the same attitudes towards the same contents* with their assertions.

Ridge's old (2006a) view implausibly classified some cases of agreement as cases of mere compatibility. Suppose Bill changes his mind about the murder but not his normative perspective. Bill, when he now asserts that the murder is wrong, expresses his disapproval of acts that promote disutility and his belief that the murder is one such act. But these are not the same attitudes Ant expresses (or would express). There is no necessary connection between the truth of their beliefs or between the satisfaction of their conative attitudes. Their attitudes are merely *compatible*. Since Ant and Bill clearly *agree* when they both assert that the murder is wrong, this is a compelling objection. The problem seems to be that, according to these particular hybrid theories, agents with different normative perspectives appear to be talking past one another. What's needed is a way to make agreeing parties express the same attitudes. A small tweak makes this possible.

Toppinen (2013) suggests that, rather than expressing two states simultaneously, normative sentences express single higher-order states composed of the kinds of lower-order states involved in Ridge's theory:

'There is reason to φ ' expresses the higher-order state of having (1) a desire or desires (of some appropriate kind) for actions insofar as they have some properties K_1, \dots, K_n , and having (2) a belief that φ has one or some of those properties. (Toppinen 2013: 254.)²

Toppinen can help himself to the assertoric theory of disagreement. Suppose one person asserts that there is reason to φ and another asserts there is not. An agent cannot consistently adopt both judgements. "That is, one cannot consistently be in the higher-order state of only having desires for things insofar as they have some properties K_1, \dots, K_n , and having the beliefs that [φ] has one or some of those properties, and none of them. Being in this higher-order state... would involve believing of [φ] that it both has and does not have a certain descriptive property. This would be inconsistent, by anyone's lights." (Toppinen 2013: 275.)

So far, so good. But how does it overcome the problem of agreement? An important part of the higher-order state view is that higher-order states are multiply realisable. Ant and the later Bill, when they both judge that the murder is wrong, are expressing precisely the same higher-order state. Of course, the two components realising this state are different in Ant's head than in Bill's, but that is fairly intuitive. We fully admit that Ant and Bill come to their judgements in different ways and for different reasons. What matters is that Ant and Bill are expressing the *same* mental state, which is the higher-order state of disapproving of the instantiation of certain properties they believe murder to have. Expressing the same attitude is a clear case of agreement.

Since Toppinen's theory can successfully accommodate both disagreement and agreement fairly naturally, anyone sympathetic to the type of hybrid expressivism espoused by early Ridge and Barker ought to adopt higher-order expressivism.³

² For a similar view, see Schroeder (2015: ch.7).

³ Ridge's later work (2014: 7 and 2015b: 8) shows he already has.

Conclusion

I began by arguing against the three main expressivist theories of disagreement. While both Stevenson and Gibbard failed to secure an intuitive extension, Ridge did not adequately address disagreement in compositional judgement. This led to the development of an assertoric theory of disagreement. I argued that parties disagree when they make assertions, or *would* make (sincere) assertions, that express judgements which cannot be simultaneously adopted by a single agent without inconsistency.

This matters for two reasons. First of all, it shows that pure expressivists have no unique problem of disagreement. Any concerns in this domain reduce to a part of the Frege-Geach problem. Secondly, it puts pressure on a family of hybrid theories such as those given by Barker (2000) and Ridge (2006a). The main source of trouble turns out to be *agreement* rather than disagreement, as the theories mistakenly identify many cases of agreement to be cases of mere compatibility. Fortunately this problem can be resolved with no singular losses by moving to the expression of higher-order states. All in all, an assertoric theory of disagreement is good news for expressivism.

~5~

Error and the Limits of Quasi-Realism

If ethical expressivism is true, moral judgements are motivational desire-like states and do not robustly represent reality in special moral ways. This gives rise to the problem of how to understand moral error. How can we be mistaken if there is no distinctive moral reality to be mistaken *about*? The standard expressivist explanation of moral doubt is couched in terms of our fear that our judgements may not survive improvements to our epistemic situation. There is a debate between Andy Egan (2007), Simon Blackburn (2009), Sebastian Köhler (2015) and Michael Ridge (2015b) on the adequacy of this explanation when it comes to the phenomenon of *fundamental* moral error. This is my contribution to the debate.

I argue, contrary to Blackburn and Ridge, that expressivism is committed to some first-order anti-realist sounding claims, and thus quasi-realism fails. If expressivism is true, none of us can coherently believe we might be fundamentally mistaken. However, contrary to Egan and Köhler, I do not think this is bad. Expressivists can still do what motivated the project of quasi-realism in the first place: they can interpret, make sense of, and vindicate *ordinary* moral discourse. I end by showing it yields some positive results for moral philosophers too – it effectively amounts to a transcendental argument against unhealthy moral scepticism.

This chapter has been written as to be independent of all my prior conclusions in this thesis. One might worry that what I say here is not compatible with the “extreme” quasi-realism defended in chapters 1 and 3, according to which moral facts are mind-independent and robustly real

natural facts. Surely we can be fundamentally mistaken about those! I agree that, if scientific realism is true, we might indeed be hopeless at finding out the moral facts. We may be brains in vats and never know which things *really* maximise pleasure (or whatever the moral properties happen to be). But the problem can be reformulated as the problem of knowing *which* natural facts are, or would be, the moral ones. In what follows it is often best to accept the following background assumption: all our relevant robustly descriptive beliefs are true. Since our moral beliefs might still be mistaken, the problem of understanding error remains.

1. The Quasi-Realist Project

As discussed in §2 of chapter 1, contemporary expressivists often want to agree with claims that were once thought definitive of realism. They want to claim there are moral truths, that these truths are mind-independent, and that we can know about them, disagree about them and – as we’ll see in the following section – that we can be *mistaken* about them. The quasi-realist project begins with minimalism about truth. Quasi-realists typically make what Ridge (2014: 200) calls “The Deflationist Gambit” – they put all their eggs in the truth minimalist’s basket in order gain moral truth on the cheap. To summarise the story once more, they propose judging that p is true is not so very different from judging that p . So, if ‘stealing is wrong’ expresses disapproval of stealing, ‘it is true that stealing is wrong’ expresses that very same disapproval. Judging q to be false is no different from judging that not q . Thus, if expressivists can solve the negation problem (Unwin 1999) and provide us with an account of what ‘stealing is not wrong’ expresses, they also have an account of what it is to say it is false that stealing is wrong.

The project is now up and running. If it is true that torture is wrong, it is hard to see why it is not a *fact* that torture is wrong – ‘it is a fact that torture is wrong’ expresses disapproval of torture too. We might also express such disapproval by saying torture exhibits the moral property of wrongness. If beliefs are simply judgements that can be true or false, then moral judgements are beliefs. If ‘torture is wrong regardless of what we think of it’ can be interpreted as the expression of disapproval towards torture in all circumstances, including those in which we like torture, then we can talk of moral properties being mind-independent. Ever and on it

goes, until the quasi-realist can, sincerely and without gritting her teeth, make those very same moral claims that “tempt people to realism.”¹

A pressing worry is that in “appearing in as many ways as possible like realists”² the expressivist *becomes* a realist. Blackburn (1993:15) was alive to this potential upshot of quasi-realism: “I take it that its success would be a measure of the difficulty of defining a genuine debate between realism and its opponents.” This is *the problem of creeping minimalism*. And if the expressivist becomes a realist, she risks inheriting the very same problems.³ Mind-independent moral properties, whether “quasi-” or not, require an explanation for how we can know about them and why they supervene on natural properties.⁴

The quasi-realist wants to be “progressively able to mimic the intellectual practices supposedly definitive of realism.”⁵ We should clarify that the intellectual practices in question are those at the first-order level. The quasi-realist’s project is to show that any ethical statement a realist can say, the expressivist can too. Again, this needs further clarification. Where speech is free, expressivists can *say* anything they like. Anyone can say there are mind-independent moral facts, and they can say the moon is made of cheese. The task is to *vindicate* these claims such that an expressivist, free from incoherence and without self-deception, can assert them sincerely.

I will argue there is a limit to quasi-realism. There are some sentences at the first-order level that a realist can say but a quasi-realist cannot. Minimalism only creeps so far. This means that the quasi-realist project, at least as initially stated by Blackburn, fails. The silver lining is that it provides us with a solution to the problem of creeping minimalism – a difference between the realist and the expressivist shows up at the level of first-order moralising. Furthermore, this difference makes epistemological challenges easier to meet. The area of first-order discourse where quasi-realism fails to mimic realism is that of moral doubt.

¹ Blackburn (1984: 171).

² Blackburn (1993: 17).

³ See Wright (1992) and especially Dreier (2004) who gave the problem of creeping minimalism its name, and see §4.5 of chapter 1 for my solution to it.

⁴ Street (2011) argues that her epistemological Darwinian dilemma, and Dreier (2015b) argues that the supervenience objection, applies just as equally to the quasi-realist as to the realist (although see Toppinen (2018) for an argument that the quasi-realist can answer the supervenience challenge in a way the realist cannot).

⁵ Blackburn (1993: 15).

2. Expressivism and Error

We do not typically think of desires as the type of mental state that can be mistaken. This contrasts starkly with beliefs. If I believe the cat is on the mat when it isn't, I am mistaken. If I desire the cat to be on the mat when it isn't, I am *unsatisfied* but I needn't have made a mistake. Moral judgements are clearly possible to be mistaken. While the early expressivists may have been happy to reject this, *quasi-realist* expressivists seek to vindicate the thought that our judgements might be wrong.

An important part of ordinary moral thought involves doubt. Our moral judgements have been wrong in the past, and there seems to be no reason to suspect we currently have everything right. We might explain error minimally – my judgement that stealing is wrong is mistaken if and only if stealing is *not* wrong. This is fair enough when thinking at the first-order level. It isn't particularly revealing, but it is no less revealing than any other first-order discourse about error. My judgement that grass is green is mistaken if and only if grass is not green. What more could one say? Instead, the problem arises at the second-order level when *interpreting* judgements about first-person potential error. Consider the following sentences and the mental states they supposedly express.

Sentence 1: *I think grass is green, but I might be mistaken.*

Expression 1: [Belief that I believe grass is green] & [Belief that this belief may not accurately represent reality]

Sentence 2: *I think stealing is wrong, but I might be mistaken.*

Expression 2: [Belief that I disapprove of stealing] & [Belief that this disapproval may not...?]

We may worry about the first conjunct in Expression 2. When a robust realist takes herself to judge stealing is wrong, presumably she thinks she is in a robustly representational state. When an ordinary speaker, untrained in metaethics, takes himself to judge stealing is wrong, presumably he has no opinion on what type of mental state that judgement amounts to. There are two ways to proceed. We might say that, if expressivism is true, then ordinary speakers and even realists implicitly take themselves to disapprove of stealing even if it requires a lot of argument before they realise it. The second way is simply to focus on what a conscientious expressivist would think they were up to when judging an opinion of theirs may be false, and whether they could continue to make the judgements in

good faith. The worry is that “proper consciousness of the activity of judgement would unmask and undermine the activity itself.”¹ The quasi-realist, then, owes us an explanation of what it is to believe one’s moral judgement – a desire-like state – might be mistaken, such that acceptance of the explanation does not force us to dismiss that healthy, common sense uncertainty of our moral opinions. The most prominent account is given by Blackburn in his *Ruling Passions* (1998).

2.1 Truth Survives Improvement; Error Does Not

When it comes to giving an expressivist picture of what it is to believe one’s moral judgements may be mistaken, the following passage from *Ruling Passions* has become ubiquitous:

Well, there are a number of things that I admire: for instance, information, sensitivity, maturity, imagination, coherence. I know that other people show defects in these respects, and that these defects lead to bad opinions. But can I exempt myself from the same possibility? Of course not (that would be unpardonably smug). So I can think that perhaps some of my opinions are due to defects of information, sensitivity, maturity, imagination, and coherence. If I really set out to investigate whether this is true, I stand on one part of the (Neurath) boat and inspect the others. (Blackburn 1998: 318.)

The idea is intuitive enough. When I worry that an opinion of mine might be mistaken, I fear for its longevity in the face of improvements to my epistemic situation. My belief that my ethical opposition to eating meat might be in error is my recognition that I might change my mind if I gained more information, or became more imaginative or more coherent.

To clarify, expressivists can be minimalists about what error actually *is*. Moral judgements are erroneous when they’re false, and there’s not much more to be said. To judge somebody else’s opinion to be erroneous is to judge it false. Furthermore, while it is not possible to judge one’s own current opinion as erroneous, we can acknowledge the possibility that we (or anyone else) *might* be in error. On Blackburn’s picture, this judgement amounts to recognising that we may change our minds after various improvements to our epistemic situation.

¹ Blackburn (1993: 3).

2.2 Is There an End to Inquiry?

If you're like me, visual aids can be very helpful. When thinking about this theory of doubt I imagine hypothetical *people* instead of uninstantiated admirable features. I picture a Great Hall of ideal observers, stretching out in several dimensions, lined up according to "scores" denoting their admirable qualities. Let us consider only two dimensions. The observers are arranged such that the ones further north are more descriptively informed than the ones further south, and the ones further east are more sensitive than the ones to the west. Let us now consider a specific judgement: the judgement that it was wrong to invade Iraq in 2003. Now, somewhere in this Great Hall are you and me. We are *instantiated* observers. Suppose I can be found at Sensitivity Level 9 and Information Level 12, with coordinates (SL9, IL12). From this epistemic vantage point, I judge the Iraq invasion to be wrong. Of course I might be mistaken, and in thinking so I approve of a range of observers from the Great Hall – say, all of those further north and east than (SL90, IL95) – and believe that some of them may not judge that it was wrong to invade Iraq. But this is not quite right, for there is a complication.

For simplicity, suppose I am unusually specific in my epistemic admirations. I totally and utterly admire being at (SL97, IL99). Let us call the hypothetical observer with these qualities 'Irene'. Irene is my ideal. The simple case is that when I doubt myself, I acknowledge the possibility that Irene disagrees with me. But Irene may not think she's perfect herself. Irene has her own opinions on improving her epistemic situation. And, if she thinks she ought to be at different coordinates, I think I'd better believe her. What I'm *really* approving of, then, is to be where Irene approves of being. And of course the observer in this position may be unsettled too, thinking *he* may have more epistemic work to do. Eventually, we might arrive at an observer – Edna – who is *settled*. Edna is happy where she is and would not approve of any change to her situation. Edna, we might say, is the end of inquiry.

Must this process to come to end? No. There are two other options. First, inquiry might go on indefinitely, such that observers always judge there is more epistemic work to be done. Every hypothetical observer has an ideal who is not themselves – I admire Irene, Irene admires Jon, Jon admires Karen, and so on forever. And while perhaps Irene disagrees with me about Iraq, maybe Jon is on my side. In this case, what we need is not

an end to the process but a point at which no further inquiry will overturn judgement. If Jon, Karen, and the infinitely many observers they approve of all think the Iraq invasion was wrong, then that's the truth.² If there is no such point, and further ideal inquiry will *always* change our minds, then we will never settle on the truth, since we could always improve and overturn our judgement. The second option is for inquiry to go in a circle. Perhaps Karen actually admires Irene. In this case, everything Irene, Jon and Karen agree on is the truth.

I will assume, again for simplicity, that inquiry comes to an end. In the case we considered above, Edna was the end of *my* inquiry. Two questions naturally arise: "Is Edna my only end?" and "Is she your end too?" Edna is surely not the only *settled* hypothetical observer. We can imagine plenty of people stuck in their ways, disapproving of any change to their situations whatsoever, believing there is no way they can improve. Most of these people are rather unadmirable creatures – creatures exemplified by so-called "ideally coherent eccentrics."³ I hope none of my ideals set me on a path that leads me to become one of *these*. But it is conceivable that Irene might approve of both Jon and Jerry, and that these two observers would take me on diverging paths leading me to different, hopefully admirable, ends of inquiry, all of which have happily settled at different coordinates in the Great Hall. Likewise, there doesn't seem to be any guarantee that your epistemic paths of improvement will lead to Edna. It might be the case that all settled observers happen to agree about everything. But it is the possibility of settling at different ends and yet making different judgements that drives the biggest challenge to quasi-realism.

3. Immunity to Fundamental Error

Andy Egan poses a trilemma for quasi-realist expressivists. He believes Blackburn was right to deny a first-person exemption to error. In general, we should not think of ourselves as having more privileged access to the truth than other people, at least without good *a posteriori* reason to think so. Egan uses the thought that we are not special as his first lemma:

² We are not *identifying* the truth with their judging so, or saying that moral truths are *made* true by the responses of ideal observers. We are simply saying that the best (hypothetical) observers believe truths. No particular order of explanation is implied.

³ Street (2009).

NO SMUGNESS: There isn't any sort of moral error to which others are subject, but against which I have an a priori guarantee of immunity. (Egan 2007: 210.)

Before moving on to Egan's next premise, we should clarify what kind of claim NO SMUGNESS is. Since expressivists are silent about error on the metaethical second-order level, Egan's premise must be a first-order claim: it is the claim that, *prima facie*, my own epistemic situation is *no better* than anyone else's. It is an evaluative claim, which on the expressivist picture is a desire-like attitude. This point is not drawn out in the literature. As a first-order claim, the expressivist is not committed to it. But, if quasi-realism is to succeed, it must be possible to vindicate the claim as *making sense* on the expressivist picture. And that means it must be possible to assign a coherent desire-like state that constitutes acceptance of the claim.

Egan's next step is to define *stability*. A given judgement of mine is stable if it would survive all changes I see as improvements. Ridge (2015b) helpfully clarifies different possible readings of stability. Take a series of changes to one's epistemic situation. Perhaps the first stage was to gain more coherence and the second stage was to gain more information. One may endorse (or not) these changes before the whole process, after the whole process, or at each stage. The best way to understand stability is in terms of endorsing each stage of changes at the moment they happen.¹ Stable judgements are those that will survive any such process of endorsed-at-the-time changes. In terms of my hypothetical ideal observers, the judgements of mine that are stable are the ones that all my ideal observers on the path to Edna (including Irene, Jon, Karen, and so on) agree with me about.

Are any of my judgements stable? I don't know for sure, but perhaps. Maybe there's no way to convince me (in ways that I endorse) to abandon my judgement that it is wrong to torture innocent people for no comparable gain in pleasure, happiness or desire satisfaction. Call this judgement 'J'. Now, it seems possible that somebody else could judge that *not J*, and for this judgement of theirs to be stable. Many philosophers, both realists and anti-realists, believe that ideally coherent eccentrics are a theoretical possibility. That is, there are potential people whose beliefs and desires are

¹ Ridge (2015b: 3) explains the 'each stage' reading of a judgement's stability is best because it is the only way to ensure that "there will be no rational way for the believer to abandon it."

internally coherent but particularly deviant. Imagine an ideally coherent Caligula who judges that torturing people is the thing to do, and that no series of endorsed-at-the-time changes would lead him to abandon that judgement.² It might be possible to change Caligula's mind in ways he *doesn't* endorse, such as by hypnosis or a bump on the head, but this has no bearing on stability, which only makes reference to changes that are thought to be improvements by the agent's lights. According to our hypothesis Caligula stably judges that J is false. But J is surely true, and so we arrive at Egan's second lemma:

FUNDAMENTAL FALLIBILITY: It's possible for people's stable moral beliefs to be mistaken. (Egan 2007: 213.)

Egan then considers whether we could be fundamentally mistaken ourselves and reaches his third lemma:

For me to be fundamentally in error, I need to have some moral view that's (a) stable, and (b) mistaken. But given Blackburn's account of moral error, this can't happen. For my moral belief that P to be stable is for it to be such that it would survive any improving change (or course of improving changes). For my moral belief that P to be mistaken is for there to be some improving change (or course of improving changes) that would lead me to abandon P. So on Blackburn's account of moral error, a moral belief is mistaken only if it's not stable. So for me to be fundamentally in error, I'd need to have some moral view that was (a) stable, and (b) not stable, which I pretty clearly can't have. ... So the quasi-realist is committed to:

FIRST-PERSON IMMUNITY: I have an a priori guarantee against fundamental moral error. (Egan 2007: 214.)

Egan's three lemmas contradict. If people such as the ideally coherent Caligula can have FUNDAMENTAL FALLIBILITY while individual expressivists are committed to FIRST-PERSON IMMUNITY, this violates NO SMUGNESS. There is a type of error that I know *a priori* I do not suffer from, but I cannot say the same of anyone else. As an expressivist I must judge that other people may be cut off from the moral truth in a way that I am not. And this seems unpardonably smug of me.

² The example of an ideally coherent Caligula who "aims solely to maximise the suffering of others" comes from Gibbard (1999: 145).

3.1 Blackburn's Reply & Köhler's Rejoinder

In Blackburn's (2009) response, he identifies the ways in which Egan's argument misrepresents the quasi-realist's commitments. Blackburn (2009: 205-6) makes the distinction between "(M) If something is entrenched in my outlook, in such a way that nothing I could recognize as an improvement would undermine it, then it is true" and "(I) If something is entrenched in my outlook, in such a way that nothing that is an improvement would undermine it, then it is true." He states that he agrees with I but disagrees with M, and M is what's needed for Egan's argument.³

Blackburn's response is important. He clarifies the quasi-realist position and what commitments he associates himself with, most notably the fact that the expressivist is silent on what moral error *is* – that is a normative question, and "instability" is an implausible answer for anyone without pragmatist leanings. But Köhler (2015: 163) argues the reply misses the mark, since Blackburn does not attempt to give a coherent state of mind expressed by the sentence 'I might be stably mistaken.' Egan's problem is best understood as a *challenge* for expressivists to characterise this judgement in expressivist-friendly terms.

It is clearly possible to think somebody *else's* opinion O is stable and yet might be mistaken. This consists in the belief that the agent's judgement O will survive any series of changes to her epistemic situation that *she* considers improvements at the time, along with the judgement that O might be false – where this is a complex state of (i) approving of end-of-inquiry-Edna and (ii) the belief that Edna may not judge that O. Köhler thinks expressivists cannot consistently believe this of themselves, revivifying FIRST-PERSON IMMUNITY and violating NO SMUGNESS. For to believe that I might be *fundamentally* mistaken in my judgement J is to think that J is stable, such that no series of changes I consider improvements will lead me to abandon it, and yet to also approve of Edna and think that I might abandon J after I become like her.

Consequently, the very thought that I might myself be in fundamental error is *inconsistent*. So, I *do* have an *a priori* guarantee against this kind of error. Expressivists are still committed to FIRST-PERSON IMMUNITY. Quasi-realism is still threatened by Egan's challenge. (Köhler 2015: 165.)

³ See Egan (2007: 214) for the slip between what Blackburn calls "M" and "I".

3.3 Ridge's Distinction: 'Could' or 'Would'?

Look closer at the interpretation on offer:

Sentence 3: *My judgement J is stable, but might be mistaken.*

Expression 3: [Belief that J will survive all changes I consider improvements] & [Approval of a process of changes] & [Belief that J might not survive the (whole) process of changes I expressed approval of]

Ridge (2015b) shows that, contrary to what Köhler says, this judgement is consistent. It is perfectly fine to think that J *might* not survive certain changes while thinking that J in fact *will* survive those changes. Thoughts of this kind are easy to come by. I believe the sun will rise tomorrow while at the same time recognising the (small) possibility that the earth will stop turning and I'll be stuck on the dark side. In light of this, Ridge (2015b: 17) constructs a version of Egan's argument couched in terms of "hyper-stability." To judge something hyper-stable is to be certain it is stable. Ridge's new problem involves this judgement that one's opinion *must* survive all possible changes one considers improvements.

Sentence 4: *My judgement J is 100% certainly stable, but it might be mistaken.*

Expression 4: [Belief that J *must* survive all changes I consider improvements] & [Approval of a process of changes] & [Belief that J might not survive the (whole) process of changes I expressed approval of]

This is now enough to guarantee a contradiction between the two belief components. It is the contradiction that Egan was looking for in order for his trilemma to apply. If this is the correct expressivist account of error, the conscientious expressivist cannot sincerely assert Sentence 4, as they would be aware that it expresses two beliefs that cannot both be true. (They know that the changes they consider improvements are precisely those they express approval of.) If the conscientious expressivist believes a judgement of theirs is hyper-stable (100% certainly stable), they are forced to say their judgement is immune to error. FIRST-PERSON IMMUNITY applies, and this is smug.

3.4 This is Smug, But So What?

Ridge (2015b: 18) agrees that this is smug, but he thinks this isn't a problem with expressivism. Instead, it's a problem with anyone who makes the judgement of hyper-stability. Anyone who is 100% certain that a judgement of theirs is stable is being very smug! How can anyone sensibly rule out the possibility of changing their minds about any of their judgements? We would think such a person smug whether we're expressivists or not. The vice of smugness arises not from the expressivist interpretation of the judgement, but from the agent's absolute certainty. All it takes is a mere flicker of humility to recognise that a judgement of yours, while it may be *likely* to be stable, has a chance of being abandoned after what you see as improving changes.

Ridge's strategy is effectively to bite the bullet: expressivism *does* imply that someone who thinks a judgement of hers is hyper-stable must also think she is not fundamentally mistaken about it (and she need not think the same of *other* people's hyper-stable judgements), and this agent is unpardonably smug. However, she is *already* smug to be so certain that there is no chance she would ever change her mind after any possible series of changes that she sees as improvements at the time. So it seems that all Egan and Köhler have going for them is the worry that someone, who smugly thinks one of her judgements is hyper-stable, must also smugly think she cannot be fundamentally mistaken. "In that case, though, the conclusion of the argument would only be that being both an expressivist and independently of one's expressivism being unpardonably smug entails that you are unpardonably smug."⁴ And that doesn't seem like a worry at all.

Beware the equivocation of the word 'smug'. It may be an epistemic vice to be absolutely certain that you will never change your mind after any possible series of seen-by-you-as-improvements, but this isn't smugness as defined by Egan in his trilemma. Egan's 'smugness' is exemplified by an agent who thinks he has a special *a priori* guarantee of first-person immunity from fundamental error that nobody else has. The vice Ridge discusses is different, and we should give it a different name. I'll call it 'misplaced certainty.' The conclusion of Egan's argument, contrary to what

⁴ Ridge (2015b: 20).

Ridge argues, is not that expressivism plus smugness entails smugness. The conclusion is that expressivism plus *misplaced certainty* entails smugness.

The reader may wonder why this matters. If expressivism is true, it means that one vice entails another – no biggie. But, first of all, it is an entailment that does not hold for a realist. Secondly, the problem can be reinstated even for agents that do not exemplify the vice of misplaced certainty.

3.5 The Anti-Realist Conditional

I do not rule many things out entirely, and nor should you. But this includes the possibility that one of our judgements is actually stable in the sense that, while unbeknownst to us, there really are no possible changes we would endorse at the time that would lead us to abandon that particular judgement. Given that this is a possibility, we can think about what would follow from it. I shall argue the conditional we are forced to accept, on the assumption that expressivism is true, is this:

Anti-Realist Conditional (ARC): Necessarily, if my judgement J is stable, then it is true.

No realist would accept this conditional. Expressivists must. This is because an expressivist cannot consistently believe its negation – that there is a possibility that their judgement J is both stable and false. Let us bring out why.

Sentence 5: *It is possible my judgement J is false and stable.*

Expression 5: [Approval of a process of changes] & [Belief that it is possible that: J might not survive the (whole) process of changes I expressed approval of *and* J must survive all changes I consider improvements]⁵

Sentence 5 expresses an inconsistent belief because, as Blackburn knows, the changes I express approval of *just are* the changes I consider improvements. It is therefore inconsistent to deny ARC. The inconsistency is hidden in the psychology, not something that can be read off the sentence. Nevertheless, insofar as we are to eliminate inconsistencies from our beliefs

⁵ The first modal operator is epistemic whereas the second and third are metaphysical. It is the belief that, for all I know, there is and is not a metaphysical chance that J won't survive until the end of inquiry. Thanks to Jimmy Lenman for this point.

we must believe ARC, even without misplaced certainty that any of our judgements are *actually* stable. We smugly have to believe we have an *a priori* first-person immunity from moral opinions that are stably false. Egan was right after all.

4. The Limits of Quasi-Realism

Instances of the Anti-Realist Conditional are first-order sentences. If expressivism is true, a “proper consciousness”¹ of the mental states these conditionals express forces us to accept them. Expressivists must accept that necessarily, if a judgement of theirs is “locked” into their perspective, then it is true. What does this mean for expressivism? It means that expressivism commits us to some first-order claims. It means that quasi-realism, at least on my reading of Blackburn, fails. What we end up with is some form of quasi-anti-realism. Exactly what form this is will depend on the first-order choices the expressivist can make.

We have seen that, despite arguments from Blackburn and Ridge to the contrary, the expressivist is committed to FIRST-PERSON IMMUNITY, the lemma that she herself has an *a priori* guarantee against fundamental moral error. This yields a choice: reject NO SMUGNESS or give up FUNDAMENTAL FALLIBILITY. To choose the latter is to judge that nobody can be fundamentally mistaken. On this view, nobody stably believes anything false – there’s hope for everyone. And there are two ways to accept this. Firstly, we could be optimistic about the idea of eventual convergence of opinion. Perhaps everyone will reach Edna at the end of their ideal inquiry. Eccentrics would never be *settled* – there would always be changes they would be happy to call improvements to their situation that would eventually lead them to Edna. This, we might say, is a kind of quasi-Kantianism.

The second way to give up FUNDAMENTAL FALLIBILITY is to relativise truth. Settled eccentrics are possible, but they are not mistaken. Their judgements are correct; they’re just different. When we talk of the moral truth, we’re talking about what we’ll find at the end of *our* inquiries, perhaps hoping our interlocutor is on a similar path. But when we consider strange creatures who moralise in very different ways and with whom we

¹ Blackburn (1993: 3).

cannot hope to have a productive moral argument, we don't *disagree* with them as such.² What we end up with here is a kind of quasi-relativism.

For my part, I find it plausible that settled ideally coherent eccentrics are possible but I also believe they can be wrong. No matter how strange a settled ideally coherent Caligula would be, I cannot bring myself to think he speaks the truth when he says it is right to torture innocent people for no comparable gain in pleasure, happiness or desire satisfaction. My favoured route is to reject NO SMUGNESS and accept that I have an *a priori* guarantee that I can correct any of my mistakes in ways I see as improvements, and yet I cannot say the same of anyone else, such as the settled Caligula. Luckily I can know *a posteriori* that other people are not cut off from the truth. As it happens, I haven't met a single person I believe *is* in such a hopeless situation. But, since I must inspect each new person I meet (including my past and future selves) to check they are not like Caligula, I am the only person that I can know *a priori* has IMMUNITY from being stably mistaken. Perhaps we can call this a kind of quasi-idealism.

4.1 This Isn't So Bad

The language in this particular area of philosophy can be misleading. Egan (drawing from Blackburn's text) gives us a very specific definition of 'smugness' that is important to distinguish from common parlance. Even though I personally reject NO SMUGNESS, I do not have any *misplaced certainty* in any of my opinions. I believe even my most firmly held judgements might be wrong. Of course, given my commitment to ARC, I do not believe I am cut off from the truth. Wherever I'm wrong, I must believe there is a *possibility* that I can improve enough to access the moral facts. But then again, I do not believe that you are cut off either. I believe you can access the moral truth just as well as I can. And I believe there are many people who are much better at ethical inquiry than I am and from whom I can learn a lot.

The word 'fundamental' is misleading too. In Egan's sense, to be fundamentally mistaken is to have an irreparably false judgement that you cannot get rid of through any process you'd be willing to call improvement. It is to be *severed* from the moral truth in a most heinous way. It is not simply

² This echoes the "relativism of distance" of Williams (1985: ch.9).

to be mistaken about fundamental truths. Given that I believe *any* of my moral judgements are up for grabs – I can't know for sure what will result from the process of moral inquiry before I've gotten anywhere near finishing it – I also believe I might be mistaken about the fundamentals.

In short, I instantiate SMUGNESS but (I hope) I'm not *smug*, and while I cannot coherently believe I might be FUNDAMENTALLY mistaken, I very much do believe I might be *fundamentally* mistaken about any of my dearest and most firmly held judgements. So why is this supposed to be bad for the expressivist? Egan (2007: 215) thinks it is because "NO SMUGNESS captures an important part of our ordinary way of thinking about morality." But who is he referring to? Ordinary people do not think about the *a priori*. They do think about error, and presumably they think they can go wrong in the same ways as everyone else they've met – they do not think they're special in having privileged access to the truth. But, as I have explained, I do not think I'm special either. In my experience, people tend to have the same kinds of faculties as I do. I know this *a posteriori*, but how else could I find out whether I'm a special or a common creature around here?

What was initially attractive about Blackburn's quasi-realist project was its mission of vindicating ordinary moral thought and discourse. Recognising the truth of expressivism should not prevent us from making the same kinds of judgements we made before philosophy distracted us. Egan (2007: 216) thinks he has shown otherwise: "If quasi-realists are obliged to reject NO SMUGNESS, then they won't be able to just go on as before – being a quasi-realist will force a major revision of our ordinary ways of thinking and talking about moral matters." Again, I strongly disagree. When I reject NO SMUGNESS, I'm going to continue *exactly* as I was before. As long as I do not have any *misplaced certainty* in any of my particular opinions, I will continue to doubt myself at appropriate times and inspect the Neurath boat while it's at sea. This vindicates *ordinary* moral practice, and that's all we as expressivists are required to do.

Egan's challenge shows that, contrary to Blackburn's hopes, expressivism does commit us to some first-order claims that a realist is able to reject. But this isn't so bad. Acceptance of Anti-Realist Conditionals does not force us to see ourselves as superior, it leaves us plenty of room for healthy scepticism, and it allows us to vindicate everything to be found in ordinary moral discourse.

4.2 Silver Linings

I have discussed why the limits to the quasi-realist project are not bad for expressivism. I will end with reasons to welcome them. To begin, it yields another answer to the problem of creeping minimalism: there are first-order claims that a realist can make but an expressivist cannot. More importantly, it means expressivists evade some of the epistemological challenges that plague realists. For imagine what it would be like if we thought the ideally coherent Caligula was just as likely to be right as we are. If I genuinely believed that, I do not think I could go on as before. It would be *realism* that was the threat to ordinary moral thought.

Moral realists I'm sure would deny that Caligula is just as likely to be right as we are. After all, Caligula is an awful person. He thinks we should torture people for the sole reason of maximising suffering! The moral realist might say we can find out *a posteriori* that Caligula's moral faculties are faulty. But how do we know it is Caligula's faculties at fault and not our own? Why think that *we* have access to the realm of moral facts? In answering this question, expressivists have a weapon realists don't: we think it because we must.

You might fear this makes it mysterious why the moral truth hangs round people like us and not people like Caligula. What's so special about us? How can it be reassuring to think I can access the moral truth *just because I'm me*?³ In response, I would first reiterate there is no commitment to the moral truth being in any way mind-dependent. The reason torture is wrong need not have anything to do with me or my moral sensibility, and judgements about what the moral truth depends on are to be assessed from the first-order level. I would then emphasise that the same applies to our epistemic access: the *reason* I can access the moral truth is not that I'm me, but that I possess certain features that make me a somewhat decent moral inquirer. Caligula lacks these features. Judgements about what makes a good inquirer (and what makes a terrible one) are to be assessed from the first-order level, and *the fact that I'm me* is a bad answer. If these judgements are first-order, what progress has been made? Well, what I have argued in this chapter is that if expressivism is true, we cannot coherently think of ourselves as hopeless inquirers. Radical scepticism undermines itself, and

³ Thanks to Elinor Mason and Alex Gregory for raising this concern and to Daniel Elstein for helpful discussion of the response to it.

so expressivists do not have to answer radically sceptical worries. Therein lies the epistemic advantage of expressivism over realism.

In summary, on the expressivist picture I can doubt the truth of every single one of my judgements. What I cannot do, on pain of inconsistency, is believe my false ones are incapable of correction. If the possibility of being severed from the truth is what Egan is fighting for, I want no part of it. It is far healthier to always believe we might be mistaken, and to always believe we can correct our mistakes.

Conclusion

I have argued that quasi-realism doesn't stretch all the way. There are some first-order claims an expressivist cannot coherently deny. Far from being a reason to reject expressivism, this is a reason to favour it, as the expressivist has the beginnings of an answer to the moral sceptic. What the moral truth *actually is*, expressivists do not say. Ask an ethicist – we remain silent on the matter. But what we *can* say to ourselves is this: find stability and you'll find truth. How do we find stability? In much the same ways as before: collect evidence, cohere your thoughts, become more sympathetic, reflect. *Improve*. Moral truth needn't be identified with stability, but we do have to believe our stable beliefs are true. After all, stability is the end of inquiry. This is not a very realist thing to say, and it is a limit on quasi-realism, but it is no bad thing. It only means expressivists must accept that, at the end of inquiry, *truth will out*. To think otherwise is to abandon hope.

Summary

Here at last we come to the end of my thesis. I have defended expressivism against the threat of creeping minimalism, thus maintaining it is in fact a distinctive position in metaethics – and without any strong commitments in the philosophy of language to boot. I have argued that the most promising form of expressivism is *higher-order expressivism*, for it can make sense of complex moral thought as well as abstract claims about moral properties. I ended with two of the biggest problems for expressivism: understanding moral disagreement and the possibility of error. I argued for theories of disagreement and fallibility that allow us to preserve everything that can be found in ordinary moral discourse.

The picture of morality that emerges is flexible and optimistic. It is flexible because it leaves much to be assessed at the first-order level. If you are like me you'll want to insist that moral truths are mind-independent, but if you think morality is constructed by us (or whatever), then we can have that debate and remain expressivists. It is a debate of *value*. The only thing off the table is a *sui generis* realm of irreducibly moral facts. Beyond that, game on.

I ended with reasons to be optimistic. Again, if you're like me you'll favour the mind-independence of morality, but this provokes the worry that moral truths are inaccessible to us. I have argued this worry is unstable. We should not have *misplaced certainty* in our moral opinions but radical scepticism is fundamentally incoherent. If expressivism is true, we have to believe we're on the right track.

And that, I hope, is a comforting thought.

References

- Adams, E. W., 1965, "A Logic of Conditionals," *Inquiry*, 8: 166–197.
- Alwood, A., 2016, "Should Expressivism be a Theory at the Level of Metasemantics?" *Thought: A Journal of Philosophy*, 5(1): 13–22.
- Anscombe, G. E. M., 1957, *Intention*, Oxford: Blackwell.
- Asay, J., 2013, "Truthmaking, Metaethics, and Creeping Minimalism," *Philosophical Studies*, 163(1): 213–232.
- Ayer, A. J., 1935, "The Criterion of Truth," *Analysis*, 3(1-2): 28–32.
- , 1936, *Language, Truth, and Logic*, London: Victor Gollancz Ltd.
- Baker, D., 2018, "Skepticism about Ought *Simpliciter*," in R. Shafer-Landau, *Oxford Studies in Metaethics*, vol. 13, Oxford: Oxford University Press.
- Baker, D. and Woods, J., 2015, "How Expressivists Can and Should Explain Inconsistency," *Ethics*, 125: 391–424.
- Barker, S., 2000, "Is Value Content a Component of Conventional Implicature?" *Analysis*, 60(3): 268–279.
- Barlassina, L. and Hayward, M., 2018, "More of Me! Less of Me! Reflexive Imperativism about Affective Phenomenal Character," manuscript under review.
- Bedke, M., 2009, "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly*, 90(2): 188–209.
- Bex-Priestley, G. and Shemmer, Y., 2017, "A Normative Theory of Disagreement," *Journal of the American Philosophical Association*, 3(2): 189–208.
- Björnsson, G., Olinder, C., Eriksson, J. and Björklund, F. (eds.), 2015, *Motivational Internalism: Contemporary Debates*, Oxford University Press.

Blackburn, 1981, "Rule-Following and Moral Realism," in S. Holtzman and C. M. Leich (eds.), *Wittgenstein: To Follow a Rule*, London: Routledge & Kegan Paul.

---, S., 1984, *Spreading the Word*, Oxford: Oxford University Press.

---, 1988, "Attitudes and Contents," *Ethics*, 98(3): 501–517.

---, 1993, *Essays in Quasi-Realism*, New York: Oxford University Press.

---, 1995, "Justification, Scepticism, and Nihilism," *Utilitas*, 7(2): 237–246.

---, 1998, *Ruling Passions*, Oxford: Clarendon Press.

---, 2009, "Truth and A Priori Possibility: Egan's Charge Against Quasi-Realism," *Australasian Journal of Philosophy*, 87(2): 201–213.

---, 2010, *Practical Tortoise Raising, and Other Philosophical Essays*, Oxford: Oxford University Press.

---, 2015, "Blessed are the Peacemakers," *Philosophical Studies*, 172(4): 843–853.

Block, N., 1986, "Advertisement for a Semantics for Psychology," in P. French, T. Uehling and H. Wettstein (eds.), *Studies in the Philosophy of Mind, Midwest Studies in Philosophy* (Volume 10), Minneapolis: University of Minnesota Press.

Boisvert, D., 2008, "Expressive-Assertivism," *Pacific Philosophical Quarterly*, 89(2): 169–203.

Bolland, B. (w, a), 1996, "An Innocent Guy," *Batman: Black and White*, 1(4), DC Comics.

Boyd, R., 1988, "How to be a Moral Realist," in G. Sayre-McCord (ed.), *Essays on Moral Realism*, Ithica: Cornell University Press.

---, 2003, "Finite Beings, Finite Goods: The Semantics, Metaphysics, and Ethics of Naturalist Consequentialism, Part II," *Philosophy and Phenomenological Research*, 67(1): 24–47.

Brink, D., 1989, *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.

---, 2001, "Realism, Naturalism, and Moral Semantics," *Social Philosophy and Policy*, 18(2): 154–176.

Brandom, R., 1994, *Making it Explicit: Reasoning, Representing and Discursive Commitment*, Cambridge, MA: Harvard University Press.

---, 2000, *Articulating Reasons*, Cambridge, MA: Harvard University Press.

- Buckwalter, W. and Turri, J., 2015, "Inability and Obligation in Moral Judgement," *PLoS One*, 10(8): E0136589.
- Burgess, A. and Sherman, B., 2014, *Metasemantics: New Essays on the Foundations of Meaning*, Oxford University Press.
- Carnap, R., 1935, *Philosophy and Logical Syntax*, London: Kegan Paul, Trench, Trubner & Co.
- Chrisman, M., 2015, *The Meaning of 'Ought': Beyond Descriptivism and Expressivism in Metaethics*, Oxford University Press.
- Copp, D., 1997, "The Ring of Gyges: Overridingness and the Unity of Reason," *Social Philosophy and Policy*, 14(1): 86–101.
- , 2001, "Realist Expressivism: A Neglected Option for Moral Realism," *Social Philosophy and Policy*, 18(2): 1–43.
- , 2009, "Realist-Expressivism and Conventional Implicature," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 4, Oxford: Oxford University Press.
- , 2014, "Can a Hybrid Theory Have it Both Ways? Moral Thought, Open Questions and Moral Motivation," in G. Fletcher and M. Ridge (eds.), *Having it Both Ways: Hybrid Theories and Modern Metaethics*, Oxford: Oxford University Press.
- , 2018, "Realist-Expressivism and the Fundamental Role of Normative Belief," *Philosophical Studies*, 175(6): 1333–1356.
- Cresswell, M. J., 1985, *Structured Meanings*, Cambridge, MA: MIT Press.
- Cuneo, T., 2014, *Speech and Morality: On the Metaethical Implications of Speaking*, Oxford: Oxford University Press.
- Cuneo, T. and Shafer-Landau, R., 2014, "The Moral Fixed Points: New Directions for Moral Nonnaturalism," *Philosophical Studies*, 171(3): 399–433.
- Davidson, D., 1973, "Radical Interpretation," *Dialectica*, 27(3-4): 313–328.
- Davis, W., 2003, *Meaning, Expression, and Thought*, Cambridge: Cambridge University Press.
- Dennett, D., 1987, *The Intentional Stance*, Cambridge, MA: MIT Press.
- Dreier, J., 1996, "Expressivist Embeddings and Minimalist Truth," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 83(1): 29–51.
- , 2004, "Metaethics and the Problem of Creeping Minimalism," *Philosophical Perspectives*, 18: 23–44.
- , 2009, "Relativism (and Expressivism) and the Problem of Disagreement," *Philosophical Perspectives*, 23(1): 79–110.

---, 2015a, "Truth and Disagreement in Impassioned Belief," *Analysis*, 75(3): 450–459.

---, 2015b, "Explaining the Quasi-Real," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 10, Oxford: Oxford University Press.

Dummett, M., 1981, *Frege: Philosophy of Language*, 2nd Edition, London: Duckworth.

Dworkin, R., 2011, *Justice for Hedgehogs*, Cambridge, MA: Harvard University Press.

Edgington, D., 1995, "On Conditionals," *Mind: A Quarterly Review of Philosophy*, 104(414): 235–329.

Egan, A., 2007, "Quasi-Realism and Fundamental Moral Error," *Australasian Journal of Philosophy*, 85(2): 205–219.

Enoch, D., 2011, *Taking Morality Seriously: A Defense of Robust Realism*, Oxford: Oxford University Press.

Eriksson, J., 2009, "Homage to Hare: Ecumenism and the Frege-Geach Problem," *Ethics*, 120: 8–35.

---, 2015, "Explaining Disagreement: a Problem for (Some) Hybrid Expressivists," *Pacific Philosophical Quarterly*, 96: 39–53.

Field, H., 1980, *Science Without Numbers*, Princeton, NJ: Princeton University Press.

---, 1994, "Disquotational Truth and Factually Defective Discourse," *Philosophical Review*, 103(3): 405–52.

Fine, K., 2001, "The Question of Realism," *Philosophers' Imprint*, 1(1).

Finlay, S., 2014, *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press. .

---, 2017, "Disagreement Lost and Found," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 12, Oxford: Oxford University Press.

Fitzpatrick, W. J., 2016, "Ontology for an Uncompromising Ethical Realism," *Topoi*, 1–11.

Geach, P., 1960, "Ascriptivism," *Philosophical Review*, 69: 221–225.

---, 1965, "Assertion," *Philosophical Review*, 74: 449–465.

Gibbard, A., 1990, *Wise Choices, Apt Feelings*, Cambridge: Harvard University Press.

---, 1993, "Reply to Blackburn," *Philosophical Issues*, 4: 67–73.

---, 1999, "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics*, 110: 140–164.

---, 2003, *Thinking How to Live*, Cambridge: Harvard University Press.

---, 2011, "How Much Realism? Evolved Thinkers and Normative Concepts," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 6, Oxford: Oxford University Press.

---, 2012, *Meaning and Normativity*, Oxford: Oxford University Press.

Graham, P., 2011, "'Ought' and Ability," *Philosophical Review*, 120(3): 337–382.

Gregory, A., 2012, "Changing Direction on Direction of Fit," *Ethical Theory and Moral Practice*, 15(5): 603–614.

Grice, H. P., 1989, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Hale, B., 1993, "Can There be a Logic of Attitudes?" in J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection*, New York, NY: Oxford University Press.

Hare, R. M., 1952, *The Language of Morals*, Oxford: Oxford University Press.

---, 1967, "Some Alleged Differences Between Imperatives and Indicatives," *Mind*, 76: 309–326.

---, 1970, "Meaning and Speech Acts," *The Philosophical Review*, 79(1): 3–24.

Harman, G., 1973, *Thought*, Princeton, NJ: Princeton University Press.

---, 2009, "Guilt-Free Morality," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 4, Oxford: Oxford University Press.

Horgan, T. and Timmons, M., 1991, "New Wave Moral Realism Meets Moral Twin Earth," *Journal of Philosophical Research*, 16: 447–465.

---, 1992a, "Troubles on Moral Twin Earth: Moral Queerness Revived," *Synthese*, 92: 221–260.

---, 1992b, "Troubles for New Wave Moral Semantics: The Open Question Argument Revived," *Philosophical Papers*, 21: 153–175.

---, 1996, "From Moral Realism to Moral Relativism in One Easy Step," *Critica*, 83: 3–39.

---, 2006, "Cognitivist Expressivism," in T. Horgan and M. Timmons (eds.), *Metaethics After Moore*, Oxford University Press.

---, 2015a, "Exploring Intuitions on Moral Twin Earth: A Reply to Sonderholm," *Theoria*, 81(4): 355–375.

---, 2015b, "Modest Quasi-Realism and the Problem of Deep Moral Error," in R. N. Johnson and M. Smith (eds.), *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, Oxford University Press.

Horwich, P., 1998, *Truth* (2nd ed.), Oxford: Clarendon.

---, 2004, *From a Deflationary Point of View*, Oxford: Clarendon Press.

Hurley, S., 1989, *Natural Reasons: Personality and Polity*, New York: Oxford University Press.

Ingram, S., 2017, "I Can't Relax! You're Driving me Quasi!" *Pacific Philosophical Quarterly*, 98(3): 490–510.

Jackson, J., 1979, "On Assertion and Indicative Conditionals," *Philosophical Review*, 88: 565–589.

---, 1987, *Conditionals*, Oxford: Basil Blackwell.

---, 1998, *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Clarendon.

Jackson, J. and Pettit, P., 1995, "Moral Functionalism and Moral Motivation," *The Philosophical Quarterly*, 45(178): 20–40.

---, 1998, "A Problem for Expressivism," *Analysis*, 58: 239–251.

James, W., 1907, "Pragmatism's Conception of Truth," *The Journal of Philosophy, Psychology and Scientific Methods*, 4(6): 141–155.

Johnson, M. and Nadow, J., 2014, "Moderate Intuitionism: A Metasemantic Account," in P. Booth and A. Rowbottom (eds.), *Intuitions*, Oxford University Press.

---, 2017, "Actual vs. Counterfactual Dispositional Metasemantics: A Reply to Andow," *Philosophia*, 45(2): 717–734.

Joyce, R., 2001, *The Myth of Morality*, Cambridge: Cambridge University Press.

---, 2016, "Moral Anti-Realism," *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), E. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/>](https://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/).

Kalderon, M., 2005, *Moral Fictionalism*, Oxford: Clarendon Press.

Kaplan, D., 1978, "Dthat," in P. Cole (ed.), in P. Cole (ed.), *Syntax and Semantics, Vol. 9, Pragmatics*, New York: Academic Press.

---, 1979, "On the Logic of Demonstratives," *Journal of Philosophical Logic*, 8(1): 81–98.

---, 1989, "Demonstratives," in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press.

Köhler, S., 2015, "What is the Problem with Fundamental Moral Error?" *Australasian Journal of Philosophy*, 93(1): 161–165.

---, 2017, "Expressivism, meaning, and all that," *Canadian Journal of Philosophy*, 1–20.

Korsgaard, C., 1996, *The Sources of Normativity*, Cambridge: Cambridge University Press.

Kripke, S., 1980, *Naming and Necessity*, Cambridge, MA: Harvard University Press.

---, 1982, *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Oxford: Blackwell.

Lance, M., and O'Leary-Hawthorne, J., 1997, *The Grammar of Meaning: Normativity and Semantic Discourse*, Cambridge: Cambridge University Press.

Laskowski, N., 2017, "The Sense of Incredibility in Ethics," *Philosophical Studies*, DOI: 10.1007/s11098-017-1007-1.

Lenman, J., 2003, "Disciplined Syntacticism and Moral Expressivism," *Philosophy and Phenomenological Research*, 66(1): 32–57.

---, 2008, "Against Moral Fictionalism," *Philosophical Books*, 49(1): 23–32.

---, 2012, "Expressivism and Constructivism," in J. Lenman and Y. Shemmer (eds.), *Constructivism in Practical Philosophy*, Oxford: Oxford University Press.

---, 2013, "Ethics Without Errors," *Ratio*, 26(4): 391–409.

---, 2014, "Gibbardian Humility: Moral Fallibility and Moral Smugness," *Journal of Value Inquiry* 48: 235–245.

---, 2018, "An Ecumenical Matter?" *The Philosophical Quarterly*, DOI: 10.1093/pq/pqy022.

Lewis, D., 1969, *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

---, 1974, "Radical Interpretation," *Synthese*, 23: 331–344.

---, 1983, "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, 61(4): 343–377.

---, 1984, "Putnam's Paradox," *Australasian Journal of Philosophy*, 62(3): 221–236.

---, 1986a, *On the Plurality of Worlds*, Oxford: Blackwell.

---, 1986b, *Philosophical Papers* (Volume 2), Oxford: Oxford University Press.

---, 1989, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society, Supplementary Volumes*, 63: 113–137.

---, 1991, *Parts of Classes*, Oxford: Basil Blackwell.

Mackie, J. L., 1946, "A Refutation of Morals," *Australasian Journal of Philosophy* 24(1-2): 77–90.

---, 1977, *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin Books.

Marsili, N., 2015, "Normative Accounts of Assertion: From Peirce to Williamson and Back Again," *Rivista Italiana di Filosofia del Linguaggio* ISSN 2036-6728: 112–30.

McDowell, J., 1979, "Virtue and Reason," *Monist*, 62: 331–50.

McGrath, M. and Frank, D., 2018, "Propositions," *The Stanford Encyclopedia of Philosophy* (Spring 2018 edition), E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/propositions/>.

McGrath, S., 2014, "Relax? Don't Do It! Why Moral Realism Won't Come Cheap," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 9, Oxford University Press.

Menzel, C., 1993, "The Proper Treatment of Predication in Fine-Grained Intensional Logic," in J. E. Tomberlin (ed.), *Philosophical Perspectives*, vol. 7, Atascadero: Ridgeview Publishing Company.

Millikan, R., 1984, *Language, Thought and Other Biological Categories*, Cambridge MA: MIT Press.

Mintz-Woo, K., 2017, "On Parfit's Ontology," *Canadian Journal of Philosophy*, DOI: 10.1080/00455091.2017.1381935.

Nagel, T., 1970, *The Possibility of Altruism*, Oxford University Press.

O'Leary-Hawthorne, J. and Price, H., 1996, "How to Stand up for Non-Cognitivists," *Australasian Journal of Philosophy*, 74(2): 275–292.

Oddie, G., 2005, *Value, Reality, and Desire*, Oxford University Press.

Ogden, C. K. and Richards, I. A., 1923, *The Meaning of Meaning*, New York: Harcourt Brace & Jovanovich.

Oliveira, L. and Perrine, T., 2017, "Cornell Realism, Explanation, and Natural Properties," *European Journal of Philosophy*, 25(4): 1021–1038.

Olson, J., 2014, *Moral Error Theory: History, Critique, Defence*, Oxford: Oxford University Press.

Parfit, D., 2011, *On What Matters*, Volumes 1 and 2, Oxford: Oxford University Press.

Peacocke, C., 2004, *The Realm of Reason*, Oxford: Oxford University Press.

Pérez Carballo, A., 2014, "Semantic Hermeneutics," in A. Burgess and B. Sherman (eds.), *Metasemantics: New Essays on the Foundations of Meaning*, Oxford University Press.

Pérez Carballo, A. and Santorio, P., 2016, "Communication for Expressivists," *Ethics*, 126(3): 607–635.

Price, H., 1994, "Semantic Minimalism and the Frege Point," in S. Tsohatzidis (ed.), *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, London: Routledge.

---, 2015, "From Quasi-Realism to Global Expressivism – and Back Again?" in R. N. Johnson and M. Smith (eds.), *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, Oxford University Press.

Prinz, J., 2004, *Gut Reactions: A Perceptual Theory of Emotion*, New York: Oxford University Press.

---, 2007, *The Emotional Construction of Morals*, Oxford: Oxford University Press.

Railton, P., 1993, "Noncognitivism about Rationality: Benefits, Costs, and an Alternative," *Philosophical Issues*, 4: 46–51.

Ridge, M., 2006a, "Ecumenical Expressivism: Finessing Frege," *Ethics*, 116(2): 302–336.

---, 2006b, "Sincerity and Expressivism," *Philosophical Studies*, 131(2): 487–510.

---, 2009, "Moral Assertion for Expressivists," *Philosophical Issues*, 19: 182–204.

---, 2013, "Disagreement," *Philosophy and Phenomenological Research*, 86: 41–63.

---, 2014, *Impassioned Belief*, Oxford: Oxford University Press.

---, 2015a, "Summary," *Analysis*, 75(3): 433–442.

---, 2015b, "I Might Be Fundamentally Mistaken," *Journal of Ethics and Social Philosophy*, 9(3): 1–21.

---, 2017, "Meeting Constitutivists Halfway," *Philosophical Studies*, DOI: 10.1007/s11098-017-0989-z.

Rosati, C., 2016, "Moral Motivation," *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/moral-motivation/>.

Rosen, G., 1998, "Blackburn's *Essays in Quasi-Realism*," *Nous*, 32(3): 386–405.

---, 2005, "Problems in the History of Fictionalism," in M. Kalderon (ed.), *Fictionalism in Metaphysics*, Oxford: Clarendon Press.

Rubin, M., 2015, "Normatively Enriched Moral Meta-Semantics," *Philosophy and Phenomenological Research*, 91(2): 386–410.

Rumbaugh, D., 1995, "Primate Language and Cognition - Common Ground," *Social Research*, 62(3): 711–730.

Salmon, N., 1989, "Tense and Singular Propositions," in J. Almog, J. Perry and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press.

Savage-Rumbaugh, S. and R. Lewin, 1994, *Kanzi: An Ape at the Brink of Human Mind*, New York: Wiley.

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

---, 2014, *Being Realistic About Reasons*, Oxford: Oxford University Press.

Schroeder, M., 2007, *Slaves of the Passions*, Oxford: Oxford University Press.

---, 2008, *Being For: Evaluating the Semantic Program of Expressivism*, Oxford: Clarendon.

- , 2009, "Hybrid Expressivism: Virtues and Vices," *Ethics*, 119(2): 257–309.
- , 2010, *Noncognitivism in Ethics*, London: Routledge.
- , 2013, "Tempered Expressivism," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 8, Oxford University Press.
- , 2014, "Does Subjectivism Have Subjectivist Consequences?" *Philosophical Perspectives*, 28(1): 278–290.
- , 2015, *Expressing Our Attitudes: Explanation and Expression in Ethics*, Volume 2, Oxford University Press.
- Searle, J., 1975, "Indirect Speech Acts", in P. Cole and J. L. Morgan (eds.), *Speech Acts*, vol. 3, New York: Academic Press.
- Searle, J., and D. Vanderveken, 1985, *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press.
- Setiya, K., 2004, "Hume on Practical Reason," *Philosophical Perspectives*, 18: 365–389.
- Shafer-Landau, R., 2003, *Moral Realism: A Defence*, Oxford: Clarendon Press.
- Silk, A., 2013, "Truth Conditions and the Meanings of Ethical Terms," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 8, Oxford: Oxford University Press.
- , 2015, "How to Be an Ethical Expressivist," *Philosophy and Phenomenological Research*, 91(1): 47–81.
- Sinclair, N., 2006, "The Moral Belief Problem," *Ratio*, 19(2): 249–260.
- , 2011, "Moral Expressivism and Sentential Negation," *Philosophical Studies*, 152(3): 385–411.
- Smith, M., 1987, "The Humean Theory of Motivation," *Mind*, 96(381): 36–61.
- , 1994, *The Moral Problem*, Oxford: Blackwell.
- Soames, S., 2010, *What is Meaning?* Princeton, NJ: Princeton University Press.
- , 2014, "Cognitive Propositions," in King, J. C., Soames, S. and Speaks, J., *New Thinking about Propositions*, Oxford University Press.
- Stevenson, C. L., 1937, "The Emotive Meaning of Ethical Terms," *Mind: A Quarterly Review of Philosophy*, 46: 14–31.
- , 1944, *Ethics and Language*, New Haven: Yale University Press.
- , 1963, *Facts and Values: Studies in Ethical Analysis*, New Haven: Yale University Press.

Stoljar, D., 1993, "Emotivism and Truth Conditions," *Philosophical Studies*, 70(1): 81–101.

Strandberg, C., 2015, "Options for Hybrid Expressivism," *Ethical Theory and Moral Practice*, 18(1): 91–111.

Street, S., 2009, "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues*, 19: 273–298.

---, 2010, "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass*, 5(5): 363–384.

---, 2011, "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have it Both Ways," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 6, Oxford: Oxford University Press.

Streumer, B., 2013, "Do Normative Judgements Aim to Represent the World?" *Ratio*, 26(4): 450–470.

---, 2017, *Unbelievable Errors: An Error Theory about All Normative Judgements*, Oxford University Press.

Stringer, R., 2018, "Realist Ethical Naturalism for Ethical Non-Naturalists," *Philosophical Studies*, 175(2): 339–362.

Sturgeon, N., 1985, "Moral Explanations," in D. Copp and D. Zimmerman (eds.), *Morality, Reason and Truth*, Totowa, NJ: Rowman and Allanheld.

---, 2003, "Moore on Ethical Naturalism," *Ethics*, 113(3): 528–556.

Suikkanen, J., 2009, "The Subjectivist Consequences of Expressivism," *Pacific Philosophical Quarterly*, 90(3): 364–387.

---, 2017, "Non-Realist Cognitivism, Truth and Objectivity," *Acta Anal*, 32: 193–212.

Tappolet, C., 2011, "Values and Emotions: Neo-Sentimentalism's Prospects," in C. Bagnoli, *Morality and the Emotions*, Oxford University Press.

Tennant, N., 2007, "Existence and Identity in Free Logic: A Problem for Inferentialism?" *Mind*, 116: 1055–1078.

Tiffany, E., 2007, "Deflationary Normative Pluralism," *Canadian Journal of Philosophy*, 37(5): 231–62.

Toppinen, T., 2013, "Believing in Expressivism," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 8, Oxford: Oxford University Press.

---, 2018, "Non-Naturalism Gone Quasi: Explaining the Necessary Connections between the Natural and the Normative," in R. Shafer-Landau, *Oxford Studies in Metaethics*, vol. 13, Oxford: Oxford University Press.

---, 2018, "Disagreement (also) for Expressivists," unpublished manuscript.

Unwin, N., 1999, "Quasi-Realism, Negation and the Frege-Geach Problem," *Philosophical Quarterly*, 49: 337-352.

Van Roojen, M., 1996, "Expressivism and Irrationality," *The Philosophical Review*, 105(3): 311-355.

---, 2006, "Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 1, Oxford: Oxford University Press.

Wallman, J., 1992, *Aping Language*, Cambridge: Cambridge University Press.

Wedgwood, R., 2007, *The Nature of Normativity*, Oxford: Clarendon.

Williams, B., 1985, *Ethics and the Limits of Philosophy*, Cambridge MA: Harvard University Press.

Williamson, T., 2000, *Knowledge and its Limits*, Oxford: Oxford University Press.

Wright, C., 1992, *Truth and Objectivity*, Cambridge MA: Harvard University Press.

